ARTICLES

ON JUDICIAL DISCRETION IN STATUTORY INTERPRETATION

H. MILES FOY, III*

TABLE OF CONTENTS

* Professor of Law, Wake Forest University. A.B. 1967, University of North Carolina at Chapel Hill; M.A. 1968, Harvard University; J.D. 1972, University of Virginia. I would like to thank Peter Metzger for his excellent assistance in researching this Article.

Statutory interpretation is disagreeable. Judges complain about it; lawyers approach it with distaste; law students shy away. The problem has much to do with the conventional theories of statutory interpretation. According to these theories, courts are supposed to use certain interpretive methods to resolve statutory questions. Unfortunately, these methods are sometimes analytically insufficient. A conscientious judge may perform all of the conventional tasks: the judge may consider the objective meaning of the relevant text, the legislative intent reflected in the text or elsewhere, the general policies and purposes behind the legislation, the traditional canons of interpretation, the "rules of clear statement," and so forth. On the basis of such considerations the judge may determine that the statute could be interpreted to mean either *A* or *B*, yet the judge may also determine that there is no persuasive conventional reason to prefer *A* over *B* or *B* over *A*. How, then, should the judge interpret the statute? Which interpretation should be preferred?

The conventional theories of statutory interpretation make no provision for cases of this kind. They implicitly assume that if a competent judge applies the conventional methods diligently and perceptively, the judge will be persuaded that interpretation *A* is legally preferable to interpretation *B*, or vice versa. The statute's legal meaning will not remain in doubt. The judge will determine the statute's meaning through a process of relentless legal reasoning grounded in conventionally prescribed considerations.

Yet refractory cases do occur. The conventional methods of statutory interpretation are analytically sufficient in most cases, to be sure, but in some cases even the most assiduous jurist will encounter frustration while attempting to find persuasive conventional reasons for preferring one interpretation of a statute over others, and in such cases the conventional theories of statutory interpretation are a fertile source of judicial embarrassment. In effect, they require honest judges to pretend that conventional methods of interpretation are decisive, even when they are not, and this encourages obfuscation and arbitrariness in the making of judicial decisions.

This Article examines this problem and proposes a modest cure—one that would require a small adjustment in the conventional theories of statutory interpretation. This Article argues that two things need to be done. First, the courts must recognize that there are cases in which the conventional methods of statutory interpretation are useful but analytically *insufficient*: The conventional methods almost always establish plausible boundaries for interpretation, but sometimes they fail to provide persuasive reasons for specific interpretive choices. Second, the courts must acknowledge that when the conventional methods are indeterminate, statutory interpretation requires the exercise of judicial discretion—prudent

choice within legal bounds.[1] That is to say, if a judge finds that a statute can be interpreted plausibly to mean either *A* or *B*, and if the judge is not persuaded on conventional grounds that one interpretation is preferable to the other, the judge must be free to declare that there are two interpretations of the statute which are equally defensible in law, and the judge must be permitted to make a prudent, discretionary choice between them. Our present theories of statutory interpretation do not expressly authorize decisionmaking of this sort. This Article argues that they should. Discretionary interpretation is inevitable in some cases, and the theories of statutory interpretation should recognize that fact.

In the end, the argument presented in this Article is simply a plea for greater realism in statutory interpretation. Even the most casual observer of judicial affairs understands that judges do exercise discretion in statutory interpretation from time to time. Discretionary interpretation is not rare. Yet the courts themselves are reluctant to admit that they ever exercise discretion when they interpret statutes,[2] and there is no established doctrine that defines (or confines) the practice. The absence of such a doctrine creates serious difficulties for the law, as will be shown.

In the discussion that follows, the conventional theories of interpretation are reviewed, their occasional insufficiency discussed, and some interesting cases that illustrate the point are examined. This Article describes how a doctrine of discretionary interpretation would work in actual practice and how it would improve the interpretive process. This Article argues that if the courts were willing to adopt a doctrine of discretionary interpretation for cases that cannot be resolved persuasively by conventional means, they would promote both clarity and rigor in statutory interpretation and strengthen the rule of law.

I. THE CONVENTIONAL THEORIES OF STATUTORY INTERPRETATION

The conventional theories of statutory interpretation are organized around three well-worn principles. The first is the concept of "legislative intent." Many judges believe that statutes should be interpreted according to the legislature's "intent" and that conscientious interpreters must therefore concern themselves with the legislative mind.[3] These judges are

1. WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 647 (2002) (defining *discretion* as a "power of free decision or choice within certain legal bounds").

2. There are exceptions, of course. Judges sometimes concede, in moments of exceptional candor, that interpretive questions are not always questions of "law." *See infra* text accompanying notes 62–70, 93–99.

3. The cases that reflect this idea are so numerous that it would be redundant to cite more than a few. The following cases are representative of the genre: *Philbrook v. Glodgett*, 421 U.S. 707, 713 (1975) ("Our objective . . . is to ascertain the congressional intent and give

not deluded. They understand that the concept of legislative intent is philosophically problematic and that the search for legislative intent is sometimes difficult, yet they believe that the ultimate purpose of statutory interpretation is to align judicial action with legislative will. Statutes, after all, are the work of the legislature. They express legislative power, not judicial power, and it is the legislature's judgment that counts. Courts must therefore make an honest effort to determine what the legislature wants, and they must resist the temptation to hijack the legislature's work under the guise of interpretation.

A second fundamental interpretive principle rests upon the assumption that statutory language has an "objective meaning." Some judges believe that statutes should be interpreted, not according to the intent of the legislative author, but according to the meaning that a reasonably intelligent reader would attribute to the statutory text, given the conventions of the English language and the relevant legal context. This objective meaning may or may not coincide with the meaning the legislature actually intended at the time of enactment, but it should be legally controlling in most instances.[4] Judges who favor this theory deserve the benefit of the doubt. They do not claim that a statute's objective meaning is always easy to determine, and occasionally they demonstrate commendable flexibility by interpreting statutes according to other principles.[5] But they insist that there are sound reasons for taking an objective approach to statutory interpretation generally, and they criticize the misguided souls who traffic in the loose currency of legislative intent.

A third fundamental principle of interpretation is the notion that preexisting law influences the legal meaning and the legal consequences of legislative action. For example, there are preexisting constitutional principles that impose substantive limitations on legislative power;[6] there are preexisting statutory schemes with which new legislation must

---

effect to the legislative will."); *Sinclair Refining Co. v. Atkinson*, 370 U.S. 195, 202–03 (1962) (where congressional intent is discernible, courts must give effect to that intent); *Flora v. United States*, 357 U.S. 63, 65 (1958) (courts must give effect to congressional intent); and *Vermilya-Brown Co. v. Connell*, 335 U.S. 377, 386 (1948) (statutory language must be construed to effectuate lawmakers' intent). *See generally* 2A STATUTES AND STATUTORY CONSTRUCTION § 45:05 (Norman J. Singer & J.D. Shambie Singer eds., 7th ed. 2007) (detailing the intent of the legislature as a method of statutory construction).

4. Justice Antonin Scalia explained this philosophy in a long essay. Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, *in* A MATTER OF INTERPRETATION 3 (Amy Guttmann ed., 1997).

5. *See* Green v. Bock Laundry Mach. Co., 490 U.S. 504, 527–30 (1989) (Scalia, J., concurring) (holding that interpretation that departs from the ordinary meaning of text is justified where ordinary meaning is "unthinkable" and there is no indication that the legislature actually intended the "unthinkable").

6. *See* 2A STATUTES AND STATUTORY CONSTRUCTION, *supra* note 3, § 45:11.

sometimes be harmonized;[7] there are preexisting rules of substantive common law that occasionally affect statutory interpretation in one way or another;[8] and there are preexisting canons and principles of interpretation that are designed for use in the interpretive process itself.[9] For "intentionalists" and "objectivists" alike, the ambient law—the law that envelopes legislative action—is an important factor in the interpretive process.

The dominant modern theories of statutory interpretation reflect various admixtures of these three elementary principles. For many years most judges embraced a soft version of intentionalism, with a drop of objectivism thrown in for good measure. They generally assumed that they were supposed to interpret statutes by determining and honoring legislative intent. They understood, of course, that cases would arise in which it would be impossible to discern specific legislative intent with respect to the specific issues they were called upon to resolve, yet they were convinced that they could deal with these cases responsibly by considering the legislature's general policies and purposes and by interpreting statutes in such a way as to advance those policies and purposes. The absence of specific legislative intent with regard to a specific interpretive issue would not defeat the interpretive enterprise; instead, sufficient guidance could usually be found in general indications of legislative will.[10]

Judges who accepted this way of thinking scrutinized statutory texts to determine what the legislature's intentions, policies, and purposes actually were, but in many cases they examined other things as well. For much of the twentieth century, especially in the federal courts, judges routinely reviewed legislative history and other extra-textual materials as they

7. *See* 2B STATUTES AND STATUTORY CONSTRUCTION §§ 51.01–51.03 (Norman J. Singer & J.D. Shambie Singer eds., 7th ed. 2008).

8. *See id.* §§ 50.1–50.5.

9. *See, e.g.*, 2A STATUTES AND STATUTORY CONSTRUCTION, *supra* note 3, §§ 47:17, 47:23 (discussing the doctrines of *ejusdem generis* and *expressio unius est exclusion alterius*, respectively).

10. *See* United States v. Bacto-Unidisk, 394 U.S. 784, 799 (1969) (stating that where statutory language is "insufficiently precise," the statute must be construed in light of statutory purpose); *see also* United States v. Bornstein, 423 U.S. 303, 310 (1976) (holding that courts must give faithful meaning to statutory language in light of evident statutory purpose); Comm'r v. Bilder, 369 U.S. 499 (1962) (asserting that statutes must be given effect in accordance with manifest congressional purpose). Indeed, during the middle decades of the twentieth century certain influential scholars came to believe that the concept of legislative "purpose" was so central to statutory interpretation that the interpretive process could best be described, not as a search for legislative "intent," but as an attempt to determine and effectuate legislative "purposes." *See* HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 1374–80 (1994).

attempted to understand the legislative mind.[11] In some instances, however, they adopted an objective, text-based approach to interpretive problems. Sometimes they found statutory language to be so plain, so specific, and so sensible that there was little room for argument about what the statute ought to mean. In such cases they were happy to honor the objective meaning of the text. The assumption here was that the objective meaning of the text probably coincided with the meaning the legislature actually had in mind.[12]

Over the last twenty years or so, more and more federal judges have adopted an objective approach to interpretive questions. Some of them profess to be largely unconcerned with actual legislative intent.[13] They insist that most statutory questions can be resolved satisfactorily on the basis of an objective reading of the relevant language. They sometimes call themselves "textualists." For them, the text, objectively considered, *is* the law. The legislative intent behind the text is irrelevant for most purposes.[14]

Other judges take a position that falls somewhere between thoroughgoing intentionalism on the one hand and thoroughgoing objectivism or textualism on the other. They profess to be concerned with actual legislative intent, but they are inclined to treat the objective meaning of the statutory text as a sufficient indicator of actual legislative intent, and they prefer to settle statutory questions on the basis of the text alone. Even in doubtful cases, they are reluctant for various reasons to accord legal weight to legislative history and other extra-textual evidence of legislative will.[15]

---

11. For a perceptive, contemporaneous account of mid-century practices concerning legislative history, see Patricia M. Wald, *Some Observations on the Use of Legislative History in the 1981 Supreme Court Term*, 68 IOWA L. REV. 195 (1983).

12. *See, e.g.*, United States v. Locke, 471 U.S. 84, 95 (1985) (stating that a literal reading of Congress's words is generally the only proper reading of those words).

13. *See, e.g.*, *In re* Sinclair, 870 F.2d 1340 (7th Cir. 1989); *see also* Frank H. Easterbrook, *Legal Interpretation and the Power of the Judiciary*, 7 HARV. J.L. & PUB. POL'Y 87 (1984).

14. Scalia, *supra* note 4, at 16–25.

15. This may well be the most widely accepted position today. As early as 1992, Justice Breyer sensed that judicial attitudes were shifting and that federal judges were placing less and less reliance on legislative history in their search for legislative intent. *See* Stephen Breyer, *On the Uses of Legislative History in Interpreting Statutes*, 65 S. CAL. L. REV. 845, 846 (1991) (noting the Court's changing reliance on legislative history). During the 1990s, within the work of the Supreme Court itself, there was a precipitous decline in the number of cases in which the Justices relied on legislative history. *See* Michael H. Koby, *The Supreme Court's Declining Reliance on Legislative History: The Impact of Justice Scalia's Critique*, 36 HARV. J. ON LEGIS. 369 (1999). Today, in the lower federal courts, judges routinely emphasize the importance of the text and often find the text to be so "plain" that recourse to legislative history is unjustified. *See, e.g.*, *In re* Hart, 328 F.3d 45 (1st Cir. 2003); *In re* Kolich, 328 F.3d 406 (8th Cir. 2003); Cairns v. Franklin Mint Co., 292 F.3d 1139 (9th Cir. 2002); Nat'l Pub. Radio, Inc. v. FCC, 254 F.3d 226 (D.C. Cir. 2001); Abdul-Akbar v. McKelvie, 239 F.3d

II. THE OCCASIONAL INDETERMINACY OF CONVENTIONAL INTERPRETIVE METHODS

What is of interest is the implicit assumption upon which all of the conventional theories of interpretation rest. All of them require the courts to employ certain interpretive methods, and all of them assume that if the courts employ these methods competently and consistently, the answers to statutory questions can be found. Is this assumption valid?

Statutory interpretation, unlike literary or historical interpretation, is a governmental process. It must satisfy the needs of the government and comply with the principles that regulate governmental action. When a court is confronted with a statutory question, it must hear the contentions of the parties and make a decisive choice. It is not permitted to embrace all possible interpretations of the statute. It must choose a single interpretation and reject others as legally incorrect. Moreover, the preferred interpretation must be case-specific. The court is not called upon to say what the statute means in general. It must choose an interpretation that resolves the specific issue presented in the controversy before it. Finally, and above all, the court's interpretation of the statute must not be arbitrary. Due process forbids arbitrary governmental action. The court is not entitled to decide the case by flipping a coin, and it may not prefer one interpretation to another because the plaintiff is better looking than the defendant. The court must have a good reason for preferring one interpretation of the statute over other possible interpretations, and it must be willing to disclose that reason to the litigants and the world at large. This is what our legal traditions require.

To be successful, a general theory of statutory interpretation must provide the courts with a conceptual framework that will allow them to perform the function described above. A successful theory of statutory interpretation must help the courts find good reasons for adopting case-specific interpretations of statutory law, and it must allow the courts to disclose those reasons candidly. The conventional theories of statutory interpretation pass this test in most instances. Competent judges can usually find good reasons for interpreting statutes in decisive, case-specific ways, such as assessing the legislature's intentions, policies, and purposes, determining the objective meaning of the statutory text, consulting existing rules of law and interpretation, or doing some combination of these things.

Yet the conventional theories of statutory interpretation do not *always* pass this test. Sometimes the intended meaning of the relevant text is too unclear to provide solid ground for case-specific interpretation; sometimes

---

307 (3d Cir. 2001).

the objective meaning of the text is intractably ambiguous; sometimes the underlying legislative policies and purposes are too diffuse or contradictory to support persuasive, case-specific inferences; sometimes the ambient law and the traditional rules of interpretation have nothing definitive to say about the precise issue the court must decide. If a judge attempts to employ conventional interpretive methods in such a case, the judge will find no good reason to prefer one interpretation of the statute over other possible interpretations, and the judge will grasp at straws in the attempt to resolve the issue.

It is instructive to compare the conventional theories of statutory interpretation with the law of evidence. Was the traffic light red or green at the time of the accident? Did the shooter intend to kill the decedent? Did toxins in the groundwater cause the plaintiff's illness? The law of evidence establishes rules and procedures for deciding questions of this kind, yet it does not assume that the answers can always be found. Sometimes the evidence will be too scanty, too evenly balanced, too contradictory, or too obscure. Sometimes the finder of fact will be unable to draw a firm conclusion about the color of the traffic light, the shooter's intent, or the etiology of the disease. The law of evidence does not deny the possibility of uncertainty concerning factual questions in general, and it provides the courts with a principled way to deal with such uncertainty. It creates special rules that spell out the legal consequences of uncertainty in a comprehensive way. These rules are called "burdens of proof."[16]

The conventional theories of statutory interpretation are quite unlike the law of evidence in this respect. If the law of evidence recognizes that factual questions may sometimes be unanswerable, the conventional theories of statutory interpretation assume that the courts will almost always be able to resolve statutory questions through the application of conventional interpretive methods, and there is no general provision for cases in which statutory questions *cannot* be resolved in this way. The assumption here is that the conventional methods will work as long as the courts employ them consistently and competently. Yet, this assumption is belied by experience. Intractable statutory ambiguity is simply a fact of legal life. Indeed, it is constitutionally unavoidable, for various reasons.[17]

This problem is neither academic nor harmless. The occasional indeterminacy of conventional interpretive methods creates grave difficulties for the judiciary. Suppose that a competent judge is called upon to resolve a statutory question. Suppose that the judge considers all of the

16. *See generally* 2 MCCORMICK ON EVIDENCE §§ 336–349 (John W. Strong ed., 5th ed. 1999).

17. *See infra* text accompanying notes 104–06.

conventional arguments that can be made in the case, and suppose that none of the arguments are ultimately persuasive. How can the judge resolve the issue? Given the restrictions imposed by conventional interpretive theory, there are only two courses of action that the judge can take, and both of them are problematic. Both are discussed below.

A. Deciding Cases on Unconventional (and Undisclosed) Grounds

A conscientious judge who is not persuaded by conventional arguments about a statute's meaning may conclude that there are other reasons—unconventional reasons—for interpreting the statute in one way or another, and the judge may be willing to decide the case on that basis. In other words, if the conventional methods of statutory interpretation are indeterminate, the judge may allow a personal sense of justice, equity, practicality, or sound public policy to determine the outcome. This is a responsible way to decide such cases, but the conventional theories of statutory interpretation make no express provision for it. On the contrary, they assume that competent judges will be able to resolve statutory questions on the basis of a process of legal reasoning involving conventional considerations—legislative intent, the objective meaning of the text, the traditional rules of interpretation, and so forth. If a judge has unconventional reasons for preferring one interpretation of a statute to another, the judge must nevertheless mount a conventional defense of the decision and must downplay or conceal the real reasons for the ultimate interpretive choice.

An important recent case, *Ledbetter v. Goodyear Tire & Rubber Co.*, illustrates this phenomenon rather clearly.[18] The case involved a sex discrimination claim under Title VII of the Civil Rights Act of 1964.[19] The plaintiff, Lilly Ledbetter, had worked for the Goodyear Tire and Rubber Company (Goodyear) for a number of years. In 1998, shortly before her retirement, she filed a complaint with the Equal Employment Opportunity Commission (EEOC), alleging that Goodyear had paid her substantially less than similarly situated male employees. Eventually, she submitted her claim to the federal district court and won a jury verdict after a trial on the merits. The district court entered judgment against Goodyear for back wages and damages, and Goodyear appealed. The U.S. Court of Appeals for the Eleventh Circuit reversed the judgment, holding that Ledbetter had failed to file her complaint within the time allowed by Title VII. The Supreme Court granted certiorari and affirmed the decision of the court of

18. 550 U.S. 618 (2007).

19. Civil Rights Act of 1964 §§ 703(a)(1), 706(e)(1), 42 U.S.C. §§ 2000e-2(a)(1), 2000e-5(e)(1) (2000).

appeals.[20]

At the heart of the controversy was a provision of Title VII that required a claimant to file a complaint "within [180] days after the alleged unlawful employment practice occurred."[21] Ledbetter proved, and the jury found, that Goodyear had set her salary at a low level because of her sex, but she was unable to prove that Goodyear had made this decision within 180 days prior to the filing of her EEOC complaint. Instead, she proved that Goodyear had made discriminatory salary-setting decisions in previous years, outside the 180-day filing period, and that Goodyear had continued to pay her at a low level during the 180-day filing period as a result of those decisions.[22]

Did the 180-day filing provision bar Ledbetter's claim? Goodyear argued that it did. According to Goodyear, an unlawful employment practice "occurred" for purposes of the statute whenever an employer made a salary-setting decision on the basis of sex; therefore, the statute required the injured employee to file her complaint within 180 days after the salary-setting decision was made. Because Ledbetter had not filed her complaint within 180 days after Goodyear had made its unlawful decisions, her complaint was time-barred.[23] Ledbetter argued in opposition that the Court should interpret the statute more broadly. Perhaps it was true that an unlawful employment practice "occurred" when an employer made a salary-setting decision on the basis of sex, but it was also true that the unlawful "practice" continued to occur as long as the employer continued to make low payments in implementation of the original decision. Ledbetter had filed her complaint at a time when Goodyear's unlawful pay practice was still continuing. The practice had not yet ceased to occur. Therefore, the complaint was timely.[24]

A closely divided Supreme Court accepted Goodyear's interpretation of the statute. Writing for a five-Justice majority, Justice Alito held that Goodyear's discriminatory salary-setting decisions were discrete events, that the statutory filing period began to run with the occurrence of these events, and that Ledbetter's complaint was untimely because it had not been filed within 180 days of the occurrence of these events. Accordingly, the Court affirmed the judgment of the court of appeals, and Ledbetter, who had suffered substantial financial losses because of unlawful conduct during the 180-day filing period, received nothing for her trouble.[25]

---

20. *Ledbetter*, 550 U.S. at 621–23.
21. 42 U.S.C. § 2000e-5(e)(1).
22. *Ledbetter*, 550 U.S. at 624–25.
23. *Id.* at 622.
24. *Id.* at 624.
25. *Id.* at 632 (majority opinion), 643–44 (Ginsburg, J., dissenting).

Justice Ginsburg filed a vigorous dissent. She argued that Ledbetter's claim was timely because it had been filed while Goodyear was still paying her at a discriminatory rate. The unlawful pay practice was still occurring; therefore, Ledbetter's complaint was not late. Three Justices agreed with Justice Ginsburg.[26]

One may sympathize with Ledbetter in this case, or one may sympathize with Goodyear, but conventional interpretive considerations did not clearly favor either party. There was no indication that Congress had actually considered the specific issue presented in the case, and it was impossible to argue the case one way or the other on the basis of clear evidence of specific legislative intent. The objective meaning of the relevant statutory language shed no light on the problem. A discriminatory salary-setting decision was clearly an unlawful employment practice within the objective meaning of the statute, and it surely occurred at the time it was made. Yet payments that were made pursuant to a discriminatory salary-setting decision were surely a continuation of the "unlawful employment practice," objectively speaking. The practice did not end with the initial decision. Thus, the question was this: Did the statute require the employee to file her complaint within 180 days after the practice *began* to occur, or did it allow her to file her complaint within 180 days after the practice *ceased* to occur? The language of the statute simply did not address this point. It provided only that the claimant was to file her complaint within 180 days after the practice occurred.

Nor was guidance to be found in the conventional rules of statutory interpretation. Justice Alito and Justice Ginsburg could cite only one conventional rule of interpretation in the course of their two opinions. It was the *Chevron* rule, which requires the courts to defer to certain administrative interpretations of ambiguous statutory language;[27] and Justice Alito cited this rule only for the purpose of noting that it did *not* apply to this case.[28]

If the demonstrable intentions of Congress, the objective meaning of the statutory language, and the conventional rules of statutory interpretation did not favor either party, upon what considerations did Justice Alito and Justice Ginsburg rely? Both attempted to rely on the Court's prior decisions interpreting the filing provision, yet the precedents themselves were in conflict. The Court had previously held that the 180-day filing

---

26. *Id.* at 643–60 (Ginsburg, J., dissenting). The Court's decision proved to be controversial and efforts were made to overturn it by legislation. Those efforts recently succeeded. *See* Lilly Ledbetter Fair Pay Act of 2009, Pub. L. No. 111-2, 123 Stat. 5 (stating specifically that this legislation is a response to the Court's decision in *Ledbetter*).

27. *Ledbetter*, 550 U.S. at 643 n.11 (majority opinion), 656 n.6 (Ginsburg, J., dissenting).

28. *Id.* at 643 n.11 (majority opinion).

provision could be interpreted broadly, as Justice Ginsburg proposed, allowing complaints to be filed long after an employer had instituted an unlawful employment practice that continued to occur over time,[29] but the Court had also held that the filing provision could be interpreted strictly in certain instances, barring complaints that were filed more than 180 days after the occurrence of discrete discriminatory acts that had continuing effects.[30]

And if the case law was in conflict, the relevant statutory policies were in conflict as well. Justice Alito's interpretation of the filing provision was consistent with the obvious statutory policy against litigating stale claims. His holding required each claimant to prove that the employer had instituted a discriminatory pay practice no more than 180 days prior to the filing of the complaint, and this tended to ensure that at the time of the filing of the complaint there would be fresh evidence of the employer's discriminatory intent, which was the central element in the employee's Title VII case. But Title VII also expressed a strong policy against sex discrimination in the workplace, and it created a remedial mechanism for the benefit of persons like Ledbetter, who had been injured by discriminatory conduct *within* the 180-day period. Justice Ginsburg's interpretation of the statute was clearly consistent with that policy. It validated the claims of employees who had been injured by discriminatory pay practices that continued during the 180-day filing period, even though the practices had begun more than 180 days prior to the filing of the complaint.

In sum, both Justice Alito's and Justice Ginsburg's interpretations of the statute were consistent with strong (and obvious) statutory policies. Nothing in the statutory text, the specific intentions of Congress, the conventional rules of interpretation, or the prior decisions of the Court required these Justices to favor one policy over the other. The legal calculations in the case were substantially in equipoise, and the Justices simply had to make a choice. But how were they to choose?

---

29. *See* Bazemore v. Friday, 478 U.S. 385, 407 (1986) (per curiam) (addressing an employment practice that had been implemented over ten years prior to the suit); *see also* Nat'l R.R. Passenger Corp. v. Morgan, 536 U.S. 101, 117–18 (2002) (noting that an employee need only file a complaint within the statutory time period of any act that is part of the hostile work environment regardless of how long it has been since the hostile work practices in general began).

30. *See, e.g.*, Lorance v. AT&T Techs., Inc., 490 U.S. 900 (1989) (pertaining to a suit by female employees who challenged employer's seniority system as discriminatory); Del. State Coll. v. Ricks, 449 U.S. 250 (1980) (addressing a professor's claim that he had been denied tenure on the grounds of national origin discrimination); United Air Lines, Inc. v. Evans, 431 U.S. 553 (1977) (involving a suit by a female flight attendant who was forced to resign after getting married).

If I had been a Justice in *Ledbetter*, I would have concluded that conventional interpretive considerations were indeterminate, and I would have decided the case on unconventional grounds. There is no way to know whether any of the nine Justices who participated in *Ledbetter* were moved by unconventional considerations, as I would have been, but I suspect that some of them were, and the final vote in the case tends to confirm this suspicion. The Court split cleanly along ideological lines. Justices Alito, Roberts, Kennedy, Scalia, and Thomas, four conservatives and one moderate, adopted Goodyear's interpretation of the statute—the strict interpretation. The four Justices who accepted the more lenient interpretation of the statute—Justices Ginsburg, Stevens, Souter, and Breyer—are moderates or liberals who have shown some sensitivity to the interests of employees from time to time. If the legal ingredients in the case were evenly balanced, perhaps these nine Justices, or some of them, simply voted their predilections, and the five outvoted the four.

This interpretation of *Ledbetter* is not a criticism of the Court or of any of the Justices who participated in the decision, and to underscore this point, I must reiterate my fundamental contention: Conscientious judges cannot always resolve statutory questions on the basis of conventional interpretive considerations. In *Ledbetter* there were two plausible interpretations of the 180-day filing provision. Each found some support in convention, yet there was no persuasive conventional reason to prefer one over the other. This meant that the Court was confronted with a choice between alternatives that were in some sense equally lawful; and if a conscientious Justice accepted either one of these two interpretations because of his or her own sense of justice, equity, practicality, or sound public policy, he or she would have violated no law or judicial duty, in my opinion. The choice was essentially discretionary. Within the bounds created by the statute, the choice could have been made for any prudent, nonarbitrary reason.

I recognize that this way of thinking about statutory interpretation is inconsistent with conventional interpretive theory, but I believe that it makes sense. Indeed, it is the only realistic way to think about a case such as *Ledbetter*.

The principal problem with *Ledbetter*, in the end, was not the judgment itself but the opinions of the Justices—both the majority opinion and the dissenting opinion. These opinions presented conventional arguments about the statute's meaning, but the arguments were not persuasive. I do not fault Justice Alito or Justice Ginsburg for this. It was impossible to argue this case persuasively on the basis of conventional considerations, and if these Justices attempted to construct conventional arguments in support of their differing views of the case, they were simply doing what the conventional theories of statutory interpretation required them to do. They

were doing the best they could, given the restrictions imposed upon them by the conventional theories, which discourage judicial candor in cases that cannot be decided persuasively on conventional grounds.

B. Deciding Cases on Trivial Conventional Grounds

There is a second strategy that a judge can follow if the judge is initially unconvinced by conventional arguments about a statute's specific meaning. Instead of deciding the case on unconventional grounds, the judge can suppress doubts and proceed in the conventional way. The judge can refuse to accept the conclusion that conventional considerations are indeterminate, and can persist in the effort to find a conventional basis for preferring one interpretation over another.

This strategy is successful some of the time, but it carries a substantial risk. If a judge insists on finding a conventional reason to justify an interpretive decision in an evenly balanced case, the judge will be tempted to attribute legal significance to triviality. A stray comment buried in the legislative history here or there,[31] or an obscure semantic distinction between one shade of objective meaning and another,[32] can tip the scales of justice, or so the judges tell us. But when life, liberty, property, and public policy hang in the balance, triviality should not be decisive. If the consequences of a judgment are weighty, the reasons for the judgment should be weighty as well; and if the reasons for a judgment are thin, justice and the appearance of justice suffer.

Consider *Chapman v. United States*.[33] The defendants in *Chapman* were convicted of distributing lysergic acid diethylamide (LSD). The case was a difficult one because of the imprecise language of the relevant statute and the peculiarities of the LSD trade. One dose of pure LSD is so light in weight that it must be sold on the street through the use of a carrier medium.[34] In some instances the pure drug is dissolved in a solvent, and the solvent is sprayed on blotter paper. The blotter paper is then cut into one-dose squares. Customers purchase the squares and ingest the drug by licking or swallowing the squares or by dropping them into beverages.[35] The defendants in *Chapman* were convicted of selling ten sheets of blotter paper bearing about fifty milligrams of pure LSD, in violation of 21 U.S.C.

---

31. *See* Kosak v. United States, 465 U.S. 848, 855 (1984) (relying on references in the legislative history, "though meager," as support for the Court's holding).

32. *See* Chisom v. Roemer, 501 U.S. 380 (1991) (examining the scope of coverage included in § 2 of the Voting Rights Act).

33. 500 U.S. 453 (1991).

34. *Id.* at 457.

35. *Id.*

§ 841(a).[36] The statute provided that the defendants should receive sentences of at least five years if they had distributed one gram or more of "a mixture or substance containing a detectable amount" of LSD.[37] The fifty milligrams that the defendants had distributed would *not* have subjected them to the mandatory five-year minimum, but the trial judge concluded that the blotter paper itself was a "mixture . . . containing a detectable amount" of LSD and that the weight of the paper (about 5.7 grams) should be added to the weight of the pure drug to determine the appropriate sentences. Accordingly, the defendants were sentenced to the mandatory minimum.[38]

The defendants appealed their sentences to the U.S. Court of Appeals for the Seventh Circuit, and the case was argued three times before that court. After a rehearing en banc, the court of appeals affirmed the judgment of the district court, with five judges dissenting.[39] The court of appeals held that the blotter paper was a "mixture . . . containing a detectable amount" of LSD and that the weight of the paper should therefore be counted in determining the sentences.[40] On certiorari, the Supreme Court affirmed the decision of the court of appeals. Writing for a seven-Justice majority, Chief Justice Rehnquist held that blotter paper stained with LSD was indeed a "mixture" within the meaning of the statute, that the weight of the blotter paper should be taken into account in determining the sentences, and that the defendants were subject to the mandatory minimum.[41] Justices Stevens and Marshall dissented.[42]

Chapman is a disturbing case. The decision ultimately turned on the interpretation of a single word: mixture. Was blotter paper bearing crystals of LSD a "mixture" in the statutory sense? There was no indication that Congress was well versed in the esoteric practices of the LSD trade, and thus there was no evidence that Congress had specifically intended for the courts to treat LSD-stained blotter paper as a mixture for the purpose of applying the sentencing scheme. There was, however, abundant evidence of congressional intent concerning sentencing in general. The statute applied to a number of controlled substances, not to LSD alone, and it was clear from the language of the statute that in most cases involving mixtures, Congress intended for punishment to be determined by the gross weight of

36. *Id.* at 455.
37. 21 U.S.C. § 841(b)(1)(B)(v) (2000).
38. *Chapman*, 500 U.S. at 455–56.
39. United States v. Marshall, 908 F.2d 1312 (7th Cir. 1990) (en banc).
40. *Id.* at 1318.
41. *Chapman*, 500 U.S. at 462, 468.
42. *Id.* at 468–77 (Stevens, J., dissenting).

the mixture, not by the net weight of the pure drug;[43] however, this did not mean that blotter paper sprinkled with LSD was a mixture. With respect to that specific question, the intentions of Congress were utterly obscure.

If the intended meaning of the word mixture was uncertain, what was the objective meaning of the word? Chief Justice Rehnquist approached this question by invoking the well-settled rule that courts should interpret statutory language according to its "ordinary" meaning.[44] He then consulted certain dictionaries, and he discovered two definitions of the word mixture that seemed to fit the case. According to these dictionary definitions, a mixture was a portion of matter consisting of two or more components retaining a separate existence, even though the particles of one were diffused among the particles of the other. Blotter paper bearing crystals of LSD was arguably a mixture in that sense, and Chief Justice Rehnquist so held.[45]

But this interpretation of the word was not the only plausible interpretation. A reasonably intelligent English speaker would not ordinarily use the word mixture to describe a necktie stained with soup or a napkin stained with cod liver oil, and a plausible argument could be made that a reasonably intelligent English speaker (or legislator) would not ordinarily use the word mixture to describe a piece of blotter paper stained with LSD. It would not be impossible to use the word mixture in that way, but such a usage would be unusual. It would not be ordinary. More appropriate language could easily be found; and if the rules of statutory interpretation require the courts to read statutes in the light of ordinary English usage, as Chief Justice Rehnquist suggested, then a plausible argument could have been made that blotter paper sprinkled with LSD was not a mixture in the statutory sense.

Thus, there were two plausible interpretations of the statute. There was Chief Justice Rehnquist's interpretation, which depended on certain dictionary definitions of the word mixture, and there was a contrary interpretation, which depended on an understanding of ordinary English usage. The Court's task was to choose between these two interpretations. At common law, the "rule of lenity" would have tilted the analysis in the defendants' favor,[46] yet the rule of lenity has apparently lost its force in federal jurisdictions. The present Supreme Court applies the rule

43. *See, e.g.*, 21 U.S.C. § 841(b)(1)(A)(i), (ii)(I), (III), (iii), (iv) (2000).

44. *Chapman*, 500 U.S. at 462 (majority opinion).

45. *Id.* at 461–62.

46. *See* United States v. Bass, 404 U.S. 336, 347–48 (1971) (cautioning against the punishment of criminal defendants when resolving statutory ambiguity); United States v. Fisher, 6 U.S. (2 Cranch) 358 (1805). *See generally* Lawrence M. Solan, *Law, Language, and Lenity*, 40 WM. & MARY L. REV. 57 (1998).

erratically or not at all, and Chief Justice Rehnquist simply refused to apply it in *Chapman*.[47]

The defendants in *Chapman* made one additional argument that deserves to be mentioned here. They claimed broadly that the statutory scheme was so problematic that major surgery was required. As written, the statute was likely to produce irrational disparities in sentencing. Two drug dealers who sold precisely the same number of doses of LSD might receive substantially different sentences, depending on the weight of the mixtures they chose to employ, and drug kingpins who sold significant quantities of the pure drug might receive lesser sentences than street-level pushers who sold the drug at retail diluted in heavy mixtures of one kind or another. The defendants argued that these potential disparities were so irrational that the constitutionality of the statute was in doubt, and they invited the Court to interpret the statute in such a way as to avoid the constitutional issue. In effect, they asked the Court to read the word mixture out of the statute in cases involving LSD, so that punishment in such cases would turn on the net weight of the pure drug, not the gross weight of any mixture with which the drug was connected.[48] This interpretation would have made it unnecessary for the Court to decide whether blotter paper sprinkled with LSD was a mixture in the statutory sense. The dissenting judges in both the Supreme Court and the court of appeals were inclined to adopt this approach,[49] but a majority of the judges in both courts were unimpressed by the constitutional argument, and left the statute as they found it.[50]

At the end of the day, with the constitutional question pushed conveniently to one side, the Court was obliged to make an unappealing choice between two equally plausible interpretations of an awkward statutory text. According to one interpretation, blotter paper sprinkled with LSD was a mixture. According to another interpretation, blotter paper sprinkled with LSD was not a mixture. Given the demise of the rule of lenity, there was no persuasive conventional reason to prefer either of these interpretations over the other, yet seven Justices suppressed all doubt and insisted that blotter paper was a mixture, even though the defendants' liberty hung in the balance. The reasons given by the Court for preferring this interpretation were so thin that the judgment looks almost arbitrary. Indeed, if the Court had simply flipped a coin and decided the case on that basis, the result would have been just as convincing, and the process would have been far more efficient. *Chapman* dishonors the law because it allowed

47. *Chapman*, 500 U.S. at 464.

48. *See id.* at 464–67.

49. *See id.* at 473–77 (Stevens, J., dissenting); United States v. Marshall, 908 F.2d 1312, 1333 (7th Cir. 1990) (en banc) (Posner, J., dissenting).

50. *Chapman*, 500 U.S. at 468 (majority opinion).

momentous consequences to turn upon an utterly trivial calculation.

III. ALTERNATIVES TO THE CONVENTIONAL THEORIES

There are powerful reasons to resist any change in the conventional way of thinking about statutory interpretation. People who believe in the rule of law take comfort in the grand propositions upon which the conventional theories rest: that a statute's meaning is determined primarily by the form of the statutory text or by the legislature's intentions, policies, and purposes, or by both; that the courts should interpret statutes by giving weight to these legislatively created things, together with preexisting rules of law and interpretation; and that the process of statutory interpretation, which inevitably involves judgment, is nonetheless a process of legal reasoning, which is grounded ultimately in a set of legally prescribed considerations.

Just as there are powerful reasons to resist any change in the conventional way of thinking, there are powerful reasons to press for change. The twin problems of arbitrariness and obfuscation, which are exemplified dramatically in cases such as *Chapman* and *Ledbetter*, are very grave indeed. The rule of law would be substantially strengthened, not weakened, if a way could be found to ameliorate these difficulties.

It would be useful to reexamine the core proposition stated above—that statutory interpretation is "a process of legal reasoning, grounded in a set of legally prescribed considerations." This proposition is a fair description of what statutory interpretation is and ought to be in most cases, but it overstates the power of conventional interpretive methods. In some cases, conventional methods do not support case-specific interpretation. They may establish boundaries for choice, but they do not always determine the choice itself. In such cases, the courts must rely on something other than legal reasoning in the conventional sense, and our theories of statutory interpretation should grant them the liberty to do precisely that. Indeed, there are precedents for a more flexible conception of the interpretive function in certain contexts, and it would be useful to take note of them before developing the argument further.

A. The "Portal-to-Portal" Case

During the early decades of the twentieth century there was a bitter dispute between iron miners and the owners of iron mines in the southeastern United States.[51] The dispute concerned the method by which the owners calculated the miners' wages. The miners began work each day

51. *See* Tenn. Coal, Iron & R.R. Co. v. Muscoda Local No. 123, 321 U.S. 590, 592–97 (1944).

by arriving at the mine, changing into working clothes, and collecting tools and equipment. They then proceeded to the "portal" of the mine, and from there they were transported underground to the "working face" of the mine, where they performed mining operations. At the end of the day they were transported back to the surface of the mine, where they stowed their tools, bathed, changed clothes, and returned to their homes. The mining companies followed the practice of paying the miners only for the work performed on the working face of the mine. They did not pay the miners for activities at the surface of the mine, and they did not pay them for the time they spent traveling within the mine to and from the working face. The miners, for their part, wanted to be paid for all of their activities and all of their time at the mine, and as a result, there was substantial unrest within the mining industry in the Southeast during the early decades of the twentieth century.[52]

The Fair Labor Standards Act (FLSA), enacted in 1938, provided that if an employee's "workweek" exceeded a certain maximum number of hours (forty-four, forty-two, or forty, depending on the circumstances), the employee was entitled to receive overtime pay at a rate of one and one-half times the employee's ordinary rate of compensation.[53] How did this requirement affect the wage-payment practices in the iron mines? The miners usually spent about eight hours a day on the working face of the mine. If, for purposes of the FLSA, their workweek included only the hours they spent on the working face, then they were entitled to receive no overtime compensation; but if their workweek included the time they spent in activities at the surface of the mine, or if it included the time they spent traveling within the mine to and from the working face, then the statutory maximum would be exceeded, and they would be entitled to receive overtime pay.

The Tennessee Coal, Iron, and Railroad Company and other mining companies brought a declaratory judgment action against certain mining unions to determine the effect of the FLSA on pay practices in the mining industry.[54] The companies argued that it was customary within the industry to measure the miners' workweek by the time spent on the working face of the mine, that this custom had guided wage negotiations in the industry for a number of years, and that Congress must have intended to affirm this custom when it enacted the FLSA.[55] The unions argued that for purposes of the FLSA the miners' workweek included all of the time the

52. *Id.* at 601–02.
53. Fair Labor Standards Act of 1938 § 7(a), 29 U.S.C. § 207(a)(1) (2006).
54. *Tenn. Coal, Iron & R.R. Co.*, 321 U.S. at 592.
55. *Id.* at 600–01.

miners spent at the mine, not just the time they spent on the working face.[56] The unions insisted that the miners were therefore entitled to overtime pay.

After a lengthy trial, the district court ruled in favor of the miners. Sitting without a jury, the district court found that the miners' travel time, as well as the time they spent at the surface of the mine obtaining and returning equipment, should be counted in determining the length of their workweek under the FLSA.[57] The companies appealed. The U.S. Court of Appeals for the Fifth Circuit partially affirmed and partially reversed the district court's judgment. The court of appeals agreed that travel time within the mine should be included in the calculation of the miners' workweek, along with the time the miners spent on the working face of the mine, but the court of appeals held that the time the miners spent obtaining and returning tools and equipment at the surface of the mine should be excluded from the calculation. In short, the court of appeals held that the miners' workweek should be calculated on a "portal-to-portal" basis.[58]

Thus, three different interpretations of the FLSA emerged over the course of the litigation. The unions claimed, and the district court found, that the miners' workweek included most of the time the miners spent at the surface of the mine and all of the time they spent within the mine. The court of appeals excluded the time the miners spent at the surface of the mine but included everything else. The mine owners took the most restrictive view: The workweek included only the time the miners spent on the working face of the mine. When the case reached the Supreme Court, the problem was to determine which one of these interpretations of the FLSA was correct.[59]

Seven of the nine Justices who heard the case took a conventional approach to the problem. Five of the Justices held that the portal-to-portal concept was consistent with the statutory language and the intentions and policies of Congress.[60] Two of them argued, to the contrary, that Congress must have intended for the FLSA to confirm the traditional pay practices in the iron mines, which credited the miners only for work done on the working face of the mine. These two Justices dissented.[61]

The two remaining Justices, Justices Frankfurter and Jackson, concurred

---

56. *Id.* at 592–93.

57. *Id.* at 593.

58. *Id.*

59. After the case reached the Supreme Court, the unions abandoned their claim that the "workweek" included time spent on the surface of the mine. Thus, the Court was called upon to decide between the "portal-to-portal" concept adopted by the court of appeals and the mine owners' contention that the "workweek" included only time spent on the "working face." *See id.* at 593 n.4.

60. *Id.* at 602.

61. *Id.* at 606–19 (Roberts, J., dissenting).

in the judgment, and they took an unconventional approach to the problem. Justice Frankfurter noted that the concept of a workweek was colloquial and that the term had no technical meaning.[62] He observed with apparent regret that Congress had created no administrative agency with authority to resolve interpretive issues arising under the FLSA.[63] This meant that the task of applying the imprecise language of the statute to "the multifarious situations in American industry" inevitably fell to the courts.[64] He then opined, remarkably, that even though the meaning of the word workweek had to be determined through "judicial proceedings," the question was *not* one of law; instead, it was one of fact.[65] The composition of the miners' workweek was to be determined in the lower courts as a matter of fact, and the findings of the lower courts as to the facts were not to be disturbed on appeal as long as they were supported by the evidence. The district court had conducted an extensive trial and had made careful findings with respect to the question of travel time. Those findings, which the circuit court had substantially approved, were supported by the evidence, and they were conclusive. Therefore, according to Justice Frankfurter, the judgment below must be affirmed.[66]

Justice Jackson agreed with Justice Frankfurter. He said that the case probably did not present "any question of law."[67] When Congress enacted the FLSA, it probably considered that "a workweek in fact should be a workweek in law"; therefore, any judicial determination of the issue was factual in nature and case-specific.[68] A decision in one case would not govern any other case, "for each establishment and industry stands on its own conditions."[69] Justice Jackson then noted that the district court had made extensive findings of fact that were supported by the evidence. He said that he would affirm the judgment below on the basis of these "controlling facts."[70]

In other words, Justice Frankfurter and Justice Jackson did not attempt to define the word *workweek* as a matter of law. In the opinion of these Justices, the usual considerations—the objective meaning of the statutory language, the legislature's intentions, policies, and purposes, and the traditional rules of interpretation—were apparently indeterminate in this case. This did not mean that the statute was an empty vessel. Frankfurter

62. *Id.* at 604 (Frankfurter, J., concurring).
63. *Id.*
64. *Id.*
65. *Id.*
66. *Id.*
67. *Id.* (Jackson, J., concurring).
68. *Id.* at 605.
69. *Id.*
70. *Id.* at 606.

and Jackson surely would have objected if the lower courts had found that the miners' workweek included leisure time spent at home. But they granted the lower courts considerable latitude within the limits established by the general concept of a workweek, and they agreed that the lower courts could resolve the issue on the basis of considerations that did not involve legal reasoning in the conventional sense.

B. Administrative Discretion in Statutory Interpretation

Justices Frankfurter and Jackson did not invent the idea that the conventional methods of statutory interpretation are sometimes indeterminate and that the precise meaning of a statute may sometimes depend on something other than legal reasoning. Indeed, there is an important field of law in which this idea, or something very much like it, has been accepted for a very long time. Within the field of administrative law, there are various doctrines that sometimes make statutory interpretation a matter of administrative discretion.

These doctrines were first introduced into American law as a critique of traditional interpretive methods. Consider, for example, an early trade regulation case, *FTC v. Gratz*.[71] During the 1910s the firm of Warren, Jones & Gratz (WJ&G) was in the business of selling various materials that were used in the marketing of cotton fiber, including steel "ties," which were used to bind cotton bales, and jute "bagging," which was used to wrap cotton bales. The Carnegie Steel Company manufactured the steel ties, and WJ&G was the Carnegie Steel Company's exclusive selling agent nationwide. WJ&G required its customers to purchase a prescribed quantity of jute bagging with every purchase of steel ties. In other words, if a customer wanted to buy steel ties, it had to buy a certain quantity of jute bagging as well.[72]

WJ&G led a quiet life until Congress passed the Federal Trade Commission Act (Act) in 1914.[73] The Act declared that "unfair methods of competition" were "unlawful," and it created a new administrative agency, the Federal Trade Commission (FTC), to enforce its provisions.[74] It authorized the FTC to issue complaints against persons who employed "unfair methods of competition in commerce,"[75] and in 1917 the FTC issued such a complaint against WJ&G and others, alleging that WJ&G had

71. 253 U.S. 421 (1920).

72. *Id.* at 428.

73. *Id.* at 422; *cf.* 15 U.S.C. §§ 41–45 (2006) (current version of the Federal Trade Commission Act).

74. *Gratz*, 253 U.S. at 422 (quoting the original Federal Trade Commission Act).

75. *Id.*

violated the Act by tying the sale of steel ties to the sale of jute bagging. The FTC conducted an administrative hearing, as the Act required. It made detailed findings concerning WJ&G's policy, and on the basis of these findings it issued a cease and desist order requiring WJ&G to abandon its policy.[76]

Seeking relief from the FTC's order, WJ&G petitioned the U.S. Court of Appeals for the Second Circuit. The Second Circuit annulled the order, and the FTC then appealed to the Supreme Court. The Supreme Court affirmed the Second Circuit's decision. It held that the FTC's complaint against WJ&G was insufficient on its face and that the cease and desist order therefore lacked a proper legal foundation.[77] Justice McReynolds, a notable conservative, delivered the opinion of the Court. Justice Brandeis, a formidable progressive, filed a lengthy dissent.

Justice McReynolds's majority opinion was entirely traditional in terms of interpretive methodology. After stating the case, he made three quick points. First, he said that it was necessary for the Court to determine the meaning of the statutory phrase "unfair methods of competition." In his view the Court, not the FTC, should have the last word concerning that issue.[78] Second, he said that the phrase *unfair methods of competition* referred to practices that were condemned by the common law; it did not refer to methods of competition "never heretofore regarded as opposed to good morals because characterized by deception, bad faith, fraud or oppression, or as against public policy because of their dangerous tendency unduly to hinder competition or create monopoly."[79] Third, he said that the FTC's complaint contained no allegation of deception, misrepresentation, or oppression, and no allegation of monopoly with respect to the sale of steel ties or jute bagging—in short, no allegation of anything that would constitute an "unfair method of competition" in the common law sense.[80] Thus, the complaint was flawed, the administrative procedure was defective, and the cease and desist order was invalid.

Embedded in this argument was the ancient assumption that courts should interpret statutory language in light of the common law. Justice McReynolds believed that the phrase *unfair methods of competition* had a specific common law meaning and that the FTC's jurisdiction therefore extended only to practices that constituted "unfair competition" in the common law sense. Furthermore, he thought that the Act should be interpreted to require the FTC to follow a procedure akin to that which

76. *Id.* at 429–30 (Brandeis, J., dissenting).
77. *Id.* at 429 (majority opinion).
78. *Id.* at 427.
79. *Id.*
80. *Id.* at 428.

was followed in the common law courts. At common law, according to traditional principles, it was necessary for a complaint to contain specific factual allegations showing that a cause of action existed. If a complaint simply gave notice of a claim without alleging facts sufficient to constitute a cause of action, it was legally defective. The FTC's complaint against WJ&G was insufficient when viewed in that light. It simply stated that the tying policy was "an unfair method of competition." It did not allege facts sufficient to establish a cause of action for monopoly, fraud, or oppression.

If Justice McReynolds's opinion was traditional in terms of its methodology, Justice Brandeis's dissenting opinion was a model of modernity. It resolved every interpretive issue by referring neither to the statutory text nor to the common law, but to the legislative history of the Act and to the decisions, interpretations, and customary practices of federal administrative agencies.[81] Justice Brandeis relied extensively on the reports of the relevant congressional committees, the publications of the Federal Bureau of Corporations, and the procedures of the Interstate Commerce Commission, and he made four main points. First, he said that Congress intended the Act to create a *novel* procedure that could address trade practices that were beyond the reach of traditional law.[82] Second, he said that the Act did not explicitly or implicitly require an FTC complaint to contain specific factual allegations such as those required in pleadings at common law. An administrative complaint under the Act would be sufficient, in his view, if it contained "a plain statement of the thing claimed to be wrong so that the respondent may be put upon his defence."[83] Third, he said that the Act did not define *unfair methods of competition* but left that matter to be determined by the FTC, *not* by the courts.[84] Finally, and most importantly, he said that if the FTC had decided in a formal trial-type hearing that a certain trade practice constituted an unfair method of competition, the role of the courts was simply to determine whether the FTC's decision was reasonable in light of the FTC's findings of fact. It was not for the courts to decide the matter anew through a process of legal reasoning.[85] The FTC had discretion to determine the effective legal meaning of the Act within appropriate legal bounds.

Justice Brandeis's argument was essentially an argument about the delegation of legislative or quasi-legislative power to the FTC. The operative statutory language—"unfair methods of competition"—had a common law meaning, to be sure, but Justice Brandeis believed that

81. *Id.* at 431–41 (Brandeis, J., dissenting).
82. *Id.* at 432.
83. *Id.* at 430.
84. *Id.* at 436.
85. *Id.* at 437–38.

Congress intended to give the FTC power to go beyond the common law to deal with problems that the common law did not adequately address. According to Justice Brandeis, Justice McReynolds's opinion was in error because it mistook the legislative purpose. The Act was designed to promote reform, and the job of defining *unfair methods of competition* therefore fell to the new administrative agency, not to the courts. The courts' role was simply to review the FTC's decisions under the Act to determine whether they were generally within the scope of the statutory grant and supported by the administrative record. The effective meaning of the statute was to be determined by the FTC in the exercise of sound administrative discretion.

Of course, Justice Brandeis lost the argument in *Gratz* and the conservatives won the day, but Brandeis's dissent was a harbinger of things to come. Within a generation or two, as the administrative state grew and matured, Brandeis's way of thinking about the relationship between the federal courts and the federal administrative agencies began to dominate the new field of administrative law. During the middle decades of the twentieth century, the federal courts began to accept the idea that they should defer to administrative interpretation of statutory law in certain circumstances. Moreover, they discovered that there were justifications for judicial deference that did not necessarily involve arguments about delegated legislative power. For example, if it was clear that Congress had considered an agency's interpretation of a statute and had implicitly ratified it through subsequent legislative action or inaction, judicial deference was sometimes appropriate;[86] and if a question of interpretation involved a technical matter within the agency's field of expertise, judicial deference was appropriate as a matter of simple prudence.[87]

Two decades ago, in *Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc.*,[88] the Supreme Court considered and restated the general principles governing judicial deference to administrative interpretation. *Chevron* involved the Clean Air Act Amendments of 1977, which required various states to establish programs to regulate "new and modified major stationary sources" of air pollution.[89] During the Administration of President Reagan, the Environmental Protection Agency (EPA) promulgated regulations

86. *See, e.g.*, Bob Jones Univ. v. United States, 461 U.S. 574, 575 (1983) (stating that subsequent congressional action and inaction implicitly ratified IRS's interpretation of the concept of "charity" under the Internal Revenue Code).

87. *See, e.g.*, NLRB v. Hearst Publ'ns, Inc., 322 U.S. 111 (1944) (in light of NLRB's expertise, courts should defer to NLRB's interpretation of word *employee* in National Labor Relations Act). *See generally* Skidmore v. Swift & Co., 323 U.S. 134 (1944).

88. 467 U.S. 837 (1984).

89. 42 U.S.C. § 7502(c)(5) (2006).

implementing this part of the statute.[90] These regulations allowed the states to treat multiple pollution-emitting devices (e.g., multiple smokestacks) as a single "stationary source" if they were located within the same plant or installation, and this had the effect of making it easier for polluting companies to comply with the relevant regulatory standards.[91] The National Resources Defense Council challenged the EPA's interpretation of the phrase *stationary source*, and when the case came before the Court, the question was whether the Court was obliged to accept the agency's interpretation of the statute or whether it was free to adopt an interpretation of its own.[92]

Writing for a unanimous Court, Justice Stevens held that judicial deference to the EPA's interpretation was warranted, and in the course of his opinion he proposed a general theory of judicial deference to administrative interpretation which drew on traditional doctrines, even as it appeared to plow new ground. Justice Stevens reasoned that when a litigant questions an agency's interpretation of the statute the agency administers, two issues may arise. The first is "whether Congress has directly spoken to the precise question at issue."[93] If the reviewing court finds that "Congress *has* directly spoken to the precise question at issue" (if the court finds that the legislative intent with respect to that issue is "clear"), then the inquiry comes to an end, "for the court, as well as the agency, must give effect to the unambiguously expressed intent of Congress."[94] The second issue arises if the court determines that Congress has *not* "directly spoken to the precise question at issue." In that event, according to Justice Stevens, "the court does not simply impose its own construction on the statute, as would be necessary in the absence of an administrative interpretation."[95] Instead, the court asks whether the agency's interpretation is "permissible."[96] To affirm the agency, the court need not conclude that the agency's interpretation is the only permissible one or that the court itself would have adopted the agency's interpretation "if the question initially had arisen in a judicial proceeding."[97] Rather, the court must defer to the agency's interpretation if the court finds the interpretation to be reasonable.[98]

Justice Stevens explained that judicial deference to the agency's

90. *Chevron*, 467 U.S. at 840.
91. *Id.*
92. *Id.* at 842–43.
93. *Id.* at 842.
94. *Id.* at 842–43 (emphasis added).
95. *Id.* at 843 (citation omitted).
96. *Id.*
97. *Id.* at 843 n.11.
98. *Id.* at 844–45.

interpretation is appropriate in such a case because the court is being asked to decide a question that the law itself does not resolve. Because Congress has *not* directly spoken to the precise question at issue, the question is essentially one of policy, according to Justice Stevens, and questions of policy are best resolved, not by judges, but by political officers in the Executive Branch.[99]

It is easy to quibble with Justice Stevens's opinion. His analysis of the general issue of judicial deference is reductive and stark.[100] Yet his central contention is clearly correct. Like Justice Frankfurter and Justice Jackson in the portal-to-portal case, and like Justice Brandeis in *Gratz*, Justice Stevens recognized that legislative action sometimes creates not a specific rule but a framework for choice, and he was right to insist that such choices must sometimes be made on grounds that are not legally prescribed. In such cases, when an administrative agency has resolved the matter in a reasonable way, the courts are wise to defer to the agency determination.

But how should the courts proceed if there is *no administrative interpretation* of the statute to which they can defer?

IV. A PROPOSAL

Consider the hypothetical posed at the beginning of this Article. A court is presented with a difficult question of statutory interpretation. There are two plausible interpretations of the statute, *A* and *B*. Conventional interpretive considerations are in equipoise, and the court cannot persuade itself that either interpretation is preferable to the other on conventional grounds. Assume further that there is no reasonable administrative interpretation of the statute to which the court can defer, as in *Chevron*, and that the question cannot properly be treated as a one of "fact," to be decided by the finder of fact on a case-by-case basis, as Justices Frankfurter and Jackson proposed in the portal-to-portal case. How should the court resolve the issue? Which interpretation of the statute should the court prefer, *A* or *B*?

Only two courses of action are open to the court in such a case, given the

99. *Id.* at 865–66.

100. His distinction between two kinds of cases—those in which Congress has "directly spoken to the precise question at issue," and those in which Congress has *not* "directly spoken to the precise question at issue"—is too simplistic. Even when Congress has not "directly spoken to the precise question at issue," it is sometimes possible to find persuasive legal reasons to interpret a statute one way or another. The absence of evidence of specific legislative intent does not inevitably convert a question of interpretation into a question of "policy," as Justice Stevens seems to suggest. Persuasive *legal* interpretation can still occur. That, at any rate, has been the traditional assumption, and it is a valid assumption, in my view.

constraints of conventional theory. If there are unconventional reasons for preferring one interpretation of the statute over the other, the court may elect to decide the case on that basis, while constructing a conventional defense of the decision and sacrificing candor in the process. (*Ledbetter* provides a clear example of this approach.) Alternatively, the court may prefer to make a quasi-arbitrary choice between *A* and *B*, preferring one interpretation over the other for conventional reasons that are thin and unconvincing. (This is surely what happened in *Chapman*.)

Yet if our theories of statutory interpretation were adjusted, the court could approach the case in a more sensible way. The court could declare that there are two plausible interpretations of the statute, *A* and *B*, and that the court finds no persuasive conventional reason to prefer *A* over *B* or *B* over *A*. The court could then hold that the statute creates a framework for judicial choice among legally permissible alternatives, in much the same way that the statute in *Chevron* created a framework for administrative choice among legally permissible alternatives, and the court could make a prudent discretionary choice between the permissible alternatives. The court would not be required to pretend that conventional considerations were determinative. The court would be required, however, to disclose fully the reasons for its choice, whatever they might be.

To understand how this approach would work in an actual case, consider both *Ledbetter* and *Chapman*. In *Ledbetter*, the plaintiff complained about an allegedly unlawful pay practice that began with discriminatory salary-setting decisions and continued for a period of years thereafter. The relevant statute required the plaintiff to file her complaint within 180 days after the alleged unlawful practice "occurred." The question was whether the 180-day filing period ran from the time the unlawful practice began to occur or from the time it ceased to occur. A conscientious judge might have concluded that there were two plausible interpretations of the statute. One would have required the plaintiff to file her complaint within 180 days after the unlawful practice began; the other would have required her to file the complaint within 180 days after the alleged unlawful practice ended. If this Article's proposal were accepted, a judge who believed that these two interpretations were equally plausible on conventional grounds would be allowed to declare her opinion openly, and would be permitted to choose between the alternative interpretations for any prudent unconventional reason while making full disclosure in the process. Such a judge might reason, for example, that a lenient interpretation of the statute would expose the employer unfairly to liability for stale claims. In that event, the judge might adopt the strict interpretation of the statute. On the other hand, the judge might adopt the lenient interpretation in the belief that a strict interpretation would be impractical and unfair to the employee, given

the realities of the workplace and the difficulty of detecting discriminatory decisionmaking by employers. In no event, however, would the judge be obligated to pretend that the final decision resulted from legal reasoning, that is to say, from conventional interpretive considerations (from an assessment of the legislature's actual intentions, policies, or purposes, from an assessment of the objective meaning of the text, from an assessment of precedent, or from an application of settled rules of interpretation), and the judge would not be obligated to pretend that the preferred interpretation was the only plausible or permissible one. The judge's only obligation would be to explain candidly why the case was ultimately decided as it was, in the exercise of interpretive discretion.

In *Chapman*, the question was whether blotter paper stained with LSD was a "mixture" for purposes of a statute that made a defendant's sentence a function of the weight of the illicit material he had distributed. A conscientious judge might well have concluded that there were two plausible interpretations of this statute. One interpretation would have treated blotter paper sprinkled with LSD as a mixture; the other would not have treated it as a mixture. If this Article's proposal were accepted, a judge who believed that these two interpretations were equally plausible would be allowed to say so, and the judge would be allowed to choose between them for any prudent reason. For example, the judge might choose not to treat blotter paper sprinkled with LSD as a mixture because the resulting sentences under the statute would then be roughly in line with the sentences imposed on distributors of other similar drugs, e.g., heroin and cocaine.[101] Alternatively, the judge might treat blotter paper sprinkled with LSD as a mixture on the ground that this treatment would tend to establish rough equality between defendants who used blotter paper as a carrier medium for LSD and defendants who used other carrier media for LSD. In no event would the judge be obliged to pretend that the preferred interpretation turned on considerations that were legally prescribed. Instead, the judge would be allowed to make a prudent, discretionary choice between the conventionally plausible alternatives and would be required to disclose fully the reasons for his choice.

A number of important benefits would flow from a doctrine that explicitly recognized the discretionary character of statutory interpretation in cases such as these. A doctrine of discretionary interpretation would encourage both rigor and clarity in judicial decisionmaking because it would require the courts to identify and explain the plausible alternative interpretations of the statute in question and the actual reasons for

101. *See* United States v. Marshall, 908 F.2d 1312, 1335 (7th Cir. 1990) (en banc) (Posner, J., dissenting).

preferring one interpretation over others. It would discourage obfuscation because it would not require the courts to construct a conventional defense of an interpretive decision that did not actually turn on conventional calculations. Finally, and perhaps most importantly, it would discourage arbitrariness, because it would relieve the courts of the obligation to attribute decisive legal force to conventional considerations that were trivial and unpersuasive in the circumstances.

The philosophical adjustment advocated in this Article is both modest and broad. The argument is not that courts should be given "discretion" to revise legislation to meet the needs of changing times, as some scholars have suggested,[102] and it is not that courts should be given a roving commission to exercise discretion in statutory interpretation in any and every case for the purpose of doing justice. Courts do not, and should not, possess that kind of power. Instead, the argument is that courts must have power to resolve uncertainty concerning the meaning of statutory texts, *and that they must have this power even in cases in which they find the conventional determinants of statutory meaning—legislative intent, objective meaning, and the like—to be indeterminate*. In such cases, for the reasons given above, the conventional theories of statutory interpretation should be adjusted to recognize that statutory interpretation sometimes requires the exercise of judicial discretion. The specific interpretive choices made in such cases cannot turn upon conventional considerations; instead, they must turn upon the court's understanding of what justice, equity, practicality, or sound public policy require in the circumstances, within the framework established by the legislative act.

Finally, it should be clear that when this Article refers to discretionary interpretation, it is referring to a power that is vested in the Judicial Branch as a whole. Like any power to interpret law, it must be exercised ultimately by the highest court in the relevant jurisdiction. Discretionary interpretive judgments made in the lower courts would be subject, under this proposal, to de novo review on appeal, just as conventional interpretive judgments are. Only in cases in which the question was essentially one of fact would the lower court's judgment be entitled to deference.[103]

V. SOME FINAL REFLECTIONS ON FUNDAMENTALS

It has never been generally recognized in theory that statutory interpretation involves discretion, as that term is used here. Moreover, the conventional theories of statutory interpretation are sensible enough, as far as they go. Surely a statute's legal meaning must depend upon the intended

102. *See, e.g.*, GUIDO CALABRESI, A COMMON LAW FOR THE AGE OF STATUTES (1982).
103. *See supra* text accompanying notes 62–70.

meaning of the text, the objective meaning of the text, reasonable inferences concerning legislative policies or purposes, existing rules of law and interpretation, or some combination of these things. And in any case, there is something disturbing about a claim that statutory interpretation may sometimes involve free judicial choice. Legislative action is supposed to create law by giving it definite form, and the courts are supposed to give legal effect to that form through the process of interpretation. The idea that a statute may have a meaning that the courts are free to create in their discretion is inconsistent with this general understanding. It suggests a kind of formlessness contrary to the general nature of legislation, and it suggests a judicial function in relation to legislation quite different from the one the courts ordinarily perform. It suggests that the courts have a lawmaking power akin to their power to create retrospective common law rules, yet it claims that this power exists in connection with the legislative process itself, which is supposed to be generally prospective in operation, and which legislators, not judges, are supposed to control.

The strength of these objections must be acknowledged, yet there are powerful countervailing arguments. Some of the benefits that would flow from explicit recognition of discretionary interpretation have already been described. In the paragraphs below, this Article will briefly describe some constitutional considerations that should encourage the courts to embrace these benefits wholeheartedly.

A. Interpretive Discretion and the Constitutional Origins of Statutory Uncertainty

Consider the separation of legislative and judicial institutions. The framers of our state and federal constitutions were not interested in promoting communication between the legislative and judicial branches, and they did not make it easy for the legislature to anticipate and control judicial behavior in particular cases. They intended to create division. In the federal system, the Constitution gives lawmaking authority to a bicameral assembly that has no single intellect, no single set of intentions. The Constitution requires this assembly to act according to a cumbersome bill-making procedure that is designed to produce a fixed form of words for the approval of 535 different congressional minds, plus the approval of the President; and it assigns to a separate branch of government, the Judicial Branch, the responsibility for interpreting and enforcing these words in particular cases. It does not allow legislators to hold judicial office.[104] It gives the courts no power to interrogate the legislature concerning its intentions, policies, or purposes or to remand statutory questions to the

104. U.S. CONST. art. I, § 6, cl. 2.

legislature for resolution. It imposes general limits on the legislature's power to deal retroactively and prospectively with specific cases and controversies,[105] and it creates no administrative mechanism by which the legislature can control judicial decisionmaking before or after the fact. Nor does it create any intermediary institution with power to reconcile and coordinate legislative and judicial decisions.[106] In short, the Constitution does nothing to ensure that the two departments will work together.

Given these constitutional structures, practices, and principles, it would be astonishing if Congress could communicate its will to the courts so precisely and effectively that the courts would always be in a position to resolve interpretive issues persuasively by appealing to demonstrable legislative intentions, policies, and purposes. It would be equally astonishing if Congress could produce statutory texts so precise and clear in their application to particular cases that the courts would always be in a position to resolve interpretive issues persuasively by appealing to the objective meaning of those texts. The Constitution itself prevents this from happening. The normal operation of the constitutional system guarantees that the intended meaning and the objective meaning of statutory texts, and the adjudicatory implications of legislative policies and purposes, will ultimately be uncertain in relation to specific statutory issues in some cases some of the time.

In light of the separation of legislative and judicial institutions, and in light of the inevitable consequences of that separation, it is remarkable that, except in the field of administrative law, the conventional theories of statutory interpretation make no general provision for cases in which the meaning of a statute is ultimately uncertain in relation to the specific issue to be decided. This omission does not necessarily make the conventional theories unconstitutional, but it surely makes them constitutionally awkward. In very difficult cases, because of the conventional theories, the courts must pretend that they are able to resolve statutory issues on the basis of legislative will or the objective meaning of the text, or on the basis of other conventional considerations, even when they are in no position to do so. The courts must resort to a kind of fiction, yet this fiction threatens justice in some cases and results in outright judicial subterfuge in others. It would be far better to recognize the problematic situation that the Constitution itself creates. On the one hand, it assigns to the courts the task of resolving uncertainty concerning the meaning of statutory texts; on the

105. *See, e.g.*, U.S. CONST. art. I, § 9, cl. 3 (Bill of Attainder and Ex Post Facto Clause); U.S. CONST. amend. V (Due Process Clause).

106. *See* JOHN HENRY MERRYMAN, THE CIVIL LAW TRADITION 39–42 (2d ed. 1985) (discussing attempts in civil law countries to create intermediary institutions for the interpretation of statutes).

other hand, it establishes a structure that sometimes makes it impossible for the courts to perform this task consistently according to the theories of interpretation upon which they traditionally rely. Because of the separation of powers, the conventional theories of statutory interpretation are not, and cannot be, analytically sufficient all of the time.

Our theories of statutory interpretation would be more consistent with the constitutional structure if they recognized the discretionary character of statutory interpretation in very hard cases. One is tempted to say that the constitutional structure delegates to the Judicial Branch the power to exercise discretionary interpretive judgment in such cases, but it would be better to use other words to describe the constitutional situation. The power to exercise discretionary interpretive judgment in very hard cases is inherently and necessarily a judicial power. It does not involve a delegation from the legislature. The Constitution necessarily assigns this power to the Judiciary, and our theories of statutory interpretation should recognize this power explicitly.

B. Interpretive Discretion and Five Judicial Virtues

Thus far, the argument for interpretive discretion has been largely negative in character. In some cases statutory interpretation should be regarded as discretionary because it cannot be otherwise and because bad things happen when courts pretend that it can be otherwise. There are positive reasons, however, for recognizing the discretionary character of statutory interpretation in some cases. When the courts, in their discretion, resolve uncertainty concerning the meaning of statutory texts, they are in some respects in a better position than the legislature itself to make judgments about what the law ought to be. They have institutional strengths that can and should be brought to bear to improve statutory law when the usual determinants of statutory meaning leave the legal consequences of legislative action in doubt. Five of these judicial virtues are discussed below.

1. Hindsight

The legislature generally makes policy for the future. It attempts to anticipate how law will affect future events, and it acts on the basis of prediction. Yet legislative foresight is imperfect, and future events can frustrate even the best laid legislative plans. The legislature's inability to foresee the future accurately is one of the causes of uncertainty in statutory law. If the legislature fails to anticipate the cases to which its statutes may apply, there is a possibility that its intentions, purposes, and language will fail to address those cases clearly. The courts, by contrast, enjoy the luxury

of hindsight. They decide cases involving historical facts that are subject to proof, and they can adjust their decisions to produce specific effects in known circumstances. In this respect, they are well positioned to compensate for deficiencies in legislative foresight. Using hindsight and prudent judgment, they can determine the legal consequences of legislative action when the legislature, looking forward, has failed to make those consequences clear.

2. Particularity

If legislative action tends to be forward looking, it also tends to be categorical. Indeed, there are constitutional principles and political realities that tend to discourage case-specific legislation.[107] If the failure of legislative foresight is one cause of uncertainty in statutory law, the categorical nature of legislative action is another. Legislative pronouncements that are models of clarity in their assertion of general rules and policies are sometimes remarkably uncertain in their bearing on particular cases and issues.[108] Here again, the courts are well positioned to compensate for an intrinsic legislative deficiency. The nature of the judicial function is such that the courts are obliged to come to terms with the particular. Even when categorical legislative action is uncertain in its relation to a particular case, the courts can and must say what the statute means in that case. The prudent resolution of uncertainty with respect to the particular case is a necessary and desirable aspect of the judicial office.

3. Detachment

When judges interpret statutes, they deal with the work of another branch of government. They usually have no institutional or personal investment in the words that appear on the pages of the statute book, no inside knowledge of the politics that supported the legislative action, and no professional commitment to the success or failure of the legislative act, aside from their general commitment to the rule of law. Judicial detachment is constitutionally valuable because, in cases of substantial uncertainty concerning either the objective or the intended meaning of a statute, it allows the courts to exercise independent judgment on the question of what the law ought to be. Sound public policy emerges over time through a process of interdepartmental dialogue, refinement, and correction, and this

107. *See* U.S. CONST. amend. V (Due Process Clause); U.S. CONST. amend. XIV, § 1 (Due Process and Equal Protection Clauses as they apply to the individual states).

108. *See, e.g.*, 15 U.S.C. § 1 (2006) ("Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal.").

requires that the legislature and the courts maintain independent points of view. The Framers built upon this principle when they separated the legislature and the courts in the first instance. Judicial independence in statutory interpretation contributes to the health of the democratic system. The point here is simply that law benefits in the long run from multiple decisionmakers in dialogue with one another.

4. Rational Explanation

Anglo-American lawyers have long believed that law should express practical reason. Blackstone asserted that reason was the essence of law: "[W]hat is not reason is not law."[109] Arbitrary action, or action which the government cannot adequately explain, has long raised suspicion in the Anglo-American legal mind. A penchant for rational explanation was characteristic of Anglo-American legal institutions at a very early date, and the process of explaining the reasons for governmental action is now an ingredient of legitimacy itself.

The legislature often attempts to explain its decisions in a general way; yet legislative decisions need to be explained precisely at the point at which they impinge on particular cases, and as a practical matter the legislature is in no position to do this comprehensively. The explanatory function belongs primarily to the courts. Statutory interpretation is essentially a process of explanation—a process by which judges explain the connection between legislative action and judicial or administrative decisions in particular cases. Through the process of interpretation—through the process of explaining the effect of legislation on particular cases—judges add value to statutory law by supplying an ingredient of legitimacy which the legislature itself is usually in no position to supply.

As noted above, there are many cases in which the modern theories of statutory interpretation actually interfere with the process of explanation. It is sometimes difficult to provide a convincing or even a minimally rational account of the legal effect of a statute in a particular case if the explanation must depend on bald assertions about legislative intent, legislative purpose, or the objective meaning of the statutory text. Yet a well-constructed theory of statutory interpretation would facilitate explanation. It would grant the courts the freedom to acknowledge the occasional indeterminacy of the elementary determinants of statutory meaning, and it would allow the courts to give other reasons for their interpretive decisions when the conventional reasons did not suffice. In this way arbitrariness would be defeated and legitimacy enhanced.

109. 1 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 70 (1765).

5. Consistency and Justice

The judicial process is designed to promote consistency in decisionmaking. The elementary notion that like cases should be treated alike is the bedrock principle, and the courts are organized in such a way that judges are supported when they adhere to this principle and corrected when they do not. To be sure, the judicial process sometimes operates inconsistently, as most human systems do, but consistency is the dominant ideal, and within the Judicial Branch there are various internal checks that tend to ensure that a measure of consistency is actually achieved. Consistency nurtures justice. It allows the governed to develop rational expectations about governmental action, it discourages arbitrary decisionmaking, and it promotes equal protection under the law. To be sure, consistency can also be the enemy of justice, especially when it discourages necessary innovation; but consistency is so closely linked with the limitation of arbitrary power, the protection of rational expectations, and the promotion of equality that the courts should not be faulted for their devotion to it.

Here again, the contrast between the legislative process and the judicial process is striking. Within the Legislative Branch there are few, if any, internal controls that promote consistency in legislative decisionmaking. Subject only to external constitutional constraints, legislative decisionmaking can be as irregular as the wind. It can, and often does, reflect political considerations that are irrelevant to the merits of the questions under consideration, and it can be haphazard and piecemeal. Judicial decisionmaking can sometimes exhibit similar qualities, but in the Judicial Branch there are traditions, procedures, and principles that guard against this.

The great virtue of the legislative process is its commitment to *inconsistency* over time, i.e., innovation. In an open society, justice can be achieved through innovation, and it is good that one branch of government is involved in the business of engineering change. But the judicial commitment to consistency in the pursuit of justice can complement the legislative commitment to innovation in the pursuit of justice, and this can occur precisely at the point where the judicial mind engages the legislative text. When the meaning of a statute is substantially in doubt, and when judges must determine the connection between the statute and a particular case, the judicial commitment to consistency can improve the legal consequences of legislatively mandated change.

CONCLUSION

In most cases it is realistic to suppose that statutes have a fairly precise,

judicially determinable meaning and that the courts are in a position to discover and declare that meaning through a process of legal reasoning. Yet there are cases in which this concept is simply unworkable. Not every question of statutory interpretation can be resolved persuasively through a process of legal reasoning, and to pretend otherwise is to perpetuate a fiction. Some legal fictions are useful; this one is not. It degrades, confuses, and corrupts the interpretive process.

Conventional interpretation is likely to fail in cases such as the ones described in this Article. Sometimes a conscientious judge will examine the relevant statute and conclude that more than one interpretation is possible and that conventional considerations do not favor a particular interpretive choice. In such a case, the judge should be free to declare that the statute creates a framework for choice, and the judge should be allowed to exercise prudent judgment within the statutory framework by preferring one permissible interpretation or another, while explaining candidly the reasons for the preference. In such a case the interpretive process is essentially a discretionary process, and it should be theorized as such.

* * *

COERCED PARTICIPATION IN CLINICAL TRIALS: CONSCRIPTING HUMAN RESEARCH SUBJECTS

LARS NOAH*

TABLE OF CONTENTS

INTRODUCTION

Biomedical research in this country depends heavily on an army of volunteers—persons willing, for one reason or another, to serve as experimental subjects. The supply of participants has not, however, kept pace with the growing demand.[1] Private research sponsors have found

* Professor of Law, University of Florida. An earlier version of this Article was presented at the annual meeting of the Law & Society Association in Denver, Colorado, and at the University of Toronto as part of its health law, ethics, and policy seminar series.

1. *See* Nancy S. Sung et al., *Central Challenges Facing the National Clinical Research Enterprise*, 289 JAMA 1278, 1279–80 (2003); Naomi Aoki, *Lack of Human Test Subjects a Bitter Pill for Drug Makers*, BOSTON GLOBE, Oct. 11, 2000, at D4; Thomas Ginsberg, *Please Stop Calling Them Drug 'Trials*,' PHILA. INQUIRER, June 21, 2006, at C1; Linda Marsa, *Clinical Trials Are Suffering: Suspicious of Medical Research, Volunteers Spurn Tests of Possibly Lifesaving Advances*, L.A. TIMES, Dec. 2, 2002, at F1; Virginia A. Smith, *Medical Testing Suffers from a Lack of Volunteers*, PHILA. INQUIRER, Apr. 6, 2004, at A1.

clever ways of recruiting new volunteers,[2] including controversial efforts at outsourcing the burden.[3] Governmental sponsors of research generally have eschewed such market-oriented solutions,[4] instead finding creative ways of encouraging otherwise unwilling subjects to "volunteer" for service. This Article describes and critiques the latter approach, focusing on a method recently developed to enroll Medicare beneficiaries in clinical trials.

This Article also provides an opportunity to evaluate official pronouncements on bioethics. This issue became prominent during the previous administration, with sometimes scathing and largely justified criticism of the President's Commission on Bioethics (PCB),[5] but potentially more radical pronouncements came from lower-level officials within the U.S. Department of Health and Human Services (HHS), particularly at the National Institutes of Health (NIH). Rather than the more easily dismissed tracts about hot-button social issues produced by ideologues serving on the PCB, these remarks about more technical issues appeared in the form of scholarly publications by respected academics doing stints in public service. For the most part, their articles did not attract the media's attention, but this very lack of visibility—coupled with a surprising absence of any sustained response by academics unaffiliated with the federal government—makes their provocative arguments potentially more insidious; surely not on a par with the infamous torture memos produced at the Department of Justice,[6] though with far broader potential impact.

---

2. *See* Trudo Lemmens & Paul B. Miller, *The Human Subjects Trade: Ethical and Legal Issues Surrounding Recruitment Incentives*, 31 J.L. MED. & ETHICS 398 (2003); Rachel Zimmerman, *Desperately Seeking Kids for Clinical Trials*, WALL ST. J., May 29, 2002, at D1.

3. *See* Marc Kaufman, *Clinical Trials of Drugs Fewer, Study Says*, WASH. POST, May 4, 2005, at A2; Saritha Rai, *Drug Companies Cut Costs with Foreign Clinical Trials*, N.Y. TIMES, Feb. 24, 2005, at C4.

4. For an exception, see Ariana Eunjung Cha, *AIDS Vaccine Testing Goes Overseas: U.S. Funds $120 Million Trial Despite Misgivings of Some Researchers*, WASH. POST, May 22, 2006, at A1.

5. *See* Lars Noah, *A Postmodernist Take on the Human Embryo Research Debate*, 36 CONN. L. REV. 1133, 1148–52 (2004); Rick Weiss, *Conservatives Draft a 'Bioethics Agenda' for President*, WASH. POST, Mar. 8, 2005, at A6.

6. *See* Ross L. Weiner, Note, *The Office of Legal Counsel and Torture: The Law as Both a Sword and Shield*, 77 GEO. WASH. L. REV. 524, 526, 536–49 (2009); Charlie Savage & Scott Shane, *Terror-War Fallout Lingers over Bush Lawyers*, N.Y. TIMES, Mar. 9, 2009, at A1. Not surprisingly, when persons with legal training move into policymaking roles, they construe constraints on government action more loosely. *See* Lars Noah, *Interpreting Agency Enabling Acts: Misplaced Metaphors in Administrative Law*, 41 WM. & MARY L. REV. 1463, 1465 (2000); Lars Noah, *The Little Agency That Could (Act with Indifference to Constitutional and Statutory Strictures)*, 93 CORNELL L. REV. 901, 917–19, 922 (2008).

I. LEVERAGING OPPORTUNITIES FOR KNOWLEDGE ACQUISITION

This Part describes three different ways that federal agencies have encouraged participation in biomedical research. It starts with a controversial Medicare initiative, which conditioned coverage for certain novel interventions on an agreement by beneficiaries to enroll in clinical trials, and then it compares and contrasts approaches used by the Pentagon and the Food and Drug Administration (FDA). The Medicare policy represents a variant of a practice that I previously had characterized as agency "arm-twisting" and critiqued primarily on *ultra vires* grounds.[7] Part II of this Article instead objects to the Medicare policy on ethical grounds.

A. Seniors

In 1996, the Health Care Financing Administration (HCFA), now the Centers for Medicare and Medicaid Services (CMS), began a bold experiment, both literally and figuratively. It conditioned payment for a new medical procedure on agreements by both beneficiaries to enroll in a randomized controlled trial (RCT) and participating surgeons not to offer the procedure to anyone who had not enrolled.[8] One decade later, after a few similar decisions,[9] CMS formalized this approach by issuing a guidance

7. *See* Lars Noah, *Administrative Arm-Twisting in the Shadow of Congressional Delegations of Authority*, 1997 WIS. L. REV. 873, 874 (explaining that arm-twisting "refers to a threat by an agency to . . . withhold a benefit in hopes of encouraging 'voluntary' compliance with a request that the agency could not impose directly"); *id.* at 899 ("Federal regulators are hardly alone in using leverage to extract voluntary commitments or concessions from private parties that they could not impose directly."); *see also Novel Procedures in FCC License Transfer Proceedings: Hearing Before the Subcomm. on Commercial & Administrative Law of the H. Comm. on the Judiciary*, 106th Cong. 29 (2000) (statement of Professor Lars Noah, University of Florida College of Law). In that article, I had focused on regulated entities (primarily companies); agency arm-twisting of beneficiaries (primarily individuals) obviously could pose more serious concerns. *Cf.* Noah, *supra*, at 903–08, 918–23 (finding parallels in criminal plea bargaining).

8. *See* Jeffrey M. Drazen, Editorial, *Surgery for Emphysema—Not for Everyone*, 345 NEW ENG. J. MED. 1126, 1127 (2001) ("One key factor set this trial apart: Medicare would no longer pay for the operation if it was performed outside the trial. Thus, prospective patients who were also Medicare recipients had only two choices if they wanted lung-volume–reduction surgery: participate in the trial or pay for the operation themselves."); *id.* (applauding this "unique" and "creative solution").

9. *See* Gina Kolata, *Medicare Covering New Treatments, but with a Catch*, N.Y. TIMES, Nov. 5, 2004, at A1 (discussing studies ordered by the Centers for Medicare and Medicaid Services (CMS) studies into off-label uses of new cancer drugs and positron emission tomography (PET) scans to diagnose Alzheimer's disease as well as proposed patient registries for implanted defibrillators, carotid stenting, and bariatric surgery); *id.* (adding that "Medicare does not intend to force studies of everything it pays for"); Rick Weiss, *Medicare to Cover Cardiac Device: Plan Raises Issue of Line Between Care and Research*, WASH. POST, Jan. 20, 2005, at A1 (explaining that the decision to cover implantable cardioverter-defibrillators (ICDs) for beneficiaries with congestive heart failure so long as they enroll in a patient

for "coverage with evidence development" (CED).[10] The CMS guidance described two forms of CED: coverage with appropriateness determination (CAD) and coverage with study participation (CSP).[11]

The 1996 study that inaugurated the CSP approach had much to commend it. Lung-volume reduction surgery for emphysema patients had become popular in the early 1990s without having undergone any rigorous study.[12] Indeed, five years later, the investigators published early results demonstrating that those in the sickest subgroup who underwent the surgery experienced greater mortality.[13] Since then, CMS has used its CED policy on only a handful of occasions,[14] including a 2005 decision to

registry "represents the most aggressive effort yet to use the federal insurance plan for the elderly as a backdoor way to learn more about what works and what does not in medicine").

10. *See* CMS, NATIONAL COVERAGE DETERMINATIONS WITH DATA COLLECTION AS A CONDITION OF COVERAGE: COVERAGE WITH EVIDENCE DEVELOPMENT (July 12, 2006), http://www.cms.hhs.gov/mcd/ncpc_view_document.asp?id=8.

11. *See id.* at pt. V. The final guidance differed in important respects from a draft version issued one year earlier, which had failed to differentiate between these two study approaches. *See* CMS, FACTORS CMS CONSIDERS IN MAKING A DETERMINATION OF COVERAGE WITH EVIDENCE DEVELOPMENT (Apr. 7, 2005), http://www.cms.hhs.gov/coverage/download/guidanceced.pdf; Sean R. Tunis & Steven D. Pearson, *Coverage Options for Promising Technologies: Medicare's 'Coverage with Evidence Development,'* 25 HEALTH AFF. 1218, 1225–26 (2006). At around the same time, CMS officials penned a brief defense of a coverage with evidence development (CED) requirement for defibrillators. *See* Mark B. McClellan & Sean R. Tunis, *Medicare Coverage of ICDs*, 352 NEW ENG. J. MED. 222, 223 (2005) (focusing on the advantages of patient registries). Other countries have adopted similar policies. *See* John Hutton et al., *Coverage with Evidence Development: An Examination of Conceptual and Policy Issues*, 23 INT'L J. TECH. ASSESSMENT IN HEALTH CARE 425, 426 (2007).

12. *See* Mark R. Tonelli et al., *Clinical Experimentation: Lessons from Lung Volume Reduction Surgery*, 110 CHEST 230, 230–32 (1996); Gina Kolata, *Questions Raised on Lung Operation*, N.Y. TIMES, Aug. 15, 2001, at A1; *see also* Lars Noah, *Medicine's Epistemology: Mapping the Haphazard Diffusion of Knowledge in the Biomedical Community*, 44 ARIZ. L. REV. 373, 387–88 (2002) ("RCTs discredit long-accepted medical treatments with disturbing regularity . . . ."); *id.* at 393–94 ("[P]hysicians may embrace new procedures and technologies prematurely, before much evidence exists to support their enthusiasm."); *id.* at 447 ("The FDA's premarket review mechanisms and other controls generate substantial information about drugs and medical devices. No similar regulatory regime exists with regard to surgical techniques and other types of therapeutic interventions . . . ."). Although not something likely to have relevance to Medicare beneficiaries, fertility treatments have followed a similar trajectory. *See* Lars Noah, *Assisted Reproductive Technologies and the Pitfalls of Unregulated Biomedical Innovation*, 55 FLA. L. REV. 603, 617–18, 665 (2003).

13. *See* Nat'l Emphysema Treatment Trial Research Group, *Patients at High Risk of Death After Lung-Volume–Reduction Surgery*, 345 NEW ENG. J. MED. 1075, 1080–82 (2001). Ultimately, the trial showed modest benefits for other patient subgroups, and CMS agreed to reimburse the surgery for such patients. *See* Scott D. Ramsey & Sean D. Sullivan, *Evidence, Economics, and Emphysema: Medicare's Long Journey with Lung Volume Reduction Surgery*, 24 HEALTH AFF. 55, 59–61 (2005). Nonetheless, surprisingly few beneficiaries thereupon underwent the procedure. *See* Gina Kolata, *Medicare Says It Will Pay, but Patients Say 'No Thanks,'* N.Y. TIMES, Mar. 3, 2006, at C1.

14. *See* Peter J. Neumann et al., *Medicare's National Coverage Decisions for Technologies,*

condition coverage of off-label uses of expensive new cancer drugs on enrollment in one of several RCTs sponsored by the National Cancer Institute.[15]

Some of the subsequent CEDs seemed harder to understand than the emphysema trial, particularly because they demanded studies of technologies that already had undergone extensive testing as a prelude to scrutiny by the FDA.[16] The latest CSP involves genetic testing of patients receiving the anticoagulant warfarin (Coumadin®). In 2007, the FDA announced revisions in the labeling of this drug to alert physicians that genetic testing might help to identify those patients who risk serious bleeding reactions because they metabolize the drug more slowly than normal—or are more sensitive to its effects—and therefore should receive a lower dose.[17] Nonetheless, two years later CMS announced that it would not cover genetic testing unless Medicare beneficiaries agreed to enroll in a clinical trial.[18] Although the cost-effectiveness of routine screening remains unclear and deserves continued investigation,[19] genetic testing for warfarin

1999–2007, 27 HEALTH AFF. 1620, 1623 (2008) ("CMS used its CED policy in seven decisions through 2007 . . . . Five of the cases involved clinical trials, while two involved data registries."). Their two most recent listed illustrations required participation in RCTs for beneficiaries with less severe impairments as a condition of access to cochlear implants and home-use oxygen. *See id.* at 1627 exh.5 (adding that, in the case of cochlear implants, "[n]o proposals for trials emerged in response"); *see also id.* at n.a ("Several cases that we call CED predate these formal [2005 and 2006] guidances."); *id.* at 1628 (mentioning a "flurry" of more recent CED proposals).

15. *See* Tanisha Carino et al., *Medicare's Coverage of Colorectal Cancer Drugs: A Case Study in Evidence Development and Policy*, 25 HEALTH AFF. 1231, 1235 (2006); *see also id.* at 1237–38 (drawing parallels to the emphysema research, but adding that most other CEDs have involved patient registries). Private insurers have done likewise in some circumstances. *See* Gina Kolata & Kurt Eichenwald, *Group of Insurers to Pay for Experimental Cancer Therapy*, N.Y. TIMES, Dec. 16, 1999, at C1.

16. CMS does not simply defer to FDA approval decisions, even though both agencies reside within the same cabinet-level department. *See* Christopher D. Zalesky, *Considering Changes to CMS's National Coverage Decision Process: Applying Lessons Learned from FDA as a Regulator of Access to Healthcare Technology*, 57 FOOD & DRUG L.J. 73, 76 (2002).

17. *See* David Brown, *For the First Time, FDA Recommends Gene Testing*, WASH. POST, Aug. 17, 2007, at A10; Bernadette Tansey, *A Specific Test for What Ails You*, S.F. CHRON., Sept. 9, 2007, at E1; *see also* Int'l Warfarin Pharmacogenetics Consortium, *Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data*, 360 NEW ENG. J. MED. 753, 754, 759–60 (2009); Mollie Roth, *The Warfarin Revised Package Insert: Is the Information in the Label "Too Thin"?*, 9 HOUS. J. HEALTH L. & POL'Y 279 (2008). *See generally* Lars Noah, *The Coming Pharmacogenomics Revolution: Tailoring Drugs to Fit Patients' Genetic Profiles*, 43 JURIMETRICS J. 1 (2002).

18. *See* Andrew Pollack, *Gene Test for Dosage of Warfarin Is Rebuffed*, N.Y. TIMES, May 5, 2009, at B3 (describing the CMS announcement as a proposal with a one month public comment period). "As many as one million or more Medicare patients a year start therapy with the drug, which is used to prevent life-threatening blood clots." *Id.*

19. *See id.* ("Some studies have shown that using the genetic test might allow the proper

sensitivity would appear to have no downside from the perspective of particular patients or their physicians.

HCFA previously had tried a different approach to encouraging study participation by the elderly. In 2000, the agency extended Medicare coverage for ancillary costs incurred by beneficiaries who chose to enroll in clinical trials,[20] but this policy expressly excluded any payment for the investigational item or service.[21] In contrast, with its CED policy CMS has offered to cover the cost of an item or service, but only when beneficiaries agree to facilitate further investigation. Some critics have argued that CMS has used the CED process for simple delay or implicit rationing of expensive new medical interventions.[22] To be sure, other federal agencies have imposed study requirements as a way of postponing difficult regulatory judgments about continued product marketing.[23] This Article

---

dose to be achieved more quickly. But Medicare said there was little evidence that doing so translated into a lower risk of blood clots or hemorrhages."); *id.* ("[K]nowing which variants of the two genes a patient has does not automatically tell the doctor what dose to give. That depends on other factors as well. Moreover, use of the genetic tests does not eliminate the need to periodically test the patient's blood-clotting propensity."); *see also* Mark H. Eckman et al., *Cost-Effectiveness of Using Pharmacogenetic Information in Warfarin Dosing for Patients with Nonvalvular Atrial Fibrillation*, 150 ANNALS INTERNAL MED. 73, 80–81 (2009); *cf.* Marie McCullough, *Hopes Rising on Finding a Better Blood Thinner*, PHILA. INQUIRER, Oct. 20, 2009, at A1 (describing new and safer substitutes).

20. *See* David Brown, *Medicare to Pay for Experimental Treatments: Clinton Aims to Bring More Seniors into Clinical Trials*, WASH. POST, June 8, 2000, at A9; *see also* Kirk Dobbins & Kay Scanlan, *Medicare's Revised Clinical Trial Policy and Clinical Trial-Related Provisions of FDAAA: What Is a Sponsor to Do?*, 62 FOOD & DRUG L.J. 695, 696–704 (2007) (discussing subsequent developments). This represented an attempt to respond to reports that RCTs included too few elderly subjects. *See, e.g.*, Laura F. Hutchins et al., *Underrepresentation of Patients 65 Years of Age or Older in Cancer-Treatment Trials*, 341 NEW ENG. J. MED. 2061, 2064–66 (1999); *see also* INST. OF MED., EXTENDING MEDICARE REIMBURSEMENT IN CLINICAL TRIALS (2000).

21. *See* Notice of Public Meeting on Medicare Coverage of Clinical Trials, 65 Fed. Reg. 60,442, 60,443 (Oct. 11, 2000).

22. *See* Sandra J. Carnahan, *Medicare's Coverage with Study Participation Policy: Clinical Trials or Tribulations?*, 7 YALE J. HEALTH POL'Y L. & ETHICS 229, 256, 258, 267–68, 272 (2007); Peter W. Groeneveld, Letter to the Editor, *Medicare Requirement for Research Participation*, 296 JAMA 2923 (2006); Carol Gentry, *Why Medicare Covers a New Lung Surgery for Just a Few Patients*, WALL ST. J., June 29, 1998, at A1. Although CMS may limit its reimbursement levels, the agency cannot consider expense when deciding whether or not to cover a medical intervention. *See* Carnahan, *supra*, at 257; *infra* note 82; *see also* Alex Berenson, *Medicare Cuts Payout on 2 Cancer Drugs*, N.Y. TIMES, Dec. 7, 2007, at C3; Andrew Pollack, *Stronger Warnings on 3 Drugs for Anemia*, N.Y. TIMES, Nov. 9, 2007, at C3; Jane Zhang, *Medicare Official Key to Spending*, WALL ST. J., Oct. 27, 2009, at A6. Thus, paying only for enrolled subjects and putting off a final coverage decision until completion of a study allows the agency to limit early access to expensive medical innovations and thereby ration care.

23. *See, e.g.*, Lars Noah & Richard A. Merrill, *Starting from Scratch?: Reinventing the Food Additive Approval Process*, 78 B.U. L. REV. 329, 382–85 (1998) (discussing "interim" food additives); *see also* Lars Noah, *The Imperative to Warn: Disentangling the "Right to Know" from the "Need to Know" About Consumer Product Hazards*, 11 YALE J. ON REG. 293, 397 (1994)

focuses, instead, on the serious ethical questions posed by the CSP policy, but first it contrasts that approach with other federal efforts to encourage individuals to participate in research.

B. Soldiers

The U.S. military has a checkered history when it comes to human experimentation. During the height of the Cold War, for example, the Department of Defense (DOD) sponsored experiments in which cancer patients and prisoners were exposed to total body radiation or plutonium in order to test the body's response, but the investigators made no effort to secure informed consent from the subjects.[24] Most military research efforts have, however, taken advantage of the large pool of active—and, for the most part, highly compliant—service members.[25]

In 1990, the FDA granted a request from the DOD for an exemption to informed consent requirements during the Gulf War in order to inoculate military personnel with unapproved treatments for biowarfare agents.[26]

(explaining comparable motivations behind risk labeling). Along similar lines, the FDA occasionally withdraws the license for a previously approved drug but allows continued use under the strictures of an investigational new drug (IND) exemption. *See, e.g.*, Forsham v. Califano, 442 F. Supp. 203, 205, 210 (D.D.C. 1977) (phenformin); Francesca Lunzer Kritz, *FDA to Weigh New Controls on Problematic Drugs: Lotronex Will Be First for Consideration by New Panel*, WASH. POST, Apr. 16, 2002, at F1 (Propulsid®); Andrew Pollack, *F.D.A. Restricts Access to Cancer Drug, Citing Ineffectiveness*, N.Y. TIMES, June 18, 2005, at C2 (reporting that patients who had benefitted from Iressa® could continue to use it and that the sponsor could continue enrolling subjects in clinical trials).

24. *See In re* Cincinnati Radiation Litig., 874 F. Supp. 796, 800–05 (S.D. Ohio 1995); *see also* CIA v. Sims, 471 U.S. 159, 162 & n.2 (1985) (LSD experiments); Barrett v. United States, 660 F. Supp. 1291, 1299 (S.D.N.Y. 1987) (Army testing of mescaline-derivatives as potential chemical warfare agents on mental hospital patients without their consent). Civilian agencies and private institutions participated in the human radiation experiments as well. *See* Stadt v. Univ. of Rochester, 921 F. Supp. 1023, 1025 (W.D.N.Y. 1996); Advisory Comm. on Human Radiation Experiments, *Research Ethics and the Medical Profession*, 276 JAMA 403 (1996); Nestor M. Davidson, Note, *Constitutional Mass Torts: Sovereign Immunity and the Human Radiation Experiments*, 96 COLUM. L. REV. 1203, 1203–04, 1226–28, 1233–35 (1996); *see also* Begay v. United States, 768 F.2d 1059, 1060–62, 1064–66 (9th Cir. 1985) (prospective epidemiological study conducted by the U.S. Public Health Service on Navajo uranium miners). *See generally* JONATHAN D. MORENO, UNDUE RISK: SECRET STATE EXPERIMENTS ON HUMANS (2000).

25. *See, e.g.*, United States v. Stanley, 483 U.S. 669, 686–89 (1987) (Brennan, J., dissenting in part) (criticizing the Army's secret LSD experiments); Jaffee v. United States, 663 F.2d 1226, 1229 (3d Cir. 1981) (en banc) (nuclear fallout); Thom Shanker, *Reports Detail Tests of Troops for Exposures*, N.Y. TIMES, July 1, 2003, at A21.

26. *See* Informed Consent for Human Drugs and Biologics; Determination That Informed Consent Is Not Feasible, 55 Fed. Reg. 52,814, 52,817 (Dec. 21, 1990) (codified as amended at 21 C.F.R. § 50.23(d) (2009)). Although some have argued that the DOD sought only to provide treatment for soldiers rather than engage in genuine research, the FDA's waiver clearly anticipated that study protocols would govern the use of the unapproved

Military officials feared that some soldiers would refuse, which then might create difficulties in the field in the event of exposure to biological and chemical weapons. Apart from doubts about the military's claim that this made it "not feasible" to secure informed consent, which represents the statutory standard for waiving the FDA's requirements,[27] Congress had imposed separate consent requirements on the DOD for "research involving a human being as an experimental subject."[28] The federal courts, however, rejected a challenge to the FDA's waiver of informed consent requirements.[29] Expressing evident displeasure with these rulings, Congress subsequently mandated that the DOD secure informed consent from military personnel before administering an investigational drug (whether or not done in connection with genuine experimentation), including an approved drug for an unapproved use, and it provided that only the President could waive this requirement.[30]

These issues returned after 2001, with concerns about bioterrorist attacks in the United States. Under a program begun in 1998 but not fully implemented until mid-2002 (shortly before the invasion of Iraq), the DOD inoculated service members and certain civilian contractor employees with anthrax vaccine adsorbed (AVA).[31] The FDA licensed the vaccine for the

products, that other human subject protections would remain in place, and that the DOD would collect data in pursuit of filing applications for marketing approval. *See id.* at 52,815–16.

27. *See* 21 U.S.C. § 355(i)(4) (2006).

28. Department of Defense Authorization Act of 1985, Pub. L. No. 98-525, § 1401(c), 98 Stat. 2492, 2615 (1984) (codified at 10 U.S.C. § 980 (2006)); *see also* Elliott J. Schuchardt, *Distinguishing Between Research and Medical Practice During Operation Desert Storm*, 49 FOOD & DRUG L.J. 271, 277–89 (1994) (concluding that the DOD had not conducted research in violation of this statute); Ruth K. Miller, Note, *Informed Consent in the Military: Fighting a Losing Battle Against the Anthrax Vaccine*, 28 AM. J.L. & MED. 325, 339–43 (2002) (defending the DOD's program).

29. *See* Doe v. Sullivan, 756 F. Supp. 12, 16 (D.D.C.) ("The fact that the DoD will collect information on the efficacy of the drugs does not transform the strategic decision to use the unapproved drugs in combat into research."), *aff'd*, 938 F.2d 1370, 1379–83 (D.C. Cir. 1991); *see also* George J. Annas, *Changing the Consent Rules for Desert Storm*, 326 NEW ENG. J. MED. 770, 772 (1992) (agreeing that the DOD was not engaging in research); Robyn Pforr Ryan, *Should Combat Troops Be Given the Option of Refusing Investigational Drug Treatment?*, 52 FOOD & DRUG L.J. 377, 393 (1997) (criticizing the waiver, noting that, although "DOD did not administer the treatment with the primary intent of generating new knowledge," the drugs were experimental in the sense that uncertainty remained about their safety and efficacy); Claire A. Milner, Comment, *Gulf War Guinea Pigs: Is Informed Consent Optional During War?*, 13 J. CONTEMP. HEALTH L. & POL'Y 199, 223–31 (1996).

30. *See* Pub. L. No. 105-85, § 766(a), 111 Stat. 1629, 1827 (1997) (codified as amended at 10 U.S.C. § 1107 (2006)); *see also* Exec. Order No. 13,139, 64 Fed. Reg. 54,175, 54,176 (Oct. 5, 1999) (announcing that the President would evaluate waiver requests using the criteria set forth in the FDA's regulation).

31. *See* Guy Gugliotta, *Pentagon to Resume Anthrax Vaccinations*, WASH. POST, June 29, 2002, at A3.

prevention of cutaneous anthrax, but it had expressed doubts about its efficacy against inhalation anthrax,[32] and, in 1996, the manufacturer submitted an investigational new drug (IND) application to undertake research that would support adding that indication to the labeling.[33] The military's Anthrax Vaccine Immunization Program (AVIP) did not, however, make any provision for securing informed consent before inoculating soldiers with AVA.[34]

A group of service members and civilian employees challenged the program, arguing that the use of a drug to protect against the risk of inhalation anthrax but licensed only to guard against cutaneous exposure was investigational and therefore required informed consent under statute unless waived by a presidential order.[35] After rejecting the government's nonjusticiability arguments,[36] and concluding that AVA remained "investigational" against inhalation anthrax,[37] a federal judge granted petitioners a preliminary injunction. The court found no merit in the DOD's claims of necessity,[38] and it concluded that, "[a]bsent an informed consent or presidential waiver, the United States cannot demand that members of the armed forces also serve as guinea pigs for experimental drugs."[39]

One week after the court's order (and eighteen years after issuing its proposal), the FDA published a final rule that found AVA safe and effective for protection against inhalation anthrax.[40] After invalidating this rule on procedural grounds, the district court issued a permanent injunction against implementation of the AVIP.[41] In response, the government invoked a newly enacted provision that authorized the use of unapproved drugs during a declared national emergency.[42] Just as the six-month

32. *See* Biological Products; Bacterial Vaccines and Toxoids; Implementation of Efficacy Review, 50 Fed. Reg. 51,002 (Dec. 13, 1985).

33. *See* Randall D. Katz, Note, *Friendly Fire: The Mandatory Military Anthrax Vaccination Program*, 50 DUKE L.J. 1835, 1853–54, 1859 (2001).

34. *See* Doe v. Rumsfeld, 297 F. Supp. 2d 119, 125 (D.D.C. 2003).

35. *See id.* at 122–23.

36. *See id.* at 126–31.

37. *See id.* at 131–34.

38. *See id.* at 134.

39. *Id.* at 135.

40. *See* Biological Products; Bacterial Vaccines and Toxoids; Implementation of Efficacy Review, 69 Fed. Reg. 255, 259 (Jan. 5, 2004).

41. *See* Doe v. Rumsfeld, 341 F. Supp. 2d 1, 19 (D.D.C. 2004); *cf.* Ammend v. BioPort, Inc., 322 F. Supp. 2d 848, 870–73 (W.D. Mich. 2004) (rejecting constitutional claims against entities that had supplied anthrax vaccine to DOD).

42. *See* Marc Kaufman, *Pentagon Boosts Plan for Anthrax Inoculations: Emergency Provisions Invoked to Revive Use*, WASH. POST, Feb. 2, 2005, at A3 (reporting that soldiers would have the right to refuse). When it authorized the use of unapproved medical products in the event of an emergency (and without distinguishing between civilians and military personnel),

authorization for emergency use of AVA expired, the FDA reissued its final order concluding that the vaccine was effective against inhalation anthrax,[43] which removed the drug from IND status and thereby avoided application of the consent requirements imposed by Congress.[44]

At some level, the Pentagon's tortuous efforts over the last couple of decades to avoid securing informed consent seem odd. After all, presumably it could discharge (dishonorably or otherwise) any service member who refused to consent, which would mean that only rarely would a soldier decline to participate.[45] Perhaps military officials realized that consent to research secured under such circumstances would not pass muster as genuinely voluntary.[46] Of course, once the FDA approves a medical product for a particular use, the special consent requirements governing experimentation become inapplicable, and the Pentagon then could force military personnel to get inoculated or face discharge in the event of refusal.[47]

Congress also had included an informed consent requirement. *See* Project BioShield Act of 2004, Pub. L. No. 108-276, § 4(a), 118 Stat. 853 (codified at 21 U.S.C. § 360bbb-3(e)(1)(A)(ii) (2006)).

43. *See* Biological Products; Bacterial Vaccines and Toxoids; Implementation of Efficacy Review; Anthrax Vaccine Adsorbed, 70 Fed. Reg. 75,180, 75,183 (Dec. 19, 2005).

44. *See* Doe v. Rumsfeld, 172 F. App'x 327, 328 (D.C. Cir. 2006) (per curiam) (rejecting the government's appeal as moot because the FDA's approval satisfied the district court's injunction).

45. *See* Neely Tucker, *Anthrax Vaccine Challenged: Two Suing Defense Department over Inoculation Policy*, WASH. POST, May 15, 2002, at A10 (reporting that approximately 500 service members had declined anthrax vaccinations and that some of those faced courts martial).

46. *See* Keri D. Brown, Comment, *An Ethical Obligation to Our Servicemembers: Meaningful Benefits for Informed Consent Violations*, 47 S. TEX. L. REV. 919, 935 (2006) (calling the "voluntariness" element arguably the biggest problem for the military and the reason they have procedures for informed consent waivers); *cf.* Scott Fontaine, *Blood Pressure Pill Slays Nightmares*, WASH. POST, Dec. 31, 2009, at A15 (reporting that military hospitals have recruited active service members and veterans to participate in genuine research of a promising off-label use of prazosin).

47. *See* George J. Annas, *Protecting Soldiers from Friendly Fire: The Consent Requirement for Using Investigational Drugs and Vaccines in Combat*, 24 AM. J.L. & MED. 245, 250, 257 & n.47 (1998); Katz, *supra* note 33, at 1848; *Pentagon Set to Vaccinate Troops, Assist in Flu Crisis*, WASH. POST, Sept. 30, 2009, at A6; *see also* Catherine L. Annas & George J. Annas, *Enhancing the Fighting Force: Medical Research on American Soldiers*, 25 J. CONTEMP. HEALTH L. & POL'Y 283, 291 (2009) ("'Questions of coercion and autonomy are particularly acute for military personnel . . . . Soldiers in the United States . . . *are legally required to take medications if ordered to for the sake of military performance*.'" (quoting Henry Greely et al., *Towards Responsible Use of Cognitive-Enhancing Drugs by the Healthy*, 456 NATURE 702, 703 (2008))); *cf. id.* at 300 (questioning this premise); *id.* at 296 n.58 (quoting a consent form that military pilots "are required to sign" before taking dextroamphetamine off-label to manage fatigue); *id.* at 308 ("The military rule should be that prescription medications should never be forced on soldiers, but should be taken only voluntarily, and only on the advice of a physician who cannot be ordered to prescribe it.").

C. Civilians

In contrast to CMS, which deals with beneficiaries and providers through reimbursement choices (and affects sellers only indirectly), and in contrast to DOD, which controls service members (and, in any event, generally does not engage in genuine medical research), the FDA exercises its authority over sellers of medical technologies (and affects patients and providers only indirectly). Thus, where CMS would find it difficult to obligate sellers to undertake further research as a condition of coverage (and certainly could not force beneficiaries to participate in such research),[48] the FDA may do so as a condition of approval. Even if undertaken by private industry at the behest of a regulatory agency, these research requirements may have the same potentially adverse impact on patients seeking access to new technologies.

The FDA requires that, before shipping an unapproved new drug to initiate human trials, sponsors file an IND application.[49] The results of these trials provide the basis for agency decisions when sponsors subsequently file an application for new drug approval (NDA).[50] Even if the FDA acts favorably on the NDA, this hardly closes the book on the safety and efficacy of a regulated product. The issuance of a product license does not magically transform an investigational medical technology into one that has matured fully and requires no additional scrutiny.[51]

---

48. Actually, the CED policy indirectly creates pressure on sellers as well as beneficiaries. *See* Kolata, *supra* note 9, at A1 ("For the first time in its history, Medicare has systematically begun to make payments for new and expensive treatments and diagnostic tests conditional on agreement by companies or other groups to pay for studies on whether these new methods actually work on the Medicare patients who get them."); *id.* ("[CMS] is using the threat of refusing to pay unless patients are in a study as a cudgel to get companies or foundations or professional groups to pay for the research."); *id.* ("[W]ith Medicare the dominant payer for elderly Americans, who are most likely to need the treatments, its clout, when it insists on studies, is substantial.").

49. *See* 21 U.S.C. § 355(i) (2006); 21 C.F.R. pt. 312 (2009); *see also* David A. Kessler, *The Regulation of Investigational Drugs*, 320 NEW ENG. J. MED. 281, 281 (1989); Richard A. Merrill, *The Architecture of Government Regulation of Medical Products*, 82 VA. L. REV. 1753, 1766–67, 1777–82 (1996). Similarly, sponsors of certain medical devices must apply for an investigational device exemption (IDE) before undertaking clinical trials. *See* 21 U.S.C. § 360j(g); 21 C.F.R. pt. 812; *see also* Martin v. Telectronics Pacing Sys., Inc., 105 F.3d 1090, 1095–96 (6th Cir. 1997).

50. *See* 21 U.S.C. § 355(b); 21 C.F.R. pt. 314.

51. *See* Annetine C. Gelijns et al., *Capturing the Unexpected Benefits of Medical Research*, 339 NEW ENG. J. MED. 693, 693 (1998) ("The end of the research-and-development process does not entail the elimination of all, or even most, of the uncertainties surrounding medical innovation."); Margaret Gilhooley, *When Drugs Are Safe for Some but Not Others: The FDA Experience and Alternatives for Products Liability*, 36 HOUS. L. REV. 927, 936 (1999) (recognizing that "the initial use of a drug is, in effect, a continuation of the testing" phase); Wayne A. Ray et al., *Evaluating Drugs After Their Approval for Clinical Use*, 329 NEW ENG. J. MED. 2029,

In particular, safety questions often arise after approval.[52] For that reason, researchers increasingly have taken advantage of databases maintained by health insurers, both public and private.[53] For instance, Medicare billing records may allow investigators to discern rates of complications associated with particular procedures.[54] Similarly, the FDA has discovered important drug side effects from retrospective reviews of Medicaid records,[55] and the agency has announced plans to make more regular use of such sources of information in the future.[56]

Patient registries provide a somewhat more structured mechanism for tracking outcomes. During the 1990s, the FDA (together with HCFA) used a registry to evaluate cardiac pacemakers.[57] The agency also has urged that

---

2029–30 (1993).

52. *See* U.S. GEN. ACCOUNTING OFFICE, FDA DRUG REVIEW: POST APPROVAL RISKS 1976–85, at 3 (1990) (concluding that more than half of all drugs approved between 1976 and 1985 had serious risks that were discovered only after approval); Karen E. Lasser et al., *Timing of New Black Box Warnings and Withdrawals for Prescription Medications*, 287 JAMA 2215, 2218–19 (2002); Robert J. Temple & Martin H. Himmel, Editorial, *Safety of Newly Approved Drugs: Implications for Prescribing*, 287 JAMA 2273, 2275 (2002); Naomi Aoki, *A Question of Speed and Safety*, BOSTON GLOBE, Nov. 28, 2001, at G1 (noting "the growing number of drugs that have been recalled in the past three years—nearly a dozen implicated in more than 1,000 deaths").

53. *See* Wayne A. Ray et al., *Adverse Drug Reactions and the Elderly*, HEALTH AFF., Fall 1990, at 114, 120; Ricardo Alonso-Zaldivar, *Medicare's Will May Be FDA's Way*, L.A. TIMES, June 5, 2005, at A1; David Brown, *Congress Seeks to Balance Drug Safety, Quick Approval*, WASH. POST, July 5, 2007, at A4.

54. *See, e.g.*, Steve Sternberg, *Higher Price for Defibrillator Implants*, USA TODAY, June 26, 2006, at 5D (reporting that such a review found higher-than-expected rates of complications and associated hospitalization costs associated with ICDs); *id.* ("[CMS] hopes to use a registry started in January 2005 to sharpen the focus on why so many complications are occurring and how to reduce their number. The registry now has records from 51,000 patients."); *see also* Stephen F. Jencks et al., *Quality of Medical Care Delivered to Medicare Beneficiaries: A Profile at State and National Levels*, 284 JAMA 1670, 1675–76 (2000).

55. *See, e.g.*, David Brown, *Blood-Pressure Drugs Linked to Birth Defects*, WASH. POST, June 8, 2006, at A12 (reporting that research funded by the FDA and using one state's Medicaid records discovered a significant increase in the risk of birth defects when pregnant women used ACE inhibitors during their first trimester).

56. *See* News Release, FDA, Health Organizations to Study Safety of Medications Taken During Pregnancy (Dec. 30, 2009), http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm195934.htm; Rob Stein, *Program Aims for Drug, Device Safety*, WASH. POST, May 23, 2008, at A2; *see also* Medicare Program; Medicare Part D Data, 71 Fed. Reg. 61,445, 61,450–52 (proposed Oct. 18, 2006) (to be codified at 42 C.F.R. pt. 423); Natasha Singer, *Public Database Is Urged to Monitor Drug Safety*, N.Y. TIMES, Nov. 24, 2009, at B2.

57. *See* Cardiac Pacemaker Registry, 52 Fed. Reg. 27,756, 27,763–64 (July 23, 1987) (codified at 21 C.F.R. pt. 805 (1999)) (making Medicare reimbursement of physicians, hospitals, and other providers contingent on providing information about the implantation or removal of these devices), *revoked*, 64 Fed. Reg. 66,105 (Nov. 24, 1999). Newer agency initiatives have focused on the establishment of "sentinel" (early warning) systems. *See* Ross Kerber, *FDA Halts Expansion of Network to Monitor Medical Device Safety*, BOSTON GLOBE, July

patients taking suspected teratogens enroll in pregnancy registries, as it did in the case of Accutane® (isotretinoin).[58] Congress recently granted the FDA explicit authority to impose these and other sorts of postapproval study requirements on manufacturers of new drugs.[59]

These programs vaguely resemble the product registration cards that accompany many consumer goods and that most purchasers discard.[60] If obligatory, in the sense that patients must register before receiving a particular drug or device, then such tracking mechanisms seem a bit more intrusive.[61] Even so, patients would face little additional burden and presumably later would remain free to decline to respond to any follow-up requests for information unless they continue using and do not want to lose access to the product at issue. Furthermore, though the FDA has the power to demand that manufacturers undertake rigorous research after approval (so-called Phase IV trials),[62] it generally has no way to encourage

---

14, 2005, at D1.

58. The manufacturer had included an enrollment form for patients to send to the Slone Epidemiology Unit at Boston University's School of Public Health that facilitated the tracking of patient compliance and adverse outcomes. *See* Allen A. Mitchell et al., *A Pregnancy-Prevention Program in Women of Childbearing Age Receiving Isotretinoin*, 333 NEW ENG. J. MED. 101, 102 (1995); *id.* at 104–05 (concluding that the system had worked fairly well, though estimating that less than half of treated women had enrolled); *see also* FDA, General Information About Pregnancy Exposure Registries, http://www.fda.gov/ScienceResearch/SpecialTopics/WomensHealthResearch/ucm134844.htm (last visited Mar. 5, 2010); Margaret A. Honein et al., *Can We Ensure the Safe Use of Known Human Teratogens?: Introduction of Generic Isotretinoin in the US as an Example*, 27 DRUG SAFETY 1069, 1073 (2004) (discussing a voluntary pregnancy registry for antiepileptic drugs). The FDA demanded a similar tracking requirement when it approved Thalomid® (thalidomide) for limited use. *See* Rita Rubin, *Thalidomide Could Guide Use of Drugs That Risk Birth Defects*, USA TODAY, July 22, 1998, at 7D; Sheryl Gay Stolberg, *Thalidomide Approved to Treat Leprosy, with Other Uses Seen*, N.Y. TIMES, July 17, 1998, at A1.

59. *See* Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, § 901(a), 121 Stat. 823, 923 (to be codified at 21 U.S.C. § 355(o)(3)(D)); *see also id.* § 905, 121 Stat. at 944 (to be codified at 21 U.S.C. § 355(k)(3)(C)(i)(III)(aa)) (directing the agency to tap into existing public databases such as that of the Medicare program); Barry Meier, *House Bill Would Create Artificial Joints Registry*, N.Y. TIMES, June 11, 2009, at B3.

60. *See* Cindy Skrzycki, *For Now, Toy-Recall Registration Isn't in the Cards*, WASH. POST, Mar. 11, 2003, at E1.

61. *Cf.* Carnahan, *supra* note 22, at 230 n.7 ("To the extent that [Medicare] coverage is contingent upon patients providing additional (beyond billing) information to a registry for research purposes, CAD may raise some of the same issues regarding voluntary informed consent that are raised with [CSP]."); Meredith Wadman, *Medicare Compels Heart Patients to Enlist in Follow-up Research*, 433 NATURE 341 (2005) (quoting Art Caplan's objection); Michael Kranish, *New Use Is Found for Thalidomide: Fighting Cancer*, BOSTON GLOBE, Oct. 20, 2002, at A28 (reporting objections to Boston University's initial policing role: "[T]he Office of Human Research Protections . . . said that if patients could lose their medicine for not responding to the BU survey, that 'failed to minimize the possibility of coercion or undue influence as required by [HHS].'").

62. *See* Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110–85,

patients to enroll in such studies.

II. FLAWS IN DEFENSES OF MEDICARE'S RESEARCH POLICY

This Part focuses on the ethical questions posed by Medicare's CSP policy, comparing and contrasting previously described instances of arguably nonconsensual research undertaken by or at the behest of the DOD and the FDA. In particular, this Part critically evaluates published defenses of CMS's approach, and it asks more broadly what such arguments may have to tell us about the nature and direction of bioethics in this country. Medicare beneficiaries who enroll in RCTs hoping to access new medical technologies do not genuinely volunteer to serve as research subjects; CSP proponents who cavalierly dismiss ethical objections to this policy have in mind a fundamentally different regime of human research protections than prevails at the present time.

A. Disregarding Concerns About Volitional Impairment

For the most part, past instances of objectionable research with humans have involved failures to disclose information.[63] If individuals do not know that they have become experimental subjects, then sponsors of the research clearly have failed to secure informed consent.[64] Even in the absence of deception, however, subjects may object if their participation was nonconsensual. At its core, informed consent requires both knowledge and volition,[65] and research violates these norms where subjects participate

---

§ 901, 121 Stat. 823, 922 (to be codified at 21 U.S.C. § 355(o)(3)); Charles Steenburg, *The Food and Drug Administration's Use of Postmarketing (Phase IV) Study Requirements: Exception to the Rule?*, 61 FOOD & DRUG L.J. 295, 325–27 (2006); Jennifer Corbett Dooren, *Drug Makers Seen as Slow to Finish Postmarket Studies*, WALL ST. J., June 1, 2005, at D4.

63. *See, e.g.*, Henry K. Beecher, *Ethics and Clinical Research*, 274 NEW ENG. J. MED. 1354 (1966) (discussing twenty-two examples of research studies conducted without consent of the subjects); William J. Curran, *The Tuskegee Syphilis Study*, 289 NEW ENG. J. MED. 730 (1973); Barron H. Lerner, *Sins of Omission—Cancer Research Without Informed Consent*, 351 NEW ENG. J. MED. 628, 629–30 (2004); Lawrence K. Altman, *Fatal Drug Trial Raises Questions About 'Informed Consent,'* N.Y. TIMES, Oct. 5, 1993, at C3; Marlene Cimons, *CDC Says It Erred in Measles Study*, L.A. TIMES, June 17, 1996, at A11; Sandy Rovner, *Ethics Concerns Raised in Schizophrenia Study*, WASH. POST, Sept. 29, 1992, at F7; *see also supra* note 24 (referencing secret radiation experiments).

64. *See generally* JESSICA W. BERG ET AL., INFORMED CONSENT: LEGAL THEORY AND CLINICAL PRACTICE (2d ed. 2001); Karine Morin, *The Standard of Disclosure in Human Subject Experimentation*, 19 J. LEGAL MED. 157 (1998).

65. *See* RUTH R. FADEN & TOM L. BEAUCHAMP, A HISTORY AND THEORY OF INFORMED CONSENT 238–39, 256–57 (1986); *id.* at 337 ("Disclosing, informing, and comprehending are the most widely discussed topics in traditional commentary on informed consent. But remaining independent of control by others is equally important for autonomous decisionmaking."); NAT'L BIOETHICS ADVISORY COMM'N, ETHICAL AND

knowingly but involuntarily.[66]

Consent, which expresses an individual's decision to volunteer to serve as an experimental subject, is widely recognized as the central ethical requirement for conducting clinical research. In its very first sentence, the Nuremberg Code insisted on "voluntary consent," with affiliated demands for adequate disclosure of information mentioned only secondarily.[67] Similarly, the International Covenant on Civil and Political Rights emphasized that "no one shall be subjected without his *free consent* to medical or scientific experimentation."[68] It may be easier to discern and

---

POLICY ISSUES IN RESEARCH INVOLVING HUMAN PARTICIPANTS 98–99 (2001), http://bioethicsprint.bioethics. gov/reports/past_commissions/nbac_human_part.pdf; Benjamin Freedman, *A Moral Theory of Informed Consent*, HASTINGS CTR. REP., Aug. 1975, at 32, 35–37; Peter H. Schuck, *Rethinking Informed Consent*, 103 YALE L.J. 899, 900 (1994) ("To say that one cannot be bound by a promise that one did not voluntarily and knowingly make is to say that the individual should be the author of her own undertakings, that a genuine respect for her dignity requires a broad deference to her choices.").

66. *See, e.g.*, Blanton v. United States, 428 F. Supp. 360, 361–63 (D.D.C. 1977) (imposing tort liability on a government hospital for administering an FDA-approved drug to a patient as part of a clinical trial to determine its effectiveness beyond the labeled shelf life after the patient had specifically declined to participate as a subject). *See generally* Robert M. Nelson & Jon F. Merz, *Voluntariness of Consent for Research: An Empirical and Conceptual Review*, 40 MED. CARE V-69 (2002).

67. Although it was only one of ten principles enunciated in the Nuremberg Code, consent received top billing:

> The *voluntary consent* of the human subject is *absolutely essential*. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to *exercise free power of choice*, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have *sufficient knowledge* and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonably to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment. The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment.

The Nuremberg Code (1947), *reprinted in* 276 JAMA 1691 (1996) (emphasis added); *see also* Evelyne Shuster, *Fifty Years Later: The Significance of the Nuremberg Code*, 337 NEW ENG. J. MED. 1436, 1439 (1997). *But cf.* Ezekiel J. Emanuel et al., *What Makes Clinical Research Ethical?*, 283 JAMA 2701, 2701–02 (2000) (noting "the near obsession with autonomy in US bioethics," but cautioning that the Nuremberg Code and other ethical guidelines "were written in response to specific events" and therefore "tend to emphasize certain ethical requirements while eliding others"); *id.* at 2706 (discussing consent).

68. International Covenant on Civil and Political Rights art. 7, Dec. 16, 1966, 999 U.N.T.S. 171, 175 (1976) (emphasis added). The Declaration of Helsinki made voluntary participation one of many requirements for research. *See* WORLD MED. ASS'N DECLARATION OF HELSINKI, ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS ¶ 22 (2008),

criticize instances of inadequate disclosure, but we also must guard against situations where researchers take advantage of the constrained choices available to fully informed individuals.

A handful of commentators have questioned the ethical propriety of Medicare's CSP policy.[69] The most focused discussion of the issue, however, offered a thorough-going defense of the approach. An article published in the *Journal of the American Medical Association* (*JAMA*) by Steven Pearson and colleagues from NIH's Department of Clinical Bioethics—Ezekiel Emanuel and Franklin Miller—found much to praise and little to criticize in this "bold initiative by the CMS to use its considerable power as a public insurer to promote efforts to improve the evidence available on critical clinical questions for many health care decision makers."[70] Considerable power indeed!

After pointing out that some CEDs such as patient registries may not qualify as "research" in the first place, Pearson et al. rightly conceded that CSPs unmistakably fall within the category.[71] When a treatment

---

http://www.wma.net/en/30publications/10policies/b3/17c. pdf; *see also id*. ¶ 24 (calling for "freely-given" consent only after specifying the need for adequate disclosures of information); George J. Annas, *The Legacy of the Nuremberg Doctors' Trial to American Bioethics and Human Rights*, 10 MINN. J.L. SCI. & TECH. 19, 24 & n.19 (2009) (describing the Declaration of Helsinki as designed to offer a more flexible set of ethical guidelines than the rigid and legalistic Nuremberg Code); *id*. at 23–26 (explaining that bioethics originated with the Code); Troyen A. Brennan, *Proposed Revisions to the Declaration of Helsinki—Will They Weaken the Ethical Principles Underlying Human Research?*, 341 NEW ENG. J. MED. 527 (1999) (cautioning against a turn toward utilitarianism).

69. *See, e.g.*, Carnahan, *supra* note 22, at 232 (calling CSP "ethically problematic"); *id*. at 262–66; *id*. at 268 (concluding that CSP likely "violates the federal regulations for the protection of human subjects"); *id*. at 270–71 (defending CAD as a preferable approach). Ms. Carnahan's article focused, however, on questions about whether CMS enjoyed the statutory authority to impose such a coverage requirement, and she made no effort to confront ethical defenses of the program authored by its primary architects.

70. Steven D. Pearson et al., *Medicare's Requirement for Research Participation as a Condition of Coverage: Is It Ethical?*, 296 JAMA 988, 990 (2006). Dr. Pearson disclosed that, "from September 2005 through June 2006, he [had] served as Special Advisor, Technology and Coverage Policy, at [CMS.]." *Id*. (Moreover, NIH and CMS are sister agencies housed within HHS that have collaborated on particular CEDs.) A subsequently published defense with one of his earlier co-authors revealed that Pearson had joined NIH's Department of Clinical Bioethics. *See* Franklin G. Miller & Steven D. Pearson, *Coverage with Evidence Development: Ethical Issues and Policy Implications*, 46 MED. CARE 746, 746 (2008). Pearson's other original co-author, Zeke Emanuel, recently left NIH to join his older brother Rahm in the White House. *See* Robert Pear, *Hard-Charging Doctor Adds Perspective to the President's Health Care Team*, N.Y. TIMES, Apr. 18, 2009, at A10.

71. *See* Pearson et al., *supra* note 70, at 989 (recognizing "active debate and disagreement among experts over whether registries and other forms of health care services research require full, partial, or no informed consent"); *id*. (conceding that some CEDs "have linked coverage to studies that all would acknowledge are research"). According to the authors: "In a registry, all eligible patients receive the treatment, and in most registries

relationship gets converted into part of a clinical trial, the patient becomes a "subject" (and the physician becomes an "investigator");[72] the subject may or may not receive the investigational intervention—randomization and blinding make the assignment a matter of chance and secrecy—and probably will have to undergo more frequent follow-up monitoring than normal.[73] Given these features of research, it becomes critical to determine whether a subject has volunteered.

Pearson et al. also conceded that, "[a]lthough CED is clearly well-intentioned, it raises several important ethical questions."[74] As they saw it, however, these questions boil down to asking whether conditioning Medicare coverage on enrollment in a clinical trial amounts to "coercion,"[75] and they concluded that it does not because (1) beneficiaries

---

patients face minimal additional research burdens while often benefiting from the information gained." *Id.*; *see also* Ancheff v. Hartford Hosp., 799 A.2d 1067, 1071–72, 1082 (Conn. 2002) (affirming a jury's conclusion that a hospital's protocol for off-label use of an antibiotic did not qualify as research); Hecht v. Kaplan, 645 N.Y.S.2d 51, 53 (App. Div. 1996) (rejecting the plaintiff's claim that the decision to perform an additional diagnostic test on a sample of her blood amounted to experimentation without consent in violation of state statute); *cf.* Schwartz v. Boston Hosp. for Women, 422 F. Supp. 53, 55–56 (S.D.N.Y. 1976) (denying a motion for summary judgment on plaintiff's claim that she had not consented to an experimental procedure, though she had agreed to the use of records concerning her obstetrical treatment at a hospital participating in a national study of pregnant diabetics, because the court found "a fact question of whether the curettage was performed for purposes of the MIH study rather than to aid in the diagnosis and treatment of Mrs. Schwartz"). *See generally* Lars Noah, *Informed Consent and the Elusive Dichotomy Between Standard and Experimental Therapy*, 28 AM. J.L. & MED. 361, 403–04 (2002). For suggestions that even patient registries may raise ethical concerns, see *supra* note 61 and accompanying text.

72. *See* Jay Katz, *Human Experimentation and Human Rights*, 38 ST. LOUIS U. L.J. 7, 15–16, 33 (1993); Franklin G. Miller et al., *Professional Integrity in Clinical Research*, 280 JAMA 1449, 1450–51 (1998).

73. *See* Donna T. Chen et al., *Clinical Research and the Physician–Patient Relationship*, 138 ANNALS INTERNAL MED. 669, 669 (2003) (explaining that "participation in some trials may include medication washout periods, biopsies, overnight hospital stays, imaging studies, blood draws, and questionnaires"); Jesse A. Goldner, *An Overview of Legal Controls on Human Experimentation and the Regulatory Implications of Taking Professor Katz Seriously*, 38 ST. LOUIS U. L.J. 63, 121–22 (1993); Franklin G. Miller, *Research Ethics and Misguided Moral Intuition*, 32 J.L. MED. & ETHICS 111, 112 (2004); Franklin G. Miller & Donald L. Rosenstein, *The Therapeutic Orientation to Clinical Trials*, 348 NEW ENG. J. MED. 1383, 1383 (2003); E. Haavi Morreim, *The Clinical Investigator as Fiduciary: Discarding a Misguided Idea*, 33 J.L. MED. & ETHICS 586, 587, 589–90 (2005).

74. Pearson et al., *supra* note 70, at 989.

75. The authors separately defended the fairness of CED in instances where required patient registries or RCTs may not exist in certain parts of the country. *See id.* ("Inequality in access to research programs is a regrettable practical reality but does not constitute an injustice."); *id.* at 990 ("[W]ithout CED there would be no coverage at all, so inconsistent access to the technology after CED, while not ideal, is not unethical. Unequal access is not remedied by denying opportunities for all."). Inequities may, however, arise for reasons unrelated to limited geographic coverage. *See* Carnahan, *supra* note 22, at 259 (discussing

would remain free to secure access without enrolling in a study if willing to pay out of pocket,[76] and (2) beneficiaries had no right to expect any coverage in the first place.[77] This pair of assumptions led the authors to conclude that the CED approach represented a win–win situation,[78] but neither premise withstands close scrutiny.

First, the notion that Medicare patients remain free to access items and services in the open market pays insufficient attention to the financial circumstances confronting most beneficiaries.[79] It would seem equally implausible to defend research using persons in poor countries by noting that they could have paid out of pocket for the health care intervention under investigation. One decade ago, placebo-controlled trials of human immunodeficiency virus (HIV) drugs in developing countries generated tremendous controversy—defenders of the research did not make the absurd point that subjects theoretically had the option of purchasing antiretrovirals; instead, they argued that subjects given a fifty percent chance of receiving drug treatment were better off than they otherwise would have been precisely because affordability barriers made it impossible for the vast majority of patients to get such treatments.[80] Whatever one

---

restrictive enrollment criteria that would exclude beneficiaries with co-morbidities and result in underpowered trials).

76. *See* Pearson et al., *supra* note 70, at 988–89.

77. *See id*. at 988 ("CED should not be viewed as coercive because Medicare patients are not entitled to new technologies that would not receive coverage in the absence of CED."); *id*. ("Before CED, therefore, evidence that did not quite reach the CMS interpretation of 'reasonable and necessary' routinely led to a denial of coverage for a new technology."); *id*. at 989 ("Medicare explicitly conveys no entitlement to insurance coverage for all new technologies, only to those technologies judged by the CMS to be 'reasonable and necessary.'"); *id*. ("[C]overage has always been routinely denied for technologies that fail to meet Medicare's interpretation of its evidentiary standards. Without CED, coverage denial would thus be the common fate of these technologies.").

78. *See id*. at 989 ("CED was designed so that patients would gain earlier access to promising but as yet unproven technologies, industry would receive payment for innovations that would not have been covered otherwise, and all health care decision makers would benefit from the generation of better evidence on the true risks, benefits, and costs.").

79. *See* Carnahan, *supra* note 22, at 235; *id*. at 265 ("Given the high cost of new health care technology, no realistic possibility of private purchase exists."); *see also* Emily Brandon, *Even with Medicare, Health Costs Pack a Wallop*, ORLANDO SENT., Mar. 11, 2010, at G2; Deborah Thorne et al., *The Increasing Vulnerability of Older Americans: Evidence from the Bankruptcy Court*, 3 HARV. L. & POL'Y REV. 87, 100 (2009); *New Formula Shows More Live in Poverty*, BOSTON GLOBE, Oct. 21, 2009, at A2 ("About 18.7 percent of Americans 65 and older, or nearly 7.1 million, are in poverty . . . ."). Insofar as CMS effectively exercises a monopoly over the health care options available to the elderly, the agency's use of that power to encourage study participation would differ little from the leverage that the Federal Bureau of Prisons might enjoy if it wanted to promote research by conditioning inmates' access to particular health services on their willingness to enroll in RCTs. *See infra* note 125 (discussing research on prisoners).

80. *See* David P. Fidler, *"Geographical Morality" Revisited: International Relations, International*

may think about their ethical propriety, the overseas HIV drug trials clearly raised eyebrows among many in the research community, while similarly structured trials involving American seniors evidently have not attracted much notice.

Second, the noncoverage baseline that Pearson et al. assumed suffers from an inevitable contingency. One CED announced shortly before CMS formulated its draft guidance represented an instance where the agency had refused coverage before intense lobbying forced it to accept conditional coverage as a compromise.[81] Historically, however, the precise criteria used in making Medicare coverage decisions—whether local or national—have proven difficult to discern.[82] Perhaps CED provides the agency with a useful half-step toward coverage in close cases, but, in the absence of this option, it seems equally likely that CMS would allow coverage,[83] deferring to the judgments of physicians[84] and local contractors,[85] unless and until it

Law, and the Controversy over Placebo-Controlled HIV Clinical Trials in Developing Countries, 42 HARV. INT'L L.J. 299, 306–13 & n.82 (2001); David Orentlicher, *Universality and Its Limits: When Research Ethics Can Reflect Local Circumstances*, 30 J.L. MED. & ETHICS 403, 404–07 (2002); Harold T. Shapiro & Eric M. Meslin, *Ethical Issues in the Design and Conduct of Clinical Trials in Developing Countries*, 345 NEW ENG. J. MED. 139, 140 (2001).

81. *See* Rick Weiss, *A Tale of Politics: PET Scans' Change in Medicare Coverage*, WASH. POST, Oct. 14, 2004, at A1.

82. *See* Medicare Program; Revised Process for Making Medicare National Coverage Decisions, 68 Fed. Reg. 55,634 (Sept. 26, 2003); Medicare Program; The National and Local Coverage Determination Review Process, 66 Fed. Reg. 54,253 (Oct. 26, 2001); Susan Bartlett Foote, *Why Medicare Cannot Promulgate a National Coverage Rule: A Case of* Regula Mortis, 27 J. HEALTH POL. POL'Y & L. 707, 711–12, 715–20 (2002); Muriel R. Gillick, *Medicare Coverage for Technological Innovations—Time for New Criteria?*, 350 NEW ENG. J. MED. 2199, 2202 (2004); Sean R. Tunis, Editorial, *Why Medicare Has Not Established Criteria for Coverage Decisions*, 350 NEW ENG. J. MED. 2196 (2004); *see also* Eleanor D. Kinney, *Medicare Coverage Decision-Making and Appeal Procedures: Can Process Meet the Challenge of New Medical Technology?*, 60 WASH. & LEE L. REV. 1461, 1471–72 (2003); *id.* at 1462 ("Medicare coverage policy for new medical technology has been a very controversial issue in the administration of the Medicare program since its inception."); *id.* at 1501 ("[T]he development of criteria for making coverage decisions has been a very intractable issue for the Medicare program since coverage surfaced as a serious policy issue in the 1980s.").

83. *See, e.g.*, Reed Abelson, *Heart Scans Still Covered by Medicare*, N.Y. TIMES, Mar. 13, 2008, at C1 (reporting that CMS dropped its earlier proposal to impose a CSP requirement on cardiac computed tomography angiography even though it remained skeptical about the usefulness of the procedure); Barnaby J. Feder, *U.S. Expands Some Stent Reimbursement Coverage*, N.Y. TIMES, Mar. 18, 2005, at C4 (reporting that Medicare broadened payment for carotid stenting six months after the FDA approved the first device for use in this procedure); Antonio Regalado, *Who Gets Health Care? Rationing in an Age of Rising Costs*, WALL ST. J., Sept. 18, 2003, at A1 (reporting that CMS took the unprecedented step of granting "new technology" status to the drug Xigris® (drotrecogin alfa), which authorized federal reimbursement for half of the cost of this expensive new treatment for sepsis).

84. *See* Lars Noah, *Ambivalent Commitments to Federalism in Controlling the Practice of Medicine*, 53 U. KAN. L. REV. 149, 165–66 (2004) (discussing Medicare's codified noninterference principle).

later concluded that new evidence cast doubt on the safety and efficacy of the medical intervention.[86]

Indeed, later in their article, Pearson et al. emphasized that CMS should use CED in only a fairly narrow range of circumstances: "ethical application of CED requires that clear criteria exist by which technologies can be identified as fitting into an evidentiary middle ground, one that might be called 'promising but unproven.'"[87] This caveat has far less to do with ethical than statutory constraints because, if "proven," then a new technology presumptively secures coverage, and "proof" has never before required the elimination of all residual uncertainty much less answers to questions entirely collateral to the value of an intervention in treating patients. Moreover, precisely because of the possibility of "underappreciated risks" with "promising but unproven" technologies,[88] CMS should endeavor to secure genuinely informed consent when using its CED policy. The authors also, however, attempted to justify extending CED to what might be called "proven but expensive" new technologies,[89]

85. *See, e.g.*, Stephanie Saul, *(Not So) Standard Procedure*, N.Y. TIMES, Dec. 17, 2008, at B1 (reporting uneven coverage decisions for CyberKnife, a device for treating prostate cancer, adding that "[t]he disparities result from a policy principle as old as Medicare itself, in which officials in Washington leave many reimbursement decisions to the discretion of 15 regional contractors around the country"); *id.* ("[O]ver the years, Medicare has resolved only about 300 such [new technology] questions with blanket national coverage rulings. Meanwhile, thousands of other coverage policies have been—and continue to be—decided region by region.").

86. *See, e.g.*, Gina Kolata, *A Study Revives a Debate on Arthritis Knee Surgery*, N.Y. TIMES, Sept. 11, 2008, at A19 (reporting that Medicare had stopped covering arthroscopic surgery for arthritis of the knee in 2003 after a study sponsored by Department of Veterans Affairs found no benefit); *see also* Denise Grady, *Studies Question Using Cement for Spine Fractures*, N.Y. TIMES, Aug. 6, 2009, at A18 ("Medicare had no national policy on vertebroplasty and had been letting states decide. They have been covering it."); *id.* ("Dr. Salive said Medicare had looked into the treatment in 2005 but found a lack of [RCTs]. . . . [I]t was too soon to tell whether the [latest negative] research would affect coverage.").

87. Pearson et al., *supra* note 70, at 990 ("This middle ground must not be so broad as to include almost any new technology . . . ."); *see also id.* at 988 ("The CMS designed CED as a coverage mechanism that could be used when promising evidence suggested that patients might benefit from a new technology, but additional evidence was needed to determine with confidence that the technology met Medicare's statutory standard for coverage . . . ."); *id.* at 990 ("[CED] is based on identifying technologies for which the supporting evidence is not strong enough to warrant an entitlement to unlimited coverage."). This suggests that CMS could not legitimately use CED solely for purposes of acquiring collateral information about older—or clearly proven newer—technologies.

88. *Id.* at 990 ("A full analysis of the ethics of CED must not ignore the possibility that CED could result in the premature and inflated use of technologies that have underappreciated risks for many patients.").

89. *See id.* ("If a new technology with 'promising' evidence would be extremely expensive if used widely without further definition of its true risks and benefits, further ethical weight would be added to the rationale for requiring evidence development as a

which aptly describes the latest CSP: genetic testing before use of warfarin does not pose any unresolved questions of safety or efficacy; instead, it confronts CMS with practical questions of affordability given uncertainties about incremental clinical utility.[90]

B. Confining the Inquiry to "Coercion" Strictly Construed

Pearson et al. use a cramped notion of coercion as existing only when threats of adverse consequences override the exercise of genuinely free choice. Citing Wertheimer's classic treatment of the subject,[91] and apparently unaware of his subsequent book devoted to the closely related matter of "exploitation,"[92] they explained that "[c]oercion occurs when a threat of some harm compels a person to act in a manner that he or she

condition of coverage."); *id.* ("[S]ince CED cannot for practical reasons be used as the coverage approach for all 'promising' technologies, cost considerations should play a valid and honest role in selecting which technologies should be prioritized for CED."); *see also* Kinney, *supra* note 82, at 1500 ("[O]ften there is a range of opinions on coverage of a medical technology depending on views of scientific evidence, costs and other factors. Ultimately, a coverage decision is a political decision that balances many factors. There really is no 'accurate' decision regarding a disputed coverage issue.").

90. *See supra* notes 17–19 and accompanying text. These are legitimate and difficult questions, and CED allows Medicare to take it slowly, but such concerns hardly justify a requirement for enrollment in clinical trials—instead, CMS candidly should admit that only a fraction of those interested in access to an expensive new technology can receive it, at least until additional information demonstrates its value. In other contexts requiring the allocation of scarce health care resources, providers have utilized various selection methods. *See* Ezekiel J. Emanuel & Alan Wertheimer, *Who Should Get Influenza Vaccine When Not All Can?*, 312 SCIENCE 854 (2006); Lawrence O. Gostin, *Medical Countermeasures for Pandemic Influenza: Ethics and the Law*, 295 JAMA 554 (2006); Lars Noah, *Triage in the Nation's Medicine Cabinet: The Puzzling Scarcity of Vaccines and Other Drugs*, 54 S.C. L. REV. 741, 754–57 (2003).

91. ALAN WERTHEIMER, COERCION (1987).

92. ALAN WERTHEIMER, EXPLOITATION (1996). In fact, Pearson's co-authors previously had discussed this work, though at times giving it an unduly narrow interpretation. *See* Jennifer S. Hawkins & Ezekiel J. Emanuel, *Clarifying Confusions About Coercion*, HASTINGS CTR. REP., Sept.–Oct. 2005, at 16, 19 & n.12; *see also* Franklin G. Miller & Howard Brody, *A Critique of Clinical Equipoise: Therapeutic Misconception in the Ethics of Clinical Trials*, HASTINGS CTR. REP., May–June 2003, at 19, 26 (noting "the core value of *protecting research participants from exploitation*"); *cf.* David B. Resnik, *Exploitation in Biomedical Research*, 24 THEORETICAL MED. 233, 236 (2003) ("[H]arm is not a necessary condition for exploitation because exploitation may occur without any harm to the exploitee. Exploitation can occur when the exploiter fails to show adequate respect for the dignity or autonomy of the exploitee."); *id.* at 242 ("[I]n recent debates about the ethics of human research, . . . many different authors have made the charge of exploitation without explaining what they mean by this word or how we should respond to this accusation."); *id.* at 252 ("[E]xploitative research may be still ethical under some circumstances. Indeed, it is likely that a great deal of biomedical research is minimally exploitative yet still morally justified."). As it happens, Wertheimer recently joined the NIH's Department of Bioethics.

would not otherwise choose."[93] Pearson et al. distinguished coercion from "compulsion" as if that marked an ethically relevant boundary line:

> It might be argued that CED is coercive because some patients might feel that to get the treatment they "need," they have no other option but to participate in research, although they would not choose to do so if insurance coverage did not require that participation. But feeling compelled to participate in research does not constitute coercion. Some patients with terminal cancer, because of their dire prognosis, may feel compelled to enter phase I research studies, but making such a choice, even though it is done under difficult circumstances and with limited options, does not constitute coercion. Coercion requires a person, organization, or policy to threaten specific "harm."[94]

They concluded their treatment of this issue as follows: "Patients contemplating their treatment options under CED may face tough choices and may feel compelled to participate in research, but they do not do so under a cloud of coercion."[95]

Even if not technically coercion, compulsion also seems to be worrisome,[96] at least where government policy aims to take advantage of

---

93. Pearson et al., *supra* note 70, at 989 (adding that "no published articles have addressed coercion in relation to insurance coverage"). "An example is that of a kidnapper demanding ransom. The kidnapped victim's family may be coerced into giving up money to avoid the threatened harm to their loved one." *Id.* So coercion to participate in research would arise only in entirely implausible circumstances such as where an investigator secures "consent" by threatening a subject with violence?! *Cf.* Hawkins & Emanuel, *supra* note 92, at 19 ("Researchers standardly make *offers* to potential subjects, not threats.").

94. Pearson et al., *supra* note 70, at 989 (endnote omitted). The authors elaborated as follows:

> It could be argued . . . that CED threatens patients with a specific harm—the withholding of unrestricted insurance coverage. If obtaining coverage for a new technology without any requirement for research participation is the "best" option for patients, any policy short of unfettered access might represent a harm and be coercive. However, this argument is unsound. The fact that patients might feel entitled to the "best" option does not mean that they are entitled.

Id. ("Because CED does not propose to deny coverage to a technology to which patients are entitled, it does not threaten them with any harm. Since there is no threat of harm, concerns that CED is coercive are mistaken or misplaced.").

95. *Id.*; *see also* Manish Agrawal & Ezekiel J. Emanuel, *Ethics of Phase 1 Oncology Studies: Reexamining the Arguments and the Data*, 290 JAMA 1075, 1080–81 (2003). *But cf.* D. Christian Addicott, *Regulating Research on the Terminally Ill: A Proposal for Heightened Safeguards*, 15 J. CONTEMP. HEALTH L. & POL'Y 479 (1999) (urging that such patients be treated as a vulnerable class); Goldner, *supra* note 73, at 130 n.414 ("[T]he fact that [RCT enrollment] may well be the only avenue for obtaining such a benefit [of access to otherwise unavailable treatment] could be viewed as a form of inherent coercion that would vitiate the voluntariness of any consent that might be obtained."); Jerry Menikoff, *The Vulnerability of the Very Sick*, 37 J.L. MED. & ETHICS 51 (2009); Brendan P. Minogue et al., *Individual Autonomy and the Double-Blind Controlled Experiment: The Case of Desperate Volunteers*, 20 J. MED. & PHIL. 43, 46–52 (1995).

96. *See* Paul S. Appelbaum et al., *Voluntariness of Consent to Research: A Conceptual Model*,

vulnerable patients' circumstances.[97] This feature (namely, state action) would serve to distinguish their cancer example. Although private sponsors of RCTs have no greater license to engage in nonconsensual research,[98] they have absolutely no incentive to drag out drug trials in order to maintain a pool of patients desperate to enroll.[99] Perhaps the FDA's entire system of licensure, which withholds approval until sponsors have undertaken adequate studies, creates the same pressure on patients anxious for early access,[100] though the nonavailability baseline in this context lacks

---

HASTINGS CTR. REP., Jan.–Feb. 2009, at 30, 34–36 (discussing problematic offers and pressures as distinct from threats); *see also id.* at 37 ("[V]oluntariness occurs along a spectrum. Since subjects more often than not offer multiple reasons for enrollment, drawing a line between voluntary and involuntary actions is often not easy."); *cf. id.* ("[A] threat to withhold the patient's Social Security disability check unless she agrees to enter a study is clearly unacceptable."). *See generally* Alan Wertheimer, *Remarks on Coercion and Exploitation*, 74 DENV. U. L. REV. 889, 890–92, 894 (1997); *id.* at 896 ("[I]n the final analysis I do not believe that much turns on whether we can legitimately say that one agreement or another is exploitative or coercive on some linguistically plausible account of these terms."); *id.* at 906 ("We can call A's offer coercive and/or exploitative, but such labels will not resolve that moral problem. Having said that, it does not follow that the best moral answer is always to allow A to propose and B to accept any proposal that would be advantageous to B and rational for B to accept."). Wertheimer offered almost two dozen brief illustrations, including one involving experimentation with prisoners, *see id.* at 895–96, which he then alluded to in the remainder of his article, *see id.* at 896–906. Wertheimer made only passing and largely equivocal subsequent references to the experimentation hypothetical, *see id.* at 900–05, and his conclusion expressed doubts about prohibiting such research, *see id.* at 906. For further discussion of existing limitations on experimentation with prisoners, see *infra* note 125.

97. *Cf.* Carnahan, *supra* note 22, at 261–62 (arguing that elderly patients may experience particular difficulties understanding consent forms and represent an especially vulnerable population); Robert L. Schwartz, *Informed Consent to Participation in Research Employing Elderly Human Subjects*, 1 J. CONTEMP. HEALTH L. & POL'Y 115, 126–27 (1985) (elaborating on the vulnerability of nursing home residents to coercion); Donna Shalala, *Protecting Research Subjects—What Must Be Done*, 343 NEW ENG. J. MED. 808, 808 (2000) (describing a "case of a woman in a nursing home who was allegedly forced to participate in a study under threat of expulsion from the home").

98. Apart from the threat of regulatory sanctions, private researchers who injure subjects may face tort claims. *See, e.g.*, Sharona Hoffman & Jessica Wilen Berg, *The Suitability of IRB Liability*, 67 U. PITT. L. REV. 365 (2005); Michelle M. Mello et al., *The Rise of Litigation in Human Subjects Research*, 139 ANNALS INTERNAL MED. 40 (2003); E. Haavi Morreim, *Litigation in Clinical Research: Malpractice Doctrines Versus Research Realities*, 32 J.L. MED. & ETHICS 474, 475, 479–80 (2004); Roger L. Jansson, Comment, *Researcher Liability for Negligence in Human Subject Research: Informed Consent and Researcher Malpractice Actions*, 78 WASH. L. REV. 229 (2003).

99. *See* Lars Noah, *Sham Petitioning as a Threat to the Integrity of the Regulatory Process*, 74 N.C. L. REV. 1, 2 n.3 (1995) (noting that even one month delay in market entry could cost a drug company $10 million).

100. *See* Richard A. Epstein, *The Erosion of Individual Autonomy in Medical Decisionmaking: Of the FDA and IRBs*, 96 GEO. L.J. 559, 579–80 (2008); Eugene Volokh, *Medical Self-Defense, Prohibited Experimental Therapies, and Payment for Organs*, 120 HARV. L. REV. 1813, 1830 n.81 (2007) ("[S]ociety would balk at a law that generally forced people to go into clinical trials,

the contingency of Medicare's supposed noncoverage baseline: Congress insisted that new drugs not reach the market unless and until adequate and well-controlled studies satisfied the FDA that the product was relatively safe and effective.[101]

If, however, the FDA unreasonably withheld approval in order to force sponsors to engage in additional studies,[102] then these two situations become harder to distinguish. No doubt the agency would defend itself by pointing out that (1) patients have no right of access to unapproved medical technologies,[103] and (2) patients may secure access by other means (for instance, from other countries).[104] Indeed, because of various exceptions adopted by the FDA over the last two decades under pressure from desperate acquired immunodeficiency syndrome (AIDS) and cancer patients, individuals have several ways of securing access to investigational products without enrolling in RCTs (and without having to pay retail when their health insurers invoke experimental exclusion clauses).[105] Would

---

and a law that forces people to go into clinical trials if they want access to the only possibly lifesaving drugs seems to be no less coercive."). As an illustration of the power of this inducement, when private research sponsors discontinue RCTs, subjects have brought litigation in an effort to secure continued access to investigational products. *See, e.g.*, Abney v. Amgen, Inc., 443 F.3d 540, 550–53 (6th Cir. 2006) (rejecting such claims); Dahl v. HEM Pharms. Corp., 7 F.3d 1399, 1404–05 (9th Cir. 1993) (holding that the plaintiffs had a contract claim entitling them to an additional one-year supply); *see also* Michael M. Grynbaum, *Judge Orders Drug Maker to Provide Experimental Treatment to Terminally Ill Teenager*, N.Y. TIMES, Aug. 21, 2008, at C3.

101. *See* 21 U.S.C. § 355 (2006); United States v. Rutherford, 442 U.S. 544, 557–58 (1979). Congress recently directed the FDA to give expanded access to unapproved drugs for individuals suffering from serious diseases, but only if doing so would not impair the conduct of preapproval clinical trials. *See* 21 U.S.C. § 360bbb(b)(3), (c)(5).

102. *Cf.* Lars Noah, *A Miscarriage in the Drug Approval Process?: Mifepristone Embroils the FDA in Abortion Politics*, 36 WAKE FOREST L. REV. 571, 577–78, 583–85 (2001) (discussing delays in the approval of RU-486); Lars Noah, *Treat Yourself: Is Self-Medication the Prescription for What Ails American Health Care?*, 19 HARV. J.L. & TECH. 359, 374–76 & n.91 (2006) (discussing FDA delays in approving the switch of emergency contraceptives to nonprescription status); Lawrence S. Makow, Note, *Medical Device Review at the Food and Drug Administration: Lessons from Magnetic Resonance Spectroscopy and Biliary Lithotripsy*, 46 STAN. L. REV. 709, 730–32 (1994) (objecting to the agency's demands for additional studies of devices used to treat gallstones).

103. *See* Abigail Alliance for Better Access to Developmental Drugs v. von Eschenbach, 495 F.3d 695 (D.C. Cir. 2007) (en banc); *see also* Jerry Menikoff, *Beyond* Abigail Alliance*: The Reality Behind the Right to Get Experimental Drugs*, 56 U. KAN. L. REV. 1045 (2008); Alissa Puckett, Comment, *The Proper Focus for FDA Regulations: Why the Fundamental Right to Self-Preservation Should Allow Terminally Ill Patients with No Treatment Options to Attempt to Save Their Lives*, 60 SMU L. REV. 635 (2007).

104. *See* Peter S. Reichertz & Melinda S. Friend, *Hiding Behind Agency Discretion: The Food and Drug Administration's Personal Use Drug Importation Policy*, 9 CORNELL J.L. & PUB. POL'Y 493, 501–02 (2000); Mary Pat Flaherty & Gilbert M. Gaul, *Millions of Americans Look Outside U.S. for Drugs*, WASH. POST, Oct. 23, 2003, at A1 (reporting that the agency largely fails to enforce the policy's limitations).

105. *See* 21 U.S.C. § 360bbb (2006); 21 C.F.R. § 312.34 (2009); Steven R. Salbu, *The*

either one of these rebuttals answer ethical objections to the FDA's hypothetical policy of delaying licensure of safe and effective products in order to ensure that desperate patients continue to enroll in ongoing RCTs?

CMS could, of course, simply deny coverage if unpersuaded by the available evidence (as Pearson et al. argued it would do if unable to make use of the CED option), which would leave proponents to sponsor additional research that eventually might change the agency's mind.[106] Alternatively, CMS could defer making a coverage determination and ask another federal agency to undertake additional research (indeed, the agency cited this power as giving it the statutory authority for the CSP policy).[107] In either case, the existing policy of covering incidental costs of Medicare beneficiaries enrolled as subjects would facilitate completion of such studies, but the cost of the investigational item or service would fall on the sponsor.[108] From the standpoint of beneficiaries seeking access to a new but not yet covered medical intervention, the incentives seem identical (enroll or pay in full); from the standpoint of CMS, especially if it underwrites the research, the outcome seems largely the same. Nonetheless, ethically these may not come to exactly the same thing insofar

---

FDA and Public Access to New Drugs: Appropriate Levels of Scrutiny in the Wake of HIV, AIDS, and the Diet Drug Debacle, 79 B.U. L. REV. 93, 113–21 (1999). Although it ultimately prevailed in the latest litigation challenging restrictions on access to investigational drugs, the FDA recently liberalized its rules. *See* Charging for Investigational Drugs Under an Investigational New Drug Application, 74 Fed. Reg. 40,872 (Aug. 13, 2009) (to be codified at 21 C.F.R. pt. 312); Expanded Access to Investigational Drugs for Treatment Use, 74 Fed. Reg. 40,900 (Aug. 13, 2009) (to be codified at 21 C.F.R. pts. 312, 316); *see also* Linda Katherine Leibfarth, Note, *Giving the Terminally Ill Their Due (Process): A Case for Expanded Access to Experimental Drugs Through the Political Process*, 61 VAND. L. REV. 1281 (2008); James P. Sikora, Note, *Providing Hope: Developing a Viable Regulatory Framework for Providing Terminally Ill Patients with Adequate Access to Investigational Drugs*, 70 U. PITT. L. REV. 191 (2008).

106. *See, e.g.*, Purva Patel, *The Word for Cyberonics Is No: Medicare Says It Won't Pay for Use of Device to Treat Depression*, HOUS. CHRON., May 5, 2007, at D1 (explaining that, although the FDA had approved the vagus nerve stimulator to treat both epilepsy and chronic depression, CMS declined to cover use of the device); Patrick Yoest, *Colon Scans Not Covered*, WALL ST. J., May 13, 2009, at D6 (reporting that CMS rejected coverage of virtual colonoscopies).

107. *See* Carnahan, *supra* note 22, at 241–42; *id.* at 269–70 (suggesting different ways of creating incentives for genuinely voluntary participation by Medicare beneficiaries in RCTs); *id.* at 269 ("CMS could also achieve its goal of generating additional data by enhancing its traditional relationship with the [Agency for Healthcare Research and Quality]."); *see also* Janet Adams et al., *Recruiting Older Adults for Clinical Trials*, 18 CONTROLLED CLINICAL TRIALS 14, 15 (1997); Thomas M. Vogt et al., *Recruitment of Elderly Volunteers for a Multicenter Clinical Trial: The SHEP Pilot Study*, 7 CONTROLLED CLINICAL TRIALS 118, 130–31 (1986) (disputing the suggestion that non-institutionalized elderly patients are difficult to recruit as subjects).

108. *See supra* notes 20–21. Of course, if CMS underwrites an RCT conducted by another agency, then it would pay for the investigational item or service as well.

as the pressure exerted on beneficiaries flows less directly from CMS.

C. Cheapening Bioethics as Legal Discourse (and Vice Versa)

One of the most striking features of the *JAMA* article by Pearson et al. has to do with its style of analysis (and tone) rather than its content. Instead of the aspirational (some would say vacuous[109]) treatment typical of bioethical issues,[110] the authors sound almost lawyerly, focusing on what seem like technicalities and semantics.[111] Conversely, as an exercise in legal analysis, their defense of the CSP policy comes across as entirely amateurish. As explained at length in the previous sections, the standard of consent to human research has more breadth than a narrow conception of coercion.

Whether understood primarily as a form of applied philosophy, an extension of professional ethics in medical practice, or as a subset of health

---

109. *See* Carl E. Schneider, *After Autonomy*, 41 WAKE FOREST L. REV. 411, 412–15, 439–40 (2006); Giles Scofield, Commentary, *The Wizard of Oughts*, 28 J.L. MED. & ETHICS 232, 233–35 (2000); Michael H. Shapiro, *Is Bioethics Broke?: On the Idea of Ethics and Law "Catching up" with Technology*, 33 IND. L. REV. 17 (1999); Sheryl Gay Stolberg, *Bioethicists Find Themselves the Ones Being Scrutinized*, N.Y. TIMES, Aug. 2, 2001, at A1 (explaining that just about anyone can call themselves a "bioethicist"); *see also* Noah, *Assisted Reproductive Technologies*, *supra* note 12, at 606 ("One could criticize some of the existing academic commentary as engaging in little more than bioethical parlor games."). *See generally* JONATHAN BARON, AGAINST BIOETHICS (2006); Larry R. Churchill, *Are We Professionals? A Critical Look at the Social Role of Bioethicists*, DAEDALUS, Fall 1999, at 253; Edward J. Imwinkelried, *Expert Testimony by Ethicists: What Should Be the Norm?*, 76 TEMP. L. REV. 91 (2003).

110. *See* Robert Gatter, *Walking the Talk of Trust in Human Subjects Research: The Challenge of Regulating Financial Conflicts of Interest*, 52 EMORY L.J. 327, 383–86, 388–89 (2003) (explaining that overly prescriptive rules may weaken the tendency of researchers to "concern themselves with the normative spirit of the law"); Jeffrey P. Kahn & Anna C. Mastroianni, Commentary, *Moving from Compliance to Conscience: Why We Can and Should Improve on the Ethics of Clinical Research*, 161 ARCHIVES INTERNAL MED. 925, 925 (2001) (warning that an undue emphasis on adherence to rules "can cause researchers to quickly lose sight of the point of research protections—the rights and interests of the subjects themselves—and the protection of subjects can quickly be lost in the shuffle of paperwork necessary to satisfy the letter, if not the spirit, of regulations"); *see also* Miller et al., *supra* note 72, at 1453–54, 1449 ("[E]ven under an ideal regulatory system, the ethics of clinical research will continue to depend significantly on the integrity of investigators.").

111. Somewhat ironically, a physician who left the post of Assistant Surgeon General in 2001—and who would have played an early role in formulating the CED policy—framed the debate in the following terms: "Lawyers, often representing the technology developers or 'denied' patients, have argued that coverage with evidence development policies are coercive, unfair, and illegal. Ethicists disagree . . . ." Douglas Kamerow, *Paying for Promising but Unproven Technologies*, 335 BRIT. MED. J. 965, 965 (2007) (simply citing the *JAMA* article by Pearson et al.). Actually, NIH's own Office of Human Research Protections (OHRP) had raised questions about the CED policy, pointing out that it would have to comply with the federal regulations governing research. *See* Tunis & Pearson, *supra* note 11, at 1227.

law,[112] bioethics typically attempts to resolve questions by reference to a set of core principles rather than by splitting hairs.[113] Even more pragmatic or skeptical strains of bioethics do not cavalierly trade away commitments to autonomy and beneficence.[114] For example, in response to calls for expanded exceptions to informed consent requirements for certain types of research,[115] commentators responded in just such a guarded fashion, urging that when in doubt we always err on the side of protecting human subjects.[116]

---

112. *See* Tom L. Beauchamp, *Does Ethical Theory Have a Future in Bioethics?*, 32 J.L. MED. & ETHICS 209, 216 (2004); Alexander Morgan Capron & Vicki Michel, *Law and Bioethics*, 27 LOY. L.A. L. REV. 25, 25–33 (1993); Edmund D. Pellegrino, *The Metamorphosis of Medical Ethics: A 30–Year Retrospective*, 269 JAMA 1158 (1993); Symposium, *Emerging Paradigms in Bioethics*, 69 IND. L.J. 945 (1994).

113. *See, e.g.*, Jerry Menikoff, *The Involuntary Research Subject*, 13 CAMBRIDGE Q. HEALTHCARE ETHICS 338, 340–44 (2004); *see also* ROGER B. DWORKIN, LIMITS: THE ROLE OF THE LAW IN BIOETHICAL DECISION MAKING 18 (1996) (criticizing the law's role in bioethics, and opining that "our [legal institutional] tools for dealing with social problems posed by rapid change in biology and medicine are limited at best"); Robert J. Levine, *Medical Ethics and Personal Doctors: Conflicts Between What We Teach and What We Want*, 13 AM. J.L. & MED. 351, 362 (1987) ("A focus on rights and rules . . . has a tendency to yield a 'minimalist ethics.'"); Lars Noah, *Deputizing Institutional Review Boards to Police (Audit?) Biomedical Research*, 25 J. LEGAL MED. 267 (2004) (objecting generally to the legalization of bioethics). *But cf.* Noah, *supra* note 5, at 1152–60 (splitting hairs in jest to make a point); Benedict Carey, *The Subject Is . . . Subjects*, N.Y. TIMES, June 15, 2004, at F1 (reporting that the American Psychological Association has urged the use of "participant" as a less impersonal term). Similar issues may arise in other contexts. *See* Steven R. Salbu, *Law and Conformity, Ethics and Conflict: The Trouble with Law-Based Conceptions of Ethics*, 68 IND. L.J. 101, 102, 130–31 (1992); *see also* LARS NOAH, LAW, MEDICINE, AND MEDICAL TECHNOLOGY: TEACHER'S MANUAL 43 (2d ed. 2007) ("[S]hould these ethical rules be construed in a lawyerly fashion or instead more capaciously to promote the broader purposes that presumably animate them?").

114. *See* Tom L. Beauchamp, *Principles and Other Emerging Paradigms in Bioethics*, 69 IND. L.J. 955, 962–66 (1994); Edmund D. Pellegrino, *Autonomy, Beneficence, and the Experimental Subject's Consent: A Response to Jay Katz*, 38 ST. LOUIS U. L.J. 55, 57–61 (1993); Susan M. Wolf, *Shifting Paradigms in Bioethics and Health Law: The Rise of a New Pragmatism*, 20 AM. J.L. & MED. 395, 396–99, 413–14 (1994). *See generally* TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS (6th ed. 2009).

115. *See, e.g.*, Robert D. Truog et al., *Is Informed Consent Always Necessary for Randomized, Controlled Trials?*, 340 NEW ENG. J. MED. 804 (1999).

116. *See* Beverly Woodward, *Challenges to Human Subject Protections in US Medical Research*, 282 JAMA 1947, 1948, 1950–52 (1999); *see also* INST. OF MED., RESPONSIBLE RESEARCH: A SYSTEMS APPROACH TO PROTECTING RESEARCH PARTICIPANTS 6 (Daniel D. Federman et al. eds., 2003) ("The protection of research participants is fundamental and should remain paramount to any research endeavor."); Michael Baram, *Making Clinical Trials Safer for Human Subjects*, 27 AM. J.L. & MED. 253, 282 (2001) ("[W]e have drifted away from traditional regard for safeguarding humans in the process of testing and advancing a new technology . . . . [N]o outcomes justify degrading the process to the point where humans are viewed as expendable resources."); Goldner, *supra* note 73, at 125 ("It may well be the case that the effect of providing such information would be that the patient may refuse to

After recognizing that the boundary between quality improvement and research has "great practical importance" under federal regulations,[117] Pearson et al. proceeded under the mistaken assumption that those regulations only prohibit coercive research. "To be sure, patients seeking access to the new technology covered by insurance only under CED have an *inducement* to participate in research; however, this is no different than seeking access to an experimental treatment only available in clinical trials."[118] Because the federal regulations require that researchers "minimize the *possibility* of coercion *or undue influence*,"[119] the authors' implicit second claim (namely, that the undoubted possibility of influence does not rise to the level of "undue") requires more careful consideration.

Although HHS did not elaborate on what it meant by the phrase "undue influence," it had borrowed this language from the well-known Belmont Report,[120] which offered the following further explanation: undue influence may occur "through an offer of an excessive, unwarranted, inappropriate or improper reward or other overture in order to obtain compliance," adding that "inducements that would ordinarily be acceptable may become undue influences if the subject is especially vulnerable."[121] The HHS regulations add that consent forms must include "[a] statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject

---

participate . . . , choosing instead to be treated with the preferred treatment method off-protocol. This, however, is not an unreasonable price to pay for respecting the individual's autonomy.").

117. *See* Pearson et al., *supra* note 70, at 989 ("If a data-gathering process is considered research, federal regulations require that a variety of procedures must be followed to protect the participants involved.").

118. *Id.* (emphasis added).

119. 45 C.F.R. § 46.116 (2009) (emphasis added); *see also id.* § 46.101(a)(1) ("Research that is conducted or supported by a federal department or agency . . . must comply with all sections of this policy."); Goldner, *supra* note 73, at 128 ("[I]t has been understood that the possibility of coercion or undue influence is a major concern in the solicitation of subjects to participate in research protocols.").

120. *See* Protection of Human Subjects; Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, Report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral, 44 Fed. Reg. 23,192 (Apr. 18, 1979).

121. *Id.* at 23,195; *see also id.* at 23,197 (explaining that vulnerable subjects "are easy to manipulate as a result of their illness or socioeconomic condition"); *id.* at 23,195 ("[I]t is impossible to state precisely where justifiable persuasion ends and undue influence begins. But undue influence would include . . . threatening to withdraw health services to which an individual would otherwise be entitled."); COUNCIL FOR INT'L ORGS. OF MED. SCI., INTERNATIONAL ETHICAL GUIDELINES FOR BIOMEDICAL RESEARCH INVOLVING HUMAN SUBJECTS (2002), http://www.cioms.ch/frame_guidelines_nov_2002.htm ("Intimidation in any form invalidates informed consent.").

may discontinue participation at any time."[122] Thus, the federal rules do not concern themselves only with the use of threats to secure the consent of research subjects.

At least one commentator has argued that offering to pay for an item or service only if a patient enrolls in a clinical trial would violate these federal regulations:

> A decision to participate in medical research cannot be truly voluntary, however, when participation is the only way to receive the service. This is particularly troublesome in light of the fact that the particular intervention has likely already been FDA-approved as safe and effective, deemed appropriate for the patient by the patient's treating physician, and considered by CMS to be sufficiently reasonable and necessary to be approved for Medicare beneficiaries, but only so long as they agree to participate in research.[123]

As noted earlier, Pearson et al. responded that such inducements in no way differ from other RCTs.[124] Indeed, even if patients have access to therapeutic substitutes, lack of insurance coverage and limited personal resources may prompt them to enroll in clinical trials as the only hope for accessing medical care.[125] If offers of free access to treatment never create the possibility of undue influence, then only excessive bonus payments

---

122. 45 C.F.R. § 46.116(a)(8); *see also* Emanuel et al., *supra* note 67, at 2707 ("[R]espect includes permitting subjects to change their mind . . . and to withdraw without penalty."); Goldner, *supra* note 73, at 128 ("A long-standing principle of informed consent to research mandates an absolute right of a subject both to refuse to participate in research and to withdraw from it once involvement has commenced.").

123. Carnahan, *supra* note 22, at 264–65 ("CMS may be engaging in coercion or undue influence in violation of federal regulations in the sense that coverage of the service is essentially the patient's reward for enrolling in the trial.").

124. *See supra* note 94 and accompanying text; *see also* Allen L. Gifford et al., *Participation in Research and Access to Experimental Treatments by HIV-Infected Patients*, 346 NEW ENG. J. MED. 1373, 1373, 1376 (2002); Sarah Hewlett, *Consent to Clinical Research—Adequately Voluntary or Substantially Influenced?*, 22 J. MED. ETHICS 232 (1996).

125. *See* Gina Kolata & Kurt Eichenwald, *For the Uninsured, Drug Trials Are Health Care*, N.Y. TIMES, June 22, 1999, at A1; *see also* Christine Grady, *Vulnerability in Research: Individuals with Limited Financial and/or Social Resources*, 37 J.L. MED. & ETHICS 19 (2009). Situational vulnerability (to coercion rather than deception) explains the general prohibition on research with prisoners to guard against the possibility that subjects would enroll in the hope of securing early release or other favorable treatment. *See* 45 C.F.R. § 46.302 (2009) (recognizing that "prisoners may be under constraints because of their incarceration which could affect their ability to make a truly voluntary and uncoerced decision whether or not to participate as subjects in research"); Rachel Wener, Comment, *Not Situated to Exercise Free Power of Choice: Human Subject Research in Prison Settings*, 26 TEMP. J. SCI. TECH. & ENVTL. L. 365, 379–83 (2007); *see also* Sydney P. Freedberg, *Questions Raised over AIDS Research on Inmates*, ST. PETE. TIMES, Mar. 19, 2000, at 1A; Mike Ward & Bill Bishop, *Becoming Guinea Pigs to Avoid Poor Prison Care: Ill Inmates Urge Each Other to Join Experiments*, AUSTIN AM.-STATESMAN, Dec. 17, 2001, at A1.

would raise any concerns.[126] One of Pearson's co-authors previously had argued forcefully against even this view, however, suggesting that payment for research participation ethically differs in no way from wages offered to employees, including compensation for jobs that may pose risks.[127]

Pearson et al. rely heavily on a variant of the argument that the greater power (here to deny coverage altogether) includes the lesser power (here to offer coverage subject to conditions).[128] In other contexts, this type of legal reasoning has fared poorly. For instance, the "unconstitutional conditions" doctrine asks whether the government inappropriately demands that an individual forego exercising a constitutionally protected right in order to secure access to a benefit. It represents a reaction to the now generally

126. *See generally* Neal Dickert & Christine Grady, *What's the Price of a Research Subject? Approaches to Payment for Research Participation*, 341 NEW ENG. J. MED. 198 (1999); Christine Grady, *Money for Research Participation: Does It Jeopardize Informed Consent?*, AM. J. BIOETHICS, Spr. 2001, at 40; Carl Elliott, *Guinea-Pigging*, NEW YORKER, Jan. 7, 2008, at 36.

127. *See* Ezekiel J. Emanuel, *Ending Concerns About Undue Inducement*, 32 J.L. MED. & ETHICS 100 *passim* (2004). Emanuel went so far as to suggest that undue inducement would never occur in clinical trials. *See id.* at 104 ("We need to stop talking about undue inducement in clinical research."); *id.* at 102 ("Because independent review of clinical research excludes trials exposing participants to excessive discomforts and risks, undue inducement plays no role in clinical research."); *see also id.* at 100 ("[C]laims of undue inducement . . . should be treated with skepticism, placing a heavy burden of proof on those advancing such charges."). This conclusion depends on a remarkably anemic definition of the concept (and without any evident attention paid to the surrounding language in the regulation much less the fact that the cited rule actually used the somewhat broader term "influence" rather than "inducement"). *See id.* at 101 ("Absent potentially serious adverse consequences of the bad judgment there is no undue inducement. . . . These characteristics differentiate undue inducement from coercion and exploitation, with which it is frequently conflated." (footnote omitted)); *see also id.* at 103 ("The charge of undue inducement may be surreptitious paternalism by risk-averse individuals over decisions properly left to autonomous individuals."). It also seemingly ignores the emphasis on ensuring voluntariness, reducing "informed consent" to a simple question of adequate disclosure. *See id.* at 103 (discussing concerns that high inducements might lead to poor comprehension). Volitional impairment in a domain where we want only genuine volunteers to participate fundamentally distinguishes medical research from the employment setting that Emanuel chooses as his ethical benchmark. *Cf. id.* at 102 ("How can it be reasonable to invite people to enroll in a particular trial for no money, but unreasonable—even unethical—to invite them to enroll in the same trial for $100, $1,000, or even $10,000?"). For a different set of responses to Emanuel's position, see Joan McGregor, *"Undue Inducement[*"*] as Coercive Offers*, AM. J. BIOETHICS, Sept.–Oct. 2005, at 24.

128. Their subsequently published article did so explicitly. *See* Miller & Pearson, *supra* note 70, at 748 ("[O]ffers of benefit may come legitimately with strings attached—that is, with conditions that one would not choose apart from the desire to receive the offered benefit."); *id.* ("For example, government institutions may offer to pay medical tuition in exchange for a specified period of family practice in a rural community or medical service in the military."). The CSP policy does not, of course, offer a simple monetary bonus to beneficiaries who volunteer to participate.

discredited distinction between rights and privileges,[129] and the often associated premise that the government's greater power not to bestow a privilege at all includes the lesser power to provide that privilege conditionally.[130]

At its base, the unconstitutional conditions doctrine attempts to identify situations where the government has impermissibly pressured a beneficiary to relinquish a constitutional right. Narrowly conceived, coercion exists only if a person is put to a choice involving an unlawful option,[131] but coercion arguably also exists where a choice leaves the person worse off than they were previously.[132] In the typical unconstitutional conditions challenge, however, the government has offered to make a person better off in a tangible sense than they were previously, and it does not force them to accept a benefit conditioned on the waiver of rights. Instead, the doctrine recognizes that, even without coercion, persons often face seriously constrained choices and that the government's offer may encourage waivers of their rights without valid consent. "Exploitation" (or "manipulation") may be more apt a term than coercion.[133]

---

129. *See* William W. Van Alstyne, *The Demise of the Right-Privilege Distinction in Constitutional Law*, 81 HARV. L. REV. 1439, 1442 (1968); *cf.* Rodney A. Smolla, *The Reemergence of the Right-Privilege Distinction in Constitutional Law: The Price of Protesting Too Much*, 35 STAN. L. REV. 69, 69 (1982) ("[T]he doctrine has shown an uncanny ability to reconstitute itself in spite of the best efforts of scholars and jurists to bury it.").

130. *See* Brooks R. Fudenberg, *Unconstitutional Conditions and Greater Powers: A Separability Approach*, 43 UCLA L. REV. 371, 519 (1995) (concluding that, although the greater-includes-the-lesser argument makes some sense, heightened judicial scrutiny is appropriate in those cases where a lesser power is separated from the greater power along a constitutionally protected dimension); John H. Garvey, *The Powers and the Duties of Government*, 26 SAN DIEGO L. REV. 209, 215–19 (1989) (discussing the limitations of this argument); Michael Herz, *Justice Byron White and the Argument That the Greater Includes the Lesser*, 1994 BYU L. REV. 227, 238–49 (same).

131. *See supra* Part II.B; *see also* WERTHEIMER, *supra* note 91, at 202–21; Daniel Lyons, *Welcome Threats and Coercive Offers*, 50 PHIL. 425, 436 (1975); Jeffrie G. Murphy, *Consent, Coercion, and Hard Choices*, 67 VA. L. REV. 79, 83 (1981) (responding to "the mistaken assimilation of all hard decisions made under pressure of grim alternatives to cases of duress or coercion").

132. *See* Robert Nozick, *Coercion*, *in* PHILOSOPHY, SCIENCE, AND METHOD 440, 447 (Sidney Morgenbesser et al. eds., 1969) (arguing that coercion exists when threatened action would make one worse off than they "would have been in the normal or natural or expected course of events"); *see also* Peter Westen, *"Freedom" and "Coercion"—Virtue Words and Vice Words*, 1985 DUKE L.J. 541, 558–93; David Zimmerman, *Coercive Wage Offers*, 10 PHIL. & PUB. AFF. 121, 124–38 (1981); *cf.* James Lindgren, *Unraveling the Paradox of Blackmail*, 84 COLUM. L. REV. 670, 701–04 (1984) (explaining that blackmail is treated as coercion even though the threatened act—disclosure of damaging but truthful information about the victim—is not considered unlawful).

133. *See, e.g.*, FADEN & BEAUCHAMP, *supra* note 65, at 258–60, 354–62; JOEL FEINBERG, HARM TO SELF 242–49 (1986) (explaining that exploitation exists where one party takes advantage of another's weakness or dependency); *see also* WERTHEIMER, *supra* note 92, at

Because of its wildly inconsistent application by the Supreme Court, the unconstitutional conditions doctrine has attracted its fair share of scholarly attention. A number of competing formulations have been suggested by commentators,[134] including one that attempts to distinguish "threats" from "offers" by reference to some baseline,[135] or one that identifies situations where the government appears to be exercising monopoly power,[136] but the Supreme Court has not explicitly embraced any of these approaches.[137] Whether or not a Medicare beneficiary successfully could assail the CSP policy as an unconstitutional condition,[138] the doctrine offers instructive

123–57 (calling unconstitutional conditions a form of exploitation); *cf.* Richard R.W. Fields, Comment, *Perks for Prisoners Who Pray: Using the Coercion Test to Decide Establishment Clause Challenges to Faith-Based Prison Units*, 2005 U. CHI. LEGAL F. 541, 558–67.

134. *See* Kathleen M. Sullivan, *Unconstitutional Conditions*, 102 HARV. L. REV. 1413 (1989) (canvassing several competing theories based on notions of coercion, corruption, and commodification, and offering instead a "systemic" theory calling for strict scrutiny of rights-pressuring conditions on government benefits because they skew the distribution of power between and among the government and governed). "[U]nconstitutional conditions doctrine responds to a constant fear that government will tend to use the strategic manipulation of gratuitous benefits to aggrandize public power." *Id.* at 1493.

135. *See* Seth F. Kreimer, *Allocational Sanctions: The Problem of Negative Rights in a Positive State*, 132 U. PA. L. REV. 1293, 1352 (1984) ("[T]he distinction between liberty-expanding offers and liberty-reducing threats turns on the establishment of an acceptable baseline against which to measure a person's position after imposition of an allocation."). Kreimer proposed that history, equality, and prediction serve as relevant baselines. *See id.* at 1359–74; *see also* Kenneth W. Simons, *Offers, Threats, and Unconstitutional Conditions*, 26 SAN DIEGO L. REV. 289, 311–17 & n.78 (1989) (rejecting history and equality in favor of a modified predictive baseline).

136. *See* Richard A. Epstein, *The Supreme Court, 1987 Term—Foreword: Unconstitutional Conditions, State Power, and the Limits of Consent*, 102 HARV. L. REV. 4, 102 (1988) (concluding that "the traditional norms prohibiting coercion and duress are insufficient to police the legal monopoly that government exercises over certain critical domains"); *see also* RICHARD A. EPSTEIN, BARGAINING WITH THE STATE 312 (1993) (concluding that "a government that has any level of monopoly power cannot be trusted to impose whatever conditions it wants").

137. *See* Lynn A. Baker, *The Prices of Rights: Toward a Positive Theory of Unconstitutional Conditions*, 75 CORNELL L. REV. 1185, 1195 (1990) (noting that all commentators concede that "the Court has yet to arrive, explicitly or implicitly, at a clear limiting principle for deciding challenges to conditions on government benefits"); Cass R. Sunstein, *Is There an Unconstitutional Conditions Doctrine?*, 26 SAN DIEGO L. REV. 337, 338 (1989) ("Whether a condition is permissible is a function of the particular constitutional provision at issue; on that score, anything so general as an unconstitutional conditions doctrine is likely to be quite unhelpful."). For recent reviews of this subject, see Mitchell N. Berman, *Coercion Without Baselines: Unconstitutional Conditions in Three Dimensions*, 90 GEO. L.J. 1 (2001); Daniel A. Farber, *Another View of the Quagmire: Unconstitutional Conditions and Contract Theory*, 33 FLA. ST. U. L. REV. 913 (2006).

138. *Cf.* Mem. Hosp. v. Maricopa County, 415 U.S. 250 (1974) (holding that a state cannot deny access to medical care because an otherwise eligible person had exercised a fundamental right to travel); Heather S. Dixon, *Pelvic Exam Prerequisite to Hormonal Contraceptives: Unjustified Infringement on Constitutional Rights, Governmental Coercion, and Bad Public*

insights for the ethical debate: semantic quibbles should not distract from efforts to judge the acceptability of conditions on public health insurance coverage that would obligate patients to "volunteer" for research, and the various indignities that come with it,[139] in order to secure access to a needed item or service.

D. Undervaluing Autonomy: Communitarian Research Ethics

Perhaps the most stunning and potentially radical justification for CSP appears in the final paragraph of the *JAMA* article by Pearson et al. when they invoked society's "interest in greater knowledge" as a relevant factor "[i]n assessing the ethics of CED."[140] As they elaborated: "Patients who share in the benefits of society, and who ask for society to pay for these benefits, arguably should share a willingness to contribute to the body of evidence that will improve the quality and value of the health care of tomorrow."[141] In our autonomy-based tradition of bioethics,[142] such an

---

Policy, 27 HARV. WOMEN'S L.J. 177, 209–17 (2004) (arguing that publicly funded family planning clinics cannot condition access to oral contraceptives on intrusive exams that serve only collateral purposes); *id.* at 231–32 (concluding that, while physicians should discuss risks and separately might encourage a pelvic exam, women retain the right to make an informed choice to use oral contraceptives without first undergoing such an exam); Andrew Zoltan, Comment, Jacobson *Revisited: Mandatory Polio Vaccination as an Unconstitutional Condition*, 13 GEO. MASON L. REV. 735 (2005) (arguing that, once an infectious disease such as smallpox has been eradicated, mandatory immunizations no longer serve a public health purpose and, if made a prerequisite for access to public education, would violate the unconstitutional conditions doctrine). *But cf.* Lars Noah, *Too High a Price for Some Drugs?: The FDA Burdens Reproductive Choice*, 44 SAN DIEGO L. REV. 231, 253 (2007) ("What if the government demanded contraceptive use as a condition of Medicaid reimbursement for drugs that create a risk of birth defects?"); *id.* at 254 (concluding that this "looks more like a nonsubsidy than a penalty because a woman receiving public assistance for drug coverage would remain free (in theory) to refuse contraception and pay for the [teratogenic] drug out of pocket").

139. *See supra* note 73. Individuals enjoy rights of bodily integrity that would allow them to decline unwanted medical interventions unless the state had some powerful justification. *See* Washington v. Glucksberg, 521 U.S. 702, 720, 724–25 (1997); *In re* Cincinnati Radiation Litig., 874 F. Supp. 796, 810–14 (S.D. Ohio 1995) (situating a subject's right to bodily integrity in substantive due process); *see also* Ken Marcus Gatter, *Protecting Patient-Doctor Discourse: Informed Consent and Deliberative Autonomy*, 78 OR. L. REV. 941, 961–82 (1999).

140. Pearson et al., *supra* note 70, at 990.

141. *Id.* Medicare beneficiaries who had made substantial contributions through separate payroll taxes would quibble with any suggestion that the program represents nothing more than government largesse, just as taxpayers who have underwritten NIH and other publicly-funded biomedical research would take issue with insinuations that they have callously free-rided on the unselfish efforts of others. *Cf.* Claude Lenfant, *Clinical Research to Clinical Practice—Lost in Translation?*, 349 NEW ENG. J. MED. 868, 868 (2003) (noting that NIH had received more than $250 billion in appropriations since 1950).

142. *See* Carl E. Schneider, *Bioethics with a Human Face*, 69 IND. L.J. 1075, 1085 (1994) (calling autonomy "the centerpiece of bioethics"). Other countries do not share our perhaps excessive preoccupation with autonomy. *See* George J. Annas & Frances H. Miller, *The*

invocation of the "greater good" would set off alarm bells.[143] It suggests a distinctly public health approach to resolving questions about human subjects protection,[144] which may make perfect sense for the types of outcomes research envisioned by the CAD policy but becomes far more troubling when extended to RCTs under the CSP policy. Indeed, if persuaded by this notion of a quid pro quo, then why not insist that all Medicare (and Medicaid) beneficiaries sign up for at least one RCT, whether or not they want access to a novel and expensive intervention?

A few commentators have offered suggestions that nicely illustrate where such an approach might take us. One scholar recently floated the idea of compulsory research service.[145] Although it has become increasingly

Empire of Death: How Culture and Economics Affect Informed Consent in the U.S., the U.K., and Japan, 20 AM. J.L. & MED. 357, 377 (1994); *id*. at 373–75 (focusing on Japan).

143. During the two decades after World War II, and notwithstanding issuance of the Nuremberg Code, biomedical researchers in this country acted quite freely—utilitarianism prevailed over the more protective norms that only emerged in the late 1960s after revelations of domestic research abuses. *See* DAVID J. ROTHMAN, STRANGERS AT THE BEDSIDE: A HISTORY OF HOW LAW AND BIOETHICS TRANSFORMED MEDICAL DECISION MAKING 51 (2d ed. 2003) ("Utilitarian justifications that had flourished under conditions of combat and conscription persisted, and principles of consent and voluntary participation were often disregarded."); *see also id*. at 37 (explaining that malaria experiments conducted by the U.S. military on prisoners were lauded at the time as promoting the war effort).

144. *See* Scott Burris et al., *Applying the Common Rule to Public Health Agencies: Questions and Tentative Answers About a Separate Regulatory Regime*, 31 J.L. MED. & ETHICS 638, 643–46 (2003) (contrasting the ethical issues that arise in biomedical research and public health investigation); *id*. at 645 ("If in Common Rule practice autonomy is a trump, or at every fork directs research and practice activity down the more autonomy-enhancing path regardless of other considerations, then there is a tension with public health."); *id*. at 638 ("A nascent public health ethics movement has articulated ethical approaches that differ with those that generated the Common Rule."); Daniel Callahan & Bruce Jennings, *Ethics and Public Health: Forging a Strong Relationship*, 92 AM. J. PUB. HEALTH 169, 170 (2002) (referring to "the predominant orientation in favor of civil liberties and individual autonomy that one finds in bioethics, as opposed to the utilitarian, paternalistic, and communitarian orientation that have marked the field of public health throughout its history"); Nancy E. Kass, *An Ethics Framework for Public Health*, 91 AM. J. PUB. HEALTH 1776, 1777–78 (2001).

145. *See* Rosamond Rhodes, *In Defense of the Duty to Participate in Biomedical Research*, 8 AM. J. BIOETHICS 37 (2008). For a range of responses to her idea, see Robert J. Levine, Editorial, *Reflections on 'Rethinking Research Ethics*,*'* 5 AM. J. BIOETHICS 1 (2005) (introducing a symposium devoted to the topic). For earlier and generally milder versions of this proposal, see Arthur L. Caplan, *Is There a Duty to Serve as a Subject in Biomedical Research?*, IRB: REV. OF HUM. SUBJECTS RES., Sept.–Oct. 1984, at 1, 4–5; Goldner, *supra* note 73, at 124–25 (noting that "the argument has been made that, at least with respect to research involving only minimal risk, there is an ethical obligation of citizens to participate in such research"); John Harris, *Scientific Research Is a Moral Duty*, 31 J. MED. ETHICS 242 (2005). For further debate on this idea, see Iain Brassington, *John Harris' Argument for a Duty to Research*, 21 BIOETHICS 160 (2007); Sarah Chan & John Harris, *Free Riders and Pious Sons—Why Science Research Remains Obligatory*, 23 BIOETHICS 161 (2009).

difficult to recruit sufficient numbers of subjects for trials,[146] conscripting people for this purpose would represent a radical solution that likely no one would take seriously.[147] In the summer of 2009, a high-level NIH official made an urgent plea seeking more than two thousand adults willing to participate in clinical trials of experimental vaccines against the novel H1N1 ("swine") flu virus.[148] Given widespread fears about that emerging pandemic, researchers had little difficulty recruiting enough subjects; if, however, an insufficient number of people volunteered, public health agencies clearly would not—and should not—have the power to draft citizens into service as unwilling guinea pigs simply because this would serve the greater good.

David Orentlicher offered a more cautious variant of the conscription proposal. In response to the difficulties caused by underenrollment in RCTs, he would allow physicians to condition continued care on their patients' willingness to enroll in trials comparing established therapies.[149] Orentlicher conceded that such a recruitment strategy would raise objections about coercion,[150] but he emphasized that patients have no right

146. *See supra* note 1.

147. *See* Richard Delgado & Helen Leskovac, *Informed Consent in Human Experimentation: Bridging the Gap Between Ethical Thought and Current Practice*, 34 UCLA L. REV. 67, 120 n.228 (1986) ("[S]ociety has not yet decided that human subjects may be conscripted (like soldiers) without their consent."); *id.* at 94 ("When the research subject does not choose freely to participate, his act loses its moral meaning. Participation in the research is not something given by the subject; rather, it is something extracted."); Epstein, *supra* note 100, at 569 ("We are no longer worried about the prospect that individuals will be conscripted into medical trials against their will."); Hans Jonas, *Philosophical Reflections on Experimenting with Human Subjects*, 98 DAEDALUS 219, 234–35 (1969); Mortimer B. Lipsett, *On the Nature and Ethics of Phase I Oncology Trials of Cancer Chemotherapies*, 248 JAMA 941, 942 (1982) (distinguishing research participation from military conscription); Robert M. Veatch, *Which Grounds for Overriding Autonomy Are Legitimate?*, HASTINGS CTR. REP., Nov.–Dec. 1996, at 43 (warning that subjugating autonomy whenever it might promote the common good "would justify conscripting people into dangerous research against the will of subjects if the social benefits were great enough"). The bioethicists at NIH evidently would dismiss such objections as reflecting a bygone era. *See* Ezekiel J. Emanuel & Christine Grady, *Four Paradigms of Clinical Research and Research Oversight*, 16 CAMBRIDGE Q. HEALTHCARE ETHICS 82 (2007) (arguing that a communitarian-based paradigm has partially displaced the protectionist approach that prevailed in the 1970s and 1980s); *infra* note 155 and accompanying text (discussing a civic obligation to participate in research recently proposed by Emanuel and a couple of his other colleagues at NIH's Department of Bioethics).

148. *See* Donald G. McNeil, Jr., *Clinical Trials for Flu Vaccine Are to Begin Soon*, N.Y. TIMES, July 23, 2009, at A4.

149. *See* David Orentlicher, *Making Research a Requirement of Treatment: Why We Should Sometimes Let Doctors Pressure Patients to Participate in Research*, HASTINGS CTR. REP., Sept.–Oct. 2005, at 20, 21–22, 27. Pediatric oncologists routinely do something along these lines with experimental interventions. *See id.* at 23; Gina Kolata & Kurt Eichenwald, *In Pediatrics, a Lesson in Making Use of Experimental Procedures*, N.Y. TIMES, Oct. 3, 1999, § 1, at 40.

150. *See* Orentlicher, *supra* note 149, at 21, 23. Even the bioethicists at NIH apparently

to continuous treatment from a particular physician.[151] In that case, however, his proposed limitation to comparative efficacy trials seems unduly narrow.[152]

Orentlicher viewed his proposal as less extreme than Medicare's CSP policy,[153] but he also conceded that, in all likelihood, it would fail to comply with our existing—and, to his mind, overly protective—research subject protections.[154] If, however, Pearson et al. offered a persuasive defense of CSP, then Orentlicher's idea would not require any alteration in those ethical safeguards, and the still more radical conscription proposals would not seem as outlandish as most commentators seem to think. As it happens, in July 2009, a group of NIH bioethicists published a piece in *JAMA* arguing that all citizens have a civic—though not (yet) compulsory—obligation to participate in biomedical research.[155] Their provocative

would object. *See* Hawkins & Emanuel, *supra* note 92, at 19 ("[I]f a physician-researcher threatened to abandon a patient or withhold necessary standard treatment unless the patient joined a study, this would clearly be coercion.").

151. *See* Orentlicher, *supra* note 149, at 25–26; *cf. id.* at 22 (conceding that the termination of an existing doctor-patient relationship for a refusal to enroll might look more like a "penalty").

152. He thought that the minimal risks associated with comparative efficacy trials made his proposal more acceptable, *see id.* at 24–25, thereby suggesting that he harbored some lingering concerns about its potentially coercive nature. In addition, though he praised the societal value of such research, *see id.* at 21, Orentlicher resisted the temptation to rest his ethical defense of the proposal on utilitarian grounds, *see id.* at 24.

153. *See id.* at 22 (calling his proposal "more cautious" because "the CMS policy affects research on tests or treatments whose efficacy has not been established for the patients being studied").

154. *See id.* ("Under current practice, it is highly unlikely that [IRB] approval would be given to a study in which physicians made participation in the study a condition for receiving treatment."); *id.* at 22–23 (quoting directly relevant language from the Declaration of Helsinki); *see also id.* at 20 ("question[ing] whether research safeguards are sometimes overly protective"); *id.* at 27 (concluding that, in some cases, "ethical safeguards can become too strict").

155. *See* G. Owen Schaefer et al., *The Obligation to Participate in Biomedical Research*, 302 JAMA 67, 67–71 (2009) (rejecting the arguments offered by other proponents of this idea based on beneficence and free-riding, instead basing the obligation on the view that generalizable medical knowledge amounts to a public good for which all members of society should contribute their fair share); *id.* at 69 (drawing an analogy to expectations that academics occasionally agree to comment on manuscripts for peer-reviewed journals); *id.* (adding that it would resemble civic obligations such as voting rather than compulsory duties); *id.* (recognizing as legitimate excuses religious convictions, excessive burdensomeness, and significant risks, and emphasizing that informed consent would remain necessary); *id.* at 70 ("[E]ncouragement would have to be given carefully; there is a risk that the patients would fear abandonment by their physician if they refused to participate."); *see also id.* at 67 (disclosing their purpose "to stimulate support for a major cultural shift in the way physicians, researchers, patients, and society at large think about participation in research"); *id.* at 70 ("One strategy to affirm and reinforce the belief that individuals have an obligation to participate in research would be a publicity campaign analogous to get-out-the-

article represents a natural extension of the justifications that they previously had offered in defense of the CSP policy, and it comes perilously close to endorsing outright conscription.[156]

CONCLUSION

CMS has discovered a creative way to use its leverage over beneficiaries in order to generate useful information. It has done so in a manner that has more in common with the Pentagon's often heavy-handed approach to the use of investigational drugs than with the FDA's more subtle and indirect methods for encouraging the production of biomedical knowledge. The CSP policy appears to run afoul of federal research regulations, which only represent ethical minima in any event. Indeed, the agency's effort to skirt those regulations and justify its ethically dubious initiative rather than to steer well clear of existing restrictions itself sets a poor example for the broader research community.

During the last decade, NIH has attracted an impressive group of bioethicists who have produced a remarkable body of scholarly work. The legacy of their efforts will come to rival even the most influential reports produced in the past by federal commissions charged with providing the government with ethical guidance. Unlike members of these commissions, however, the bioethicists employed by NIH serve a client, and some of their published work defending federal initiatives bears a troubling resemblance to that produced by their professional counterparts in legal departments serving other agencies. Advocacy pieces produced by these bioethicists—serving as apologists for the work of their institutional employers—should draw sustained attention and, if necessary, serious rebuttals from independent scholars. Otherwise, the current bioethical party line could lead us down a worrisome path in society's relentless pursuit of biomedical advance.

---

vote efforts, which have helped convince 90% of US citizens that there is a duty to vote.").

156. *See id.* at 70 ("The situation is in some ways analogous to a wartime call to arms in which . . . soldiers to fight are needed."); *id.* at 71 ("[J]oining the army is more risky and time-consuming than any clinical trial that has been approved by a well-functioning institutional review board."); *see also* Schuck, *supra* note 65, at 924 ("The autonomy principle is deeply entrenched in our culture and law; few exceptions to it—compulsory immunization and military conscription are the major examples—have been recognized." (footnotes omitted)); *cf.* LARS NOAH, LAW, MEDICINE, AND MEDICAL TECHNOLOGY 176 (2d ed. 2007) ("Should participation in biomedical research be viewed as a civic duty akin to serving on a jury?!"); C.D. Herrera, *Universal Compulsory Service in Medical Research*, 24 THEORETICAL MED. 215, 223–25 (2003) (imagining a system that resembles jury duty).

THROUGH THE DOUGHNUT HOLE: REIMAGINING THE SOCIAL SECURITY CONTRIBUTION AND BENEFIT BASE LIMIT

PATRICIA E. DILLEY*

TABLE OF CONTENTS

* Professor of Law, University of Florida Levin College of Law. I am grateful to the University of Florida Levin College of Law for summer-research-grant support for this project. I thank my able research assistants, all students in the University of Florida Levin College of Law Graduate Tax Program—Elizabeth Faist, Josh Kendrick, Marla DaVee, Jay Riffkin, and Martina Nethery.

INTRODUCTION

One of the most peculiar popular reactions during the health care reform debate of 2009 was the repeated assertion from many senior citizens that they did not want "government-provided health care," accompanied in virtually the same breath by a vociferous warning to politicians to "keep your hands off my Medicare."[1] The perception that Medicare is something other than government-provided health care indicates the political strength of the earnings-based entitlement and contributory payroll tax financing for Medicare and Social Security. These beneficiaries clearly feel that Medicare coverage belongs to them—is something they worked for, is something they have some sort of ownership interest in, and is not really provided by the government.[2] This public sense of individual ownership does not attach to other government programs, whether it be national defense or the interstate highway system, despite the fact that all are supported by taxes paid by Americans in one setting or another.

This singular view of Social Security and Medicare might seem bizarre to tax analysts who consider the Social Security payroll tax, or Federal Insurance Contributions Act (FICA) tax, to be not just one source of overall government revenue but also an inequitable, or at least unwise, one, primarily because it applies the same tax rate to all levels of earnings, and to a lesser extent because it taxes only the lower part of those earnings—up to the contribution and benefit base (limited to $106,800 in earnings for 2010).[3] Given the uneven economic impact of the tax system supporting

---

1. *See, e.g.*, Carolyn Lochhead, *Overhaul Must Be Fully Funded, Not Worsen Deficits, Obama Says*, S.F. CHRON., July 25, 2009, at A1 ("Obama shot back against complaints that he is leading the nation to 'government-run health care,' calling such complaints part of an 'ancient ideological battle' that ignores the fact that 60 percent of Americans already receive government-provided health care under Medicare, Medicaid and veterans' benefits."); Bob Moos, *Medicare Changes Coming into Focus as Health Care Overhaul Moves Forward*, DALLASNEWS.COM, Oct. 18, 2009, http://www.dallasnews.com/sharedcontent/news/healthscience/stories/101809dnbusmedicare.4472819.html ("Ender said she was flabbergasted this summer when she heard seniors and others vehemently oppose any 'public option' insurance plan, but at the same time demand that lawmakers keep their hands off Medicare. 'Don't they realize that the traditional Medicare program *is* public insurance?' she said.").

2. *See, e.g.*, Posting of Bob Cesca to The Huffington Post, http://www.huffingtonpost.com/bob-cesca/get-your-goddamn-governme_b_252326.html (Aug. 5, 2009, 06:45 EST) ("Another argument I've heard, by the way, is that seniors and veterans have *earned* their socialist health care. To which I usually respond: *I see. So socialized health care is a reward for a job well done? Can I quote you?*").

3. *See, e.g.*, THOMAS L. HUNGERFORD, CONG. RESEARCH SERV., INCREASING THE SOCIAL SECURITY PAYROLL TAX BASE: OPTIONS AND EFFECTS ON TAX BURDENS (2009); MELISSA M. FAVREAULT & GORDON B. T. MERMIN, URBAN INST., ARE THERE OPPORTUNITIES TO INCREASE SOCIAL SECURITY PROGRESSIVITY DESPITE UNDERFUNDING? (2008); Thomas L. Hungerford, *How Increasing the Payroll Tax Base Affects*

these programs, it may seem odd that the public has not developed the same phobia toward the payroll tax that it seems to have toward the income tax, much the fairer tax by most standards. Most "non-expert" taxpayers have little complaint about the FICA tax rate being essentially a flat tax (although most would of course prefer it to be a lower rate); however, the wage base limit or "tax cap" is widely excoriated by taxpayers as extremely unfair. In their eyes, it is plainly unjust that the wealthiest taxpayers pay no more in FICA taxes than someone earning just at the wage base.[4]

This sense of unfairness has likely been exacerbated over the last two decades of stagnant real-wage growth, coupled with increasing income and wage inequality.[5] Each year, the FICA contribution and benefit base reflects a decreasing percentage of wages in the national economy, despite the automatic indexing provision that raises the base limit each year according to increases in average wages. The goal of indexing is to keep about 90% of wages in the economy within the limit and therefore part of the base for benefit accrual and FICA taxation, but the current provision has proved unequal to the challenge.[6]

The most recent worldwide economic crisis has created a new context for an old argument, as shrinking payrolls from higher unemployment are temporarily reducing near-term Social Security surpluses. Tax analysts, political commentators, and some members of Congress have variously called for payroll tax holidays to stimulate the economy, for different ways to finance Social Security that would lessen the tax pressure on employment, and ultimately, for wholesale reductions in program benefits

---

Tax Burdens, 115 TAX NOTES 643, 644–46 (2007); Martin J. McMahon, Jr., *The Matthew Effect and Federal Taxation*, 45 B.C. L. REV. 993, 1025–26 (2004).

4. Recent polling data indicates that 83% of Americans support elimination of the Social Security tax cap "so that workers earning more than [$106,800] would pay Social Security tax on their entire salary just like everyone else." VIRGINIA P. RENO & JONI LAVERY, NAT'L ACAD. OF SOC. INS., ECONOMIC CRISIS FUELS SUPPORT FOR SOCIAL SECURITY: AMERICANS' VIEWS ON SOCIAL SECURITY 13 (2009), http://www.nasi.org/sites/default/files/research/Economic_Crisis_Fuels_Support_for_Social_Security.pdf.

5. *See, e.g.*, Kyle Mudry & Justin Bryan, *Individual Income Tax Rates and Shares, 2006*, STAT. INCOME BULL., Winter 2009, at 5, 12, *available at* http://www.irs.gov/pub/irs-soi/09winbulinincome.pdf (stating that despite three years of tax rate increases, the rate paid by the top 1% decreased); Justin Bryan, *Individual Income Tax Returns, 2006*, STAT. INCOME BULL., Fall 2008, at 5, 10, *available at* http://www.irs.gov/pub/irs-soi/08fallbulintax.pdf (discussing that "[f]or 2006, average tax rates increased for each income category as incomes went up to AGI of $5 million or less," but that the average tax rate for income categories above $5 million decreased); Scott Hollenbeck & Maureen Keenan Kahr, *Ninety Years of Individual Income & Tax Statistics, 1916–2005*, STAT. INCOME BULL., Winter 2008, at 136, 144 tbl.1, *available at* http://www.irs.gov/pub/irssoi/16-05intax.pdf.

6. *See infra* notes 80–85.

to reduce long-term costs of Social Security.[7] Even before the depth and breadth of the financial crisis was fully realized, however, the issue of Social Security's possible long-term financing shortfall, and the use of wage-base-limit increases to address it, was raised during the presidential campaign of 2008.

As a candidate, President Obama suggested resolving at least part of the possible long-term financing problems for Social Security by raising the contribution and benefit base limit for some taxpayers.[8] Details are a bit fuzzy, but generally the idea was to raise or eliminate the base only for workers with earnings in excess of $250,000, thus creating a gap—a "doughnut hole," so to speak—of no additional FICA tax liability for workers with earnings above the current-indexed base—set at $106,800 for 2010—but below $250,000.[9] The political attraction of this proposal is

7. *See, e.g.*, Hendrik Hertzberg, *Not Insane*, NEW YORKER, Mar. 23, 2009, at 23–24, *available at* http://www.newyorker.com/talk/comment/2009/03/23/090323taco_talk_hertzberg (noting that pundit David Frum and Senate Minority Leader Mitch McConnell support a payroll-tax holiday); Kimberly Palmer, *David Walker Explains Social Security's Future*, U.S.NEWS.COM, June 16, 2009, http://www.usnews.com/money/personal-finance/articles/2009/06/16/david-walker-explains-social-securitys-future.html (quoting the former U.S. Comptroller General suggesting, "We should consider adding on top [of Social Security] a supplement—an automatic individual savings account.").

8. *See* Jason Furman & Austan Goolsbee, Op-Ed., *The Obama Tax Plan*, WALL ST. J., Aug. 14, 2008, at A13, *available at* http://online.wsj.com/article/SB121867201724238901.html (noting that Obama was considering plans that would ask individuals making over $250,000 to pay more in total payroll taxes); Senator Barack Obama, Remarks at the AARP Life@50+ National Expo (Sept. 6, 2008), http://www.barackobama.com/2008/09/06/remarks_of_senator_barack_obam_70.php [hereinafter Obama Remarks] (proposing to cut and eliminate taxes for working families and seniors); Foon Rhee, *Candidates Offer Social Security Plans*, BOSTON GLOBE, June 14, 2008, at A5, *available at* http://www.boston.com/news/nation/articles/2008/06/14/candidates_offer_social_security_plans/. The proposal was modified in mid-August of 2008 to take effect much later, in 2017, which matches the point at which Social Security expenditures are estimated to begin exceeding total yearly revenues. BD. OF TRUSTEES, FED. OLD-AGE AND SURVIVORS INS. AND FED. DISABILITY INS. TRUST FUNDS, THE 2008 ANNUAL REPORT OF THE BOARD OF TRUSTEES OF THE FEDERAL OLD-AGE AND SURVIVORS INSURANCE AND FEDERAL DISABILITY INSURANCE TRUST FUNDS 18 (2008), http://www.ssa.gov/OACT/TR/TR08/tr08.pdf [hereinafter 2008 BD. OF TRUSTEES REPORT]. The 2009 Trustees Report now indicates that expenses will exceed revenue beginning in 2016. BD. OF TRUSTEES, FED. OLD-AGE AND SURVIVORS INS. AND FED. DISABILITY INS. TRUST FUNDS, THE 2009 ANNUAL REPORT OF THE BOARD OF TRUSTEES OF THE FEDERAL OLD-AGE AND SURVIVORS INSURANCE AND FEDERAL DISABILITY INSURANCE TRUST FUNDS (2009), http://www.ssa.gov/OACT/TR/2009/tr09.pdf [hereinafter 2009 BD. OF TRUSTEES REPORT].

9. *See* Soc. Sec. Admin., Contribution and Benefit Base, http://www.ssa.gov/OACT/COLA/cbb.html (last visited Apr. 26. 2010). Interestingly, it appears that the "doughnut hole" feature originated with the campaign of John Edwards

fairly obvious—it addresses the aforementioned popular notion of the unfairness of the FICA "wage cap" and also fulfills the President's campaign promise to not raise taxes on anyone making less than $250,000 per year. Little has been heard about the proposal since the election, and it is unclear whether it will see the light of day in any future Obama Administration proposals for Social Security. Nonetheless, the suggestion provides an opening for examination of the contribution and benefit base with fresh eyes.

Arguments about the general concept of a tax base have most often focused on the notion of a "comprehensive tax base" and, more recently, on whether we should tax consumption as a base, rather than income.[10] Whatever the criteria for the "best" tax system are—efficiency, social welfare, distribution of tax burden, etc.—the debates frequently center on the question of the appropriate definition of the tax base: what exactly should be taxed, no matter what kind of rate is assessed? If income is to be taxed, the inevitable next inquiry is what constitutes income and how comprehensive that definition should be.[11] If consumption is to be taxed, the first inquiry, before addressing the appropriate rate level, is likely to be what kinds of consumption should be exempt in order to prevent the tax from imposing too onerous a burden on the most vulnerable segments of society.[12]

Much less attention has been paid in tax theory debates to the Social

---

when he was still in the race for the Democratic nomination. Teddy Davis, *Obama Floats Social Security Tax Hike*, ABCNEWS.COM, Sept. 22, 2007, http://abcnews.go.com/Politics/Story?id=3638710&page=1 (quoting Edwards as saying, "I do think we need to have a bubble above $97,000, probably up to about $200,000, so we don't raise taxes on middle-class families. . . . But, above the $200,000, these millionaires on Wall Street ought to be paying their Social Security taxes").

10. *See, e.g.*, Deborah A. Geier, *The Taxation of Income Available for Discretionary Use*, 25 VA. TAX REV. 765, 767 (2006) ("The signature tax policy tension of the last two decades (at least) has been whether the federal tax base ought to reach 'income' or only 'consumption'"); Edward J. McCaffery, *A New Understanding of Tax*, 103 MICH. L. REV. 807, 920–38 (2005) (arguing for a progressive postpaid consumption task as a way to implement a fair timing of taxation); Daniel Shaviro, *Beyond the Pro-Consumption Tax Consensus*, 60 STAN. L. REV. 745, 746–47 (2007) ("In the last ten or so years, [the tax policy debate] has increasingly come to denote instead replacing the income tax with a consumption tax.").

11. For a more recent discussion on the comprehensive tax base, see David A. Weisbach & Jacob Nussim, *The Integration of Tax and Spending Programs*, 113 YALE L.J. 955 (2004). As those authors note, several seminal works on the topic include Boris I. Bittker, *A "Comprehensive Tax Base" as a Goal of Income Tax Reform*, 80 HARV. L. REV. 925 (1967); R.A. Musgrave, *In Defense of an Income Concept*, 81 HARV. L. REV. 44 (1967); and Joseph A. Pechman, *Comprehensive Income Taxation: A Comment*, 81 HARV. L. REV. 63 (1967).

12. *See* McCaffery, *supra* note 10, at 812 (suggesting that a postpaid consumption tax is the "fairest and least arbitrary" tax system because it "burdens some but not all uses of capital and its yield, and for normatively attractive reasons").

Security contribution and benefit base, even though many American workers pay more in payroll taxes than income taxes each year.[13] True, economists and tax analysts have repeatedly criticized the FICA tax for its regressivity, but the range of analysis is limited mainly to the economic impact of a tax on wages, with less consideration of what the base itself should look like from any other policy perspective.[14] What is different about the Social Security payroll contributions and earnings-based benefit system that leads analysts to marginalize it in the bigger picture of tax analysis?

Three elements of the FICA contribution and benefit base in particular differentiate it from other tax bases: (1) the inclusion of only wages, not other sorts of income, in the base;[15] (2) the dollar limit that results in inclusion of less than 100% of all wages in the base; and (3) the use of the same base for both benefits earned and contributions paid. Perhaps the last element is the major reason the Social Security contribution and benefit base has remained essentially unchanged in structure for seventy years: it is part of a closed system that both requires revenues from a special levy to be dedicated to a single spending purpose and ties those expenditures to earnings recorded under the same limited tax system.[16] The FICA structure's purpose is not simply to raise revenue but also to provide a method of financing that echoes the values underlying the system for old-age income benefits: those who work for a lifetime are the ones who earn entitlement to benefits.

Payroll-tax financing has made Social Security's revenue flow less susceptible to political manipulation precisely because it is part of this closed system, which includes an internal savings mechanism in the form of

---

13. *See* ANDREW CHAMBERLAIN & GERALD PRANTE, TAX FOUND., WHICH TAXES WEIGH MOST HEAVILY ON AMERICANS WITH DIFFERENT INCOMES?, (2007), http://www.taxfoundation.org/files/ff80.pdf ("On average, federal payroll taxes per household actually outweighed personal income taxes in 2004—$7,069 per household compared to $7,062.").

14. *See, e.g.*, CITIZENS FOR TAX JUSTICE, AN ANALYSIS OF ELIMINATING THE CAP ON EARNINGS SUBJECT TO THE SOCIAL SECURITY TAX & RELATED ISSUES 3–4 (2006), http://www.ctj.org/pdf/socialsecuritytaxearningscapnov2006.pdf (suggesting expanding the Social Security tax base to adjusted gross income).

15. Of course, most dedicated taxes are applied to a specific kind of income and expenditure—for example, the Highway Trust Fund is financed by federal gasoline taxes. 26 U.S.C. § 9503 (2006).

16. *See, e.g.*, Charles Krauthammer, Op-Ed., *Forget 2042—The Real Crisis for Social Security Comes in 13 Years,* PITTSBURGH POST-GAZETTE, Feb. 19, 2005, at A-10 (stressing that the Social Security system is "pay-as-you-go"); Robert Novak, Op-Ed., *McCain Could Score Big with Payroll Tax Cut*, CHI. SUN-TIMES, Mar. 27, 2008, at 19 ("[T]he heavy payroll tax revenues not only provide enough money for Social Security but fund other programs, as well.").

yearly surpluses retained and held as dedicated government bonds to be used to pay benefits whenever yearly revenues may fall short.[17] The flip side of this stability, however, is that the entire system becomes the target of political attacks in any year payroll-tax revenues are projected to fall short of projected yearly benefit payments, even when trust-fund reserves are adequate for decades to bridge any financing gaps.[18] In addition, the unified contribution and benefit structure seems to restrict, and possibly distort, thinking on options for changing the system's financing to meet changing economic circumstances.[19]

Current calls for "reform" of Social Security have little to do with any generally perceived need for change in the way the program delivers benefits. Rather, the limitations of payroll tax financing and the current wage base have created an opening for budget hawks and longtime opponents of Social Security to argue that the program must be targeted for reductions to address the federal deficit.[20] I suggest that the limits of the current base and payroll-tax system do not mean that the very necessary benefit system is too expensive but rather that we should examine other ways to increase dedicated revenues to fully fund the system if the current system's revenues fall short at some point. The problem is not that benefits

17. *See, e.g.*, 42 U.S.C. § 401(a) (2006) (establishing a Federal Old-Age and Survivors Insurance Trust Fund); 42 U.S.C. § 1395i(a) (2006) (creating a Federal Hospital Insurance Trust Fund); 2009 BD. OF TRUSTEES REPORT, *supra* note 8, at 2 (reporting that the trust funds to support Social Security are adequately financed for the next twenty-five years, but that the combined funds are projected to become exhausted in 2037).

18. *See* Lori Montgomery, *Lawmakers Seeking Consensus on Social Security Overhaul*, WASH. POST, May 6, 2009, at A14 (describing potential negotiations between both political parties regarding major changes to Social Security).

19. *See* AARP PUB. POLICY INST., REFORM OPTIONS FOR SOCIAL SECURITY 6–7 (2008), http://assets.aarp.org/rgcenter/econ/i3_reform.pdf (discussing that trust-fund asset returns could be increased by investing in nongovernment securities.) *See also* Letter from David M. Walker, Comptroller General of the U.S., to Rep. Bill Thomas, Chairman, Comm. on Ways and Means, U.S. House of Representatives (May 6, 2005), http://www.gao.gov/new.items/d05649r.pdf (outlining options for social security reform, including tapping different revenue streams and increasing investment returns through broader investing and individual accounts).

20. The recent creation of the National Commission on Fiscal Responsibility and Reform by President Obama is seen by many as a result of pressure by long-time "deficit hawks" to focus on cuts in Social Security and Medicare as a primary way to address the national debt. *See, e.g.*, Posting of James Ridgeway to Mother Jones, http://motherjones.com/mojo/2010/02/obamas-stealth-entitlement-commission (Feb. 19, 2010, 00:33 PST); *see also* Posting of John D. McKinnon to Washington Wire, http://blogs.wsj.com/washwire/2010/02/18/left-and-right-take-aim-at-alan-simpson/ (Feb. 18, 2010, 18:20 EST) (discussing former Sen. Alan Simpson, named co-chair of the commission, and his history of supporting dramatic cuts in Social Security); Posting of Dean Baker to TPMCafé, http://tpmcafe.talkingpointsmemo.com/2010/02/17/alan_simpson_a_man_who_intensely_wants_to_cut_soci/ (Feb. 17, 2010, 04:41 EST).

are too generous—far from it—but rather that revenues are falling short of system needs.[21] In contrast to the problems of Medicare and the health system generally, the costs of the Social Security cash-benefit system have not exceeded expectations (the number of baby boomers reaching retirement age beginning in 2005 was essentially known from the time of their births, after all), but payroll-tax income has not kept pace with expenditures.[22]

If we are to think creatively about how to resolve any future financing issues for Social Security, it is critical to bear in mind that while the right to benefits is earned individually, benefits are paid for on a social basis. The amount of payroll taxes collected from or on behalf of any individual worker has nothing to do with her eventual benefit entitlement.[23] The payroll tax is a group-financing mechanism, not an individual investment or payment for individual benefits, and in fact is not the sole source of revenue for the program.[24] It is often overlooked in debates over Social Security's future financial path that the original designers did not contemplate a payroll tax as the primary financing mechanism at all, and certainly not once the program reached maturity. Indeed, substantial revenues from nonpayroll-tax sources have long been part of the total financing of the system.[25]

For example, the Social Security trust funds receive general income tax

---

21. For a clear demonstration of how low Social Security benefits are, see PATRICIA E. DILLEY, NAT'L ACAD. OF SOC. INS., RESTORING OLD AGE INCOME SECURITY FOR LOW-WAGE SINGLE WORKERS, (2009), http://www.nasi.org/sites/default/files/research/Patricia_Dilley_January_2009_Rockefeller.pdf.

22. The average Social Security benefit payable to retired workers in January 2009 was $1,153 per month. *See* Soc. Sec. Admin., Find an Answer to Your Question, http://ssa-custhelp.ssa.gov/cgi-bin/ssa.cfg/php/enduser/std_adp.php?p_faqid=310 (last visited May 8, 2010). For a discussion of the system's financing arc in the future, see 2009 BD. OF TRUSTEES REPORT, *supra* note 8, at 2 (predicting that the annual cost of Social Security will begin to exceed tax income in 2016 and that the system will become insolvent in 2037).

23. *See* 42 U.S.C. § 402(a) (2006) (stating that every fully insured individual who has attained age 62 and filed an application for benefits "shall be entitled to an old-age insurance benefit for each month" beginning with the first month in which that individual has reached retirement age and ending with the month preceding the month of his death); *id.* § 414(a) (defining the term *fully insured individual* as anyone with the required number of quarters of coverage, normally forty, by the time of application for benefits); *id.* § 413 (defining *quarter of coverage* for years before 1978 as a calendar quarter in which the individual was paid $50 or more in wages, and for years after 1977, "each portion of the total of the wages paid and self-employment income credited" that equals the amount required for a quarter of coverage that year).

24. SOC. SEC. ADMIN., SOCIAL SECURITY: UNDERSTANDING THE BENEFITS 6–9 (2010), *available at* http://www.ssa.gov/pubs/10024.pdf (explaining the breakdown of paying into the system and receiving benefits out of the system). *See* 2009 BD. OF TRUSTEES REPORT, *supra* note 817, at 37 (noting that financial securities produce another stream of revenue).

25. 2009 BD. OF TRUSTEES REPORT, *supra* note 8, at 37.

revenues representing the federal government's employer share of payroll taxes for federal employees covered by Social Security as well as the revenues realized from taxation of Social Security benefits received by higher income beneficiaries.[26] Both of these revenue sources represent a further socialization of the costs of Social Security over all taxpayers, unassociated with any individual taxpayer's benefit accrual. This is not to say the contributory principle is unimportant, but focusing on the difference between earning benefits individually and paying for them as a society may yield some fresh insights on how the concept of the contribution and benefit base might be productively redesigned.

The heart of the matter is the very nature and basis for entitlement to Social Security benefits—the prevailing assumption that workers become entitled to benefits because they pay for that coverage through payroll taxes is simply wrong, both as a matter of philosophical principle and of law. Entitlement to Social Security benefits is attained by working, not by paying taxes, in keeping with the program's fundamental premise that all those who work for most or all of their lives are entitled to at least basic income security in their old age. The distinction is not mere semantics: separating entitlement based on effort from financing needs is a critical step to developing more flexible and equitable solutions to future financing problems, beginning with a fresh look at the contribution and benefit base limit.

The President's wage-base proposal has created an opportunity for a wholesale reimagining of the base limit. Beyond any specific merits or drawbacks this proposal might have, it serves as a convenient starting point for an exploration of the notion of the contribution and benefit base primarily from a Social Security programmatic perspective. The President has recently reiterated his support for increases in the contribution and benefit base to resolve, in whole or in part, any long-term or short-term financing issues for Social Security, but there may be equally compelling reasons to raise the base or change its calculations whether or not Social Security ultimately requires additional financing.[27]

---

26. *See* 26 U.S.C. § 86(a)–(b) (2006) (providing for taxation of Social Security benefits for beneficiaries in designated higher income brackets); Social Security Act Amendments of 1983, Pub. L. No. 98-21, § 121(e), 97 Stat. 65, 83–84 (requiring that the amount of revenues equivalent to the "aggregate increase in tax liabilities" attributable to taxation of Social Security benefits be appropriated at least quarterly to the Social Security trust funds).

27. *See* SOC. SEC. ADMIN., UPDATE 2010 (2010), http://www.socialsecurity.gov/pubs/10003.pdf (showing that the 2010 wage base has not been stepped up from that of 2009). For the President's continued support for increasing the wage base to address long-term Social Security financing, see President Barack Obama, Townhall in Henderson, Nevada (Feb. 19, 2010), *in* REALCLEARPOLITICS.COM, http://www.realclearpolitics.com/articles/2010/02/19/obamas_townhall_in_henderson_n

Part I of this Article explores the original design of the Social Security contribution and benefit base as part of its benefit and financing system in an attempt to understand why a wage tax was chosen as the financing mechanism, why a limit was placed on the wages and earnings that would count, and why both benefits and taxes are tied to the same base. One reason for the muddled discussion of raising the wage base is perhaps the lack of understanding of its origins and function in Social Security; the frequent appeals for fidelity to "original principles" are too often based on either extremely sketchy knowledge of or thinly disguised hostility to the actual basic principles of social insurance generally and of the U.S. Social Security system in particular.

Part II discusses the current configuration of the contribution and benefit base, which is widely perceived as unfair by the people whose wages are entirely covered by it, in the context of Social Security program principles as well as of tax policy. The contribution and benefit base limit has been increased on an ad hoc basis many times throughout the history of the program—most recently by a schedule of increases enacted as part of the 1977 Social Security Amendments, the last of which occurred in 1992 (the schedule was accelerated in the 1983 Social Security Amendments).[28] Currently, the dollar limit of the base is indexed to the increase in average wages each year, so the question is whether there is any programmatic basis for increasing the dollar amount further, increasing it only for workers with earnings greater than a certain level, or eliminating the limit altogether. The importance of the contributory principle to the Social Security program is undeniable, but it is unclear that the present level of the base for both benefits and taxes is completely consistent with program goals or with at least some definitions of tax fairness.

Part III uses the doughnut hole proposal as a starting point for examination of possible alternatives to simply raising the contribution and benefit base beyond the indexed increases already provided under current

---

evada_104485.html. In response to a question on Social Security, the President suggests,

> So what we've said is, well, . . . doesn't it make sense to maybe have that payroll tax [base] cut off at a higher level, or have people—maybe you hold people harmless till they make $250,000 a year, but between $250,000 and a million or something, they start paying payroll tax again—just to make sure that the fund overall is solvent. . . . That's not the only way of fixing it, but if you made a slight adjustment like that, then Social Security would be there well into the future and it would be fine.

Id. A brief disclaimer: I will not be discussing the Medicare program, although I will address, for comparison's sake, the Medicare portion of the FICA tax (since 1984, 1.45% of the 7.65% total) and the elimination in 1993 of the wage base for that portion.

28. *See* Social Security Amendments of 1977, Pub. L. No. 95-216, § 101, 91 Stat. 1509, 1510–12 (increasing the tax rates to offset deficit); § 331, 91 Stat. at 1541–42 (reducing benefit increases); Social Security Amendments of 1983, Pub. L. No. 98-21, § 101, 97 Stat. 65, 67–70 (changing coverage for newly hired federal employees).

law. A host of enforcement and tax-equity issues are obviously raised by the idea of increasing or eliminating the wage base only for high-wage workers, but I focus mainly on the question of whether there is a coherent basis in Social Security program theory for anything like a doughnut hole wage-base configuration. Even if the proposal is never revived by the Obama Administration, examination of the reactions to it is useful in revealing both the political agendas and the lack of understanding of Social Security on the part of many of its critics. Ultimately, the primary objection to a doughnut hole structure may be practical rather than theoretical: Congress may have learned some hard lessons from the reactions to the legislated coverage gap in the Medicare prescription-drug plan about being careful when creating gaps in either tax or benefit structures that may have unintended consequences leading to uncertainty, anger, and gamesmanship behavior in taxpayers.[29] The distributional and political results may not be worth the accompanying static.

While the doughnut hole proposal might create as many problems as it would solve, it does suggest that the time may be right to redesign the concept of the wage base to raise additional revenue for Social Security and achieve the desired distributional results without invoking the notion of a gap in taxation at all. These solutions will require reexamination of actual, as opposed to politically distorted, fundamental principles of Social Security. This is not simply a question of efficient and fair tax policy: it is essential to consider how changing or eliminating the limit on the base, or decoupling the contribution base from the benefit base after a certain wage or income level, would affect the function and political viability of Social Security and the payroll tax. Payroll-tax financing, both the fixed rate and the automatically indexed wage base, provides political strength and certainty to Social Security but also imposes rigidity and lack of flexibility in the face of changing economic conditions.

Part IV analyzes some possibilities for reimagining and reformulating the current contribution and benefit base, mainly focusing on the notion of decoupling the tax base for financing purposes from the earnings base for benefit-accrual purposes while still retaining the contributory principle that has traditionally been the foundation of Social Security's widespread public support across income and class lines. While the contributory financing

29. *See, e.g.*, Robert Pear, *Medicare Beneficiaries Confused and Angry over Gap in Drug Coverage*, N.Y. TIMES, July 30, 2006, at A14, *available at* http://www.nytimes.com/2006/07/30/washington/30medicare.html?_r=1&ex=1154404800&en=584f50d7ee778a42&ei=5087%0A ("The gap, the notorious 'doughnut hole,' is upsetting many beneficiaries, and it has become a potent symbol as politicians debate the merits of the new program.").

system, as discussed below in Part I, is an integral element of the political economy of Social Security, there may be no compelling rationale for continuing to largely restrict the program's financing to payroll tax revenues.

Some critics of proposals to raise the base for contributions but not for benefit calculations have charged that such a separation would violate a fundamental principle of the earned entitlement of social insurance.[30] The same objection is also frequently raised against suggestions for partial financing of Social Security from non-FICA, general-tax-revenue sources, despite the fact that it already receives substantial nonpayroll-tax revenues. For reasons I discuss below, I think these criticisms are overstated and frequently based on a misunderstanding of the role of the contribution and benefit base in the Social Security program. The more serious objections to any separation of the base for benefits from the base for taxes are political, having to do with public support for the program, which may or may not be grounded in an accurate understanding of program principles. There are valid political as well as substantive policy points against increasing the wage base without increasing the benefit base, but it is not an open-and-shut case.

Much of the resistance to the idea of expanding partial general-tax-revenue financing is connected to the fear of the political consequences of loosening the bonds between contributory financing and the earned right to a benefit. The erroneous notion, exploited by conservative opponents of the program, that Social Security benefits are an individual investment rather than an earned right to a portion of the future productivity of society as a whole has hampered creative approaches to financing that would equalize the tax burden by requiring more from the upper-income taxpayers who benefit disproportionately from the economic and social stability that Social Security underwrites.

There is a strong case for leaving the wage base essentially unchanged for the time being, a decision President Obama appears to have reached during the campaign when he suggested a 2019 effective date for his base increase proposal.[31] The larger issue that hangs over any discussion of changes to the Social Security base for policymakers, if not for tax theorists, is whether or when the system will need additional payroll-tax revenue to

30. *See, e.g.*, Mickey Kaus, *Obama's "Mission Accomplished": What His New Faux-Presidential Seal Symbolizes*, SLATE, June 21, 2008, http://www.slate.com/id/2193674/#bigdonut (describing reactions to proposed Social Security reforms).

31. *See* Larry Kudlow, *One-on-One with Austan Goolsbee, Obama's Econ Man*, CNBC.COM, Aug. 28, 2008, http://www.cnbc.com/id/26441455 (noting that the marginal rate would increase to 39.6% in "2019 at the earliest"); *see also* Furman & Goolsbee, *supra* note 8 (detailing Obama's suggestions for changing Social Security).

fully finance benefits as required by current law.[32] The flurry of interest at the 2009 White House Fiscal Responsibility Summit in "fixing Social Security" demonstrates the tenacity of the belief in the need to overfinance current Social Security benefits in order to secure the program's financial future thirty years from now.[33] However, depending on how one views the possibility that future congresses or presidential administrations would fail to honor the Treasury's obligations to redeem Social Security bonds and provide the cash necessary to fully pay benefits, there is a strong argument that there is no need to talk about increasing the contribution and benefit base or any other aspect of the payroll taxes right now, or at least to have any such increase go into effect any time soon.[34]

At the heart of continuing discussion of raising revenue now or years or even decades before the system actually needs any additional cash to pay benefits is the chimera of advance funding, a goal that runs counter to the fundamental "pay as you go" financing structure of the program.[35] The suggestion that Social Security needs more immediate revenue is a political question, not a programmatic or even tax-policy question, and the fact that Social Security financing is being discussed at all right now represents a victory of propaganda over analysis and a fundamental misunderstanding of Social Security itself.

---

32. For a discussion of the possible effects of the current economic crisis on Social Security's financial stability, see Robert Greenstein, President, Ctr. on Budget & Policy Priorities, Remarks at the Fiscal Responsibility Summit (Feb. 23, 2009), http://www.whitehouse.gov/assets/blog/Fiscal_Responsibility_Summit_Report.pdf; *see also The Economic Outlook and Budget Challenges: Hearings Before the H. Comm. on the Budget*, 111th Cong. 44–49 (2009) (statement of Alice M. Rivlin, Ph.D, Senior Fellow, Brookings Institution).

33. *See* Joe Conason, *Reform Healthcare—and Leave Social Security Alone*, SALON, Feb. 23, 2009, http://www.salon.com/opinion/conason/2009/02/23/fiscal_responsibility_summit/ (arguing that the Administration should leave entitlement programs alone as Social Security will "be solvent on its own for decades to come").

34. *See generally* 2009 BD. OF TRUSTEES REPORT, *supra* note 8.

35. For example, at a presidential campaign event in July 2008, John McCain attacked "pay-as-you-go" claiming that "Americans have got to understand that we are paying present-day retirees with the taxes paid by young workers in America today. And that's a disgrace. It's an absolute disgrace, and it's got to be fixed." Larry Rohter, *The Candidates Speak Off the Cuff, and Trouble Quickly Follows*, N.Y. TIMES, July 11, 2008, at A15. Of course, it is clear that pay-as-you-go financing is not a new phenomenon but rather was the intention since the program's inception. As the Committee on Economic Security noted, "Expressed differently, the plan we advocate amounts to having each generation pay for the support of the people then living who are old." COMM. ON ECON. SEC., REPORT OF THE COMMITTEE ON ECONOMIC SECURITY (1935), *reprinted in Economic Security Act: Hearing on H.R. 4120 Before the H. Comm. on Ways and Means,* 74th Cong. 45 (1935), http://www.ssa.gov/history/reports/ces5.html [hereinafter CES REPORT].

I. ORIGIN AND PURPOSE OF THE WAGE BASE LIMIT

The Social Security program—in the very limited original form of the Old Age Insurance—was initially enacted in 1935 as a response to the widespread financial crisis caused by the collapse of the international financial system, which severely shook both stock markets and financial institutions. The general economic effects of the crash and of the Great Depression in the 1930s affected almost all Americans, but the impact on the elderly was concentrated and devastating: their savings were lost when banks crashed, their pensions, for the lucky few who had any, were likely to have dried up, and their children, hit hard with lost employment, were much less able to help fill in economic gaps or even provide them a place to live.[36] It was impossible to claim that poverty and economic desperation were a result of individual shiftless or spendthrift behavior when economic collapse left at least a quarter of working-age men unemployed and essentially penniless.[37] A social response to a societal economic collapse was required.

Social Security was thus born out of economic necessity, but not as an instrument of immediate poor relief, which was the purpose of a different program altogether.[38] The primary objective of Social Security, particularly in the form that finally took effect after the 1939 fundamental

36. The Depression's impact, declared economist Paul H. Douglas in 1936, "increasingly convinced the majority of the American people that individuals could not by themselves provided adequately for their old age and that some form of greater security should be provide by society." PAUL H. DOUGLAS, SOCIAL SECURITY IN AMERICA 6–7 (1936). "The Depression deprived millions of older workers of jobs; it seemed unlikely that they would ever reenter the labor force." W. ANDREW ACHENBAUM, SOCIAL SECURITY: VISIONS AND REVISIONS 16 (1986). "A Massachusetts Census for Unemployment (1934) indicated an overall unemployment rate of 25.2%; the percentages for those aged sixty to sixty-four and sixty-five to sixty-nine were 27.2% and 29.8%, respectively." *Id.* at 201 n.21 (citing INDUSTRIAL RESEARCH DEP'T, UNIV. OF PA. WHARTON SCH. OF FIN. AND COMMERCE, UNEMPLOYMENT IN PHILADELPHIA FAMILIES, APRIL, 1931, at 20 (spec. rep. no. 1–8, 1931) and Herman B. Byer, *Employment Conditions and Unemployment Relief,* 43 MONTHLY LAB. REV. 1150, 1157–61 (1936)). Firms were unable to honor pension obligations and savings were lost. *See id.* at 16–17. "By 1934, over half of the elderly in America were impoverished. . . . Records of almshouses in 121 urban areas revealed that between 1929 and the end of 1933, the populations in those institutions jumped by almost 75 percent." NANCY J. ALTMAN, THE BATTLE FOR SOCIAL SECURITY: FROM FDR'S VISION TO BUSH'S GAMBLE 23 (2005).

37. *See* ACHENBAUM, *supra* note 36, at 16 (noting that the Depression posed a threat to everyone's futures and therefore the public became "more responsive to the problems of those growing older").

38. The Social Security Act of 1935 instituted the "Grants to States for Old-Age Assistance" program, which granted funds to each state, subject to certain requirements, to provide financial assistance for the elderly poor. *See* Social Security Act of 1935, Pub. L. No. 74-271, 49 Stat. 620, 620–22.

revisions, was to prevent future poverty in old age for workers who had spent a lifetime working, as well as for their spouses or surviving spouses.[39] It is important to keep this focus in mind when examining the financing mechanism of the payroll tax and the limitations of the wage base.

When the Roosevelt Administration sent Congress the original set of proposals that became the Social Security Act of 1935, the Old Age Insurance program was proposed to be financed through mandatory contributions from employees and an equivalent excise tax that employers would pay on employee earned wages, but there was no specific limit on the wages subject to the levy.[40] However, while manual laborers would be covered regardless of their level of earnings, workers earning more than $3,000 per year in nonmanual labor were exempt from coverage under the system.[41] The House Ways and Means Committee included the concept of the contribution and benefit base in its version of the legislation, expanding the number of workers covered to reach all workers in industrial or service work but limiting the "contribution and benefit base" to $3,000 per year. Thus, workers making more than that base figure would have essentially partial Social Security benefit accrual and taxation, as is the case today.[42]

---

39. *See generally* H.R DOC. NO. 76-110, at 1–2 (1939), *available at* http://www.ssa.gov/history/fdrstmts.html#1939b.

40. Roosevelt's advisors and many of the original designers of the program considered using general income-tax revenues rather than a dedicated wage tax to pay for Social Security. However, the income tax would have had to be greatly expanded to working people if it was not to be funded solely by the upper class, which then paid almost all income taxes. This was not a viable option at a time when only the wealthy were foreseen to ever pay income taxes and in addition would not have had Roosevelt's desired political effect of creating an "earned entitlement" that could never be taken away from workers. *See* Carolyn C. Jones, *Class Tax to Mass Tax: The Role of Propaganda in the Expansion of the Income Tax During World War II*, 37 BUFF. L. REV. 685, 689–91 (1988–89).

41. *See* JANEMARIE MULVEY & DEBRA B. WHITMAN, CONG. RESEARCH SERV., SOCIAL SECURITY: RAISING OR ELIMINATING THE TAXABLE EARNINGS BASE 1 (2008), http://aging.senate.gov/crs/ss9.pdf (noting that only very rarely would a manual laborer exceed $3,000 in yearly earnings at that time).

42. "The term 'wages' does not necessarily apply to the total remuneration received from the employer by the employee; the term includes only the first $3,000 of wages received by an employee from his employer with respect to employment during the calendar year." H.R. REP. NO. 74-615, pt. II, at 21 (1935), *available at* http://www.ssa.gov/history/reports/35housereport.html. Thus, the Committee on Economic Security report, CES REPORT, *supra* note 35, focused on covering workers in low paying jobs, as very few manual laborers would have had more than $3,000 in wages at that time, while the House bill, which ended up being essentially what was finally enacted, focused on covering all workers, but only wages of those workers up to a certain point. It is interesting that the House bill was essentially more economically democratic than the Administration's approach; covering workers who are in low-wage jobs presupposes more or less fixed employment at that level, whereas covering all workers (or at least all workers in industrial employment) but only earnings up to a fixed level allows for both mobility and fluctuation in earnings levels from year to year.

The original 1935 benefit formula was tied to the $3,000 per year base but was to be applied to all cumulative covered earnings over the worker's career, not counting more than $3,000 per year.[43] This version of Social Security never actually took effect as no benefits were paid prior to the enactment of the 1939 Amendments and the benefit structure was modified in those amendments to be based on average, rather than cumulative, wages up to the base[44] and to be substantially more progressive as well.[45] The payroll contribution made by employees and the excise tax paid by employers was also limited by the $3,000 base and collected through wage withholding, an innovation made necessary by the widespread coverage of workers who for the first time were subject to a federal obligation.[46]

Prior to the enactment of Social Security, the federal income tax affected a small minority of U.S workers, almost solely those at the top end of the

---

43. Benefits were to be calculated at 1/2 of 1% of the first $3,000 earned plus 1/12 of 1% of the next $42,000 earned plus 1/24 of 1% of the amount earned exceeding $45,000. *See* Social Security Act of 1935 § 202, 49 Stat. at 623.

44. The change from cumulative to average covered earnings as a base to which the formula was applied was done deliberately to achieve somewhat higher ultimate benefits for lower wage workers. *See* ACHENBAUM, *supra* note 36, at 32. In a 1938 report, the Advisory Council on Social Security stated the following:

> In addition, the Council believes that careful study should be given to the substitution of an *average* wage formula for the *accumulated* wage formula incorporated in the present Act. An average wage formula would more readily permit an increase in the early benefit payments and enable eventual costs to be kept within the limits prescribed under Recommendation II. Furthermore, in Recommendation VI the Council is on record as approving the average wage formula for computing survivorship benefits. By basing all benefits under Title II upon average wages, simplicity of understanding and administration is achieved as well as a consistent and related pattern of benefit payments.

ADVISORY COUNCIL ON SOC. SEC., REPORT OF THE 1938 ADVISORY COUNCIL ON SOCIAL SECURITY TO THE SOCIAL SECURITY BOARD AND THE SENATE FINANCE COMMITTEE (1938), http://www.socialsecurity.gov/history/reports/38advise.html.

45. The formula applied to the average monthly wage (AMW) was 40% of the first $50 plus 10% of the next $200, with the total result increased by 1% for each year with at least $200 of creditable wages. The result of applying that benefit formula to the AMW was the "primary insurance amount," or PIA, and all other Social Security benefits to be paid on the worker's account (spousal benefits, for example) were (and still are) calculated as a percentage of PIA. *See* GEOFFREY KOLLMANN, CONG. RESEARCH SERV., SOCIAL SECURITY: SUMMARY OF MAJOR CHANGES IN THE CASH BENEFIT PROGRAM: 1935–1996, at 2-3 http://www.ssa.gov/history/pdf/crs9436.pdf.

46. For a discussion of the innovation of wage withholding, see Joseph J. Thorndike, *Historical Perspective—The Price of Reorganization: Fewer Audits and Tax Forgiveness*, Sept. 2, 2002, TAXHISTORY.ORG, http://www.taxhistory.org/thp/readings.nsf/ArtWeb/9A29924C03AB9E1E85256DFE005981F9?OpenDocument ("World War II brought two major changes to the federal tax system. First, it dramatically expanded the individual income tax, boosting the number of taxpayers sevenfold in just six years. Second, it introduced wage withholding to help new taxpayers meet their obligations.").

income scale.[47] The insistence of the designers of Social Security on direct employee contributions to the system required a more expansive payment mechanism than the income-tax model could provide. In addition, it seemed important to distinguish Social Security contributions—under FICA—from taxes and to keep the whole system as far from the IRS as possible.[48]

Therefore, employers were enlisted in the cause of enforcement and collection of FICA contributions: charged with withholding the employee contributions and then forwarding both employee contributions and employer shares to the Bureau of Old Age Benefits for processing.[49] This collection system had a number of effects all by itself, not least making this contribution extremely visible to both workers and employers, and inspiring in workers contributing to the system a sense of connection to their future Social Security benefits.[50] The early information given to workers and the public generally about Social Security deliberately characterized the employee share of FICA as a contribution rather than a tax to emphasize each individual's relationship to the system and to his eventual entitlement to benefits.[51]

While the general point of the first Social Security Act in its entirety was to alleviate economic hardship for working people fallen on hard times, the original $3,000 figure for the contribution and benefit base for the Old Age Insurance portion of the Act was high enough to cover most American wages even though large categories of workers were initially left out of the system altogether.[52] Less than 10% of salaries exceeded $3,000 per year at

47. *See* Jones, *supra* note 40, at 689 (stating that only about 3.7% of the total population paid federal income taxes under FDR).

48. *See* ACHENBAUM, *supra* note 36, at 28 (explaining that social security officials released a campaign to assure American workers that "the taxes they were paying were like insurance premiums").

49. *See* W. ELLIOT BROWNLEE, FUNDING THE MODERN AMERICAN STATE, 1941–1995: THE RISE AND FALL OF THE ERA OF EASY FINANCE 92 (1996).

50. Nancy J. Altman, *Social Security and the Low-Income Worker,* 56 AM. U. L. REV. 1139, 1152–53 (2007) ("Nevertheless, to the extent that people have made specific monetary payments to ensure receipt of their own benefit, the moral obligation of government to honor the promises made is much stronger than it would be otherwise. Americans appropriately have a sense of contributing toward their own retirement and feel good about receiving those benefits. This sense of entitlement contributes to the program's success.").

51. ALTMAN, *supra* note 36, at 33–34 (discussing FDR's intent that the social security program be conceived of as an insurance program).

52. Farm workers and minority workers were especially affected. *See generally* Dorothy A. Brown, *Race and Class Matters in Tax Policy*, 107 COLUM. L. REV. 790 (2007). "Policymakers expected that all workers would someday participate in the old-age insurance plan, but practical administrative and constitutional considerations persuaded them to limit coverage at first. Roughly 9.4 million workers (including farmers, domestic servants, and government employees) were excluded from the new program." ACHENBAUM, *supra* note 36,

the time, and only 9% of the population made more than $2,500 a year in 1939, so the administration's 1935 proposal effectively covered the entire wage of all manual laborers in industrial jobs, few of whom would make as much as $3,000 per year, as well as most nonmanual workers in industrial work.[53] The focus was on getting people benefit coverage and only to a lesser extent on how to pay for those benefits.[54]

The drafters of the 1935 House bill that set the contribution and benefit base at $3,000, however, made an explicit decision to tie the base for earnings covered by Social Security for benefit purposes to the base for tax purposes.[55] When that original, very limited Social Security program was rewritten from the ground up in the 1939 Social Security Act Amendments and expanded into a true social insurance program, the contribution and benefit base was one of the few elements to carry over essentially unchanged.[56]

A critical point is that neither the 1935 program nor the 1939 amended program, which essentially established the basic program that is in operation today, contained any direct relationship between benefits paid out and amount of taxes paid in. It is true that the original legislation was based on a contributory annuity model, similar to private annuities

---

at 23. "That most of the poorest workers—such as Southern blacks—were excluded from coverage suggests that policymakers were willing to make politically expedient compromises." *Id.*

53. REPORT OF THE SOCIAL SECURITY BOARD, H.R. DOC. NO. 76-110, at 8–9 (1939), *available at* http://www.ssa.gov/history/reports/38ssbadvise.html. The 1939 Census showed that only 9% of the population made more than $2,500 per year, so clearly almost all of the wages of those workers covered by the program were included in the contribution and benefit base. Many categories of workers were left out of the original Act, for example, farm laborers, so that far fewer than 90% of all workers were actually covered by the original bill. *See* BUREAU OF THE CENSUS, U.S. DEP'T OF COMMERCE, POPULATION—SPECIAL REPORTS: EDUCATIONAL ATTAINMENT BY WAGE OR SALARY INCOME: 1940 (1946), http://www.census.gov/population/socdemo/education/p46-5/p46-5.pdf. Coverage was greatly expanded in the 1939 Act and later legislation.

54. "In 1935, the designers of Social Security, President Franklin Roosevelt's Committee on Economic Security, did not recommend a maximum level of taxable earnings in its plan, and the draft bill that President Roosevelt sent to the Hill did not include one. The bill emphasized who was to be covered by the system, not how much wages should be taxed. Being in the midst of the Depression, the Administration's attention was on the large number of aged people living in poverty." *See* MULVEY & WHITMAN, *supra* note 41, at 1.

55. *See generally* H.R. REP. NO. 74-615, pt. 2, at 19–22, 29–33 (1935) (establishing a system where the old-age benefits are paid directly from the federal Treasury, which is authorized to collect taxes on wages not exceeding $3,000).

56. *See* Social Security Act Amendments of 1939, Pub. L. No. 76-379, § 209, 53 Stat. 1360, 1373–78. The 1939 amendments added two new categories of benefits: dependent's benefits and survivor's benefits. In addition, the amendments "increased benefit amounts and accelerated the start of monthly benefit payments." *See* Soc. Sec. Admin., History: 1939 Amendments, http://www.ssa.gov/history/1939amends.html (last visited April 27, 2010).

purchased through insurance companies today, although with guaranteed benefits based on a benefit formula applied to cumulative earnings under the system.[57] However, that model was essentially abandoned only two years after contributions began to be collected and before any benefits were paid. The 1939 Amendments remade the 1935 Old Age Insurance program, which would have provided proportional benefits only for workers, into Old Age and Survivors Insurance (OASI), a true social-insurance program with a weighted benefit formula and spousal and survivor benefits.[58]

The common element, from 1935 up to the present, is that benefits are calculated based on earnings covered by Social Security, while taxes withheld are an entirely separate system, with no connection to benefits paid out.[59] Critics of Social Security are fond of comparing it to individual annuities or investment plans, but a more accurate private system comparison, albeit not precisely similar in all respects, is the employer-sponsored defined-benefit pension plan, in which benefits are accrued based on years of employment and financed by employer contributions to a trust, based on estimates of future financing needs.[60] The question that should be asked is why was payroll-tax financing a feature of the program at all given the partition of benefit accrual from system financing?

After the major revisions of the 1939 Social Security Amendments, the benefit calculation became more weighted toward low-wage workers, and survivor benefits were added, making the relationship between contributions paid in and benefits paid out even more remote and the system more of a true "social insurance" program.[61] Nonetheless, it is also

57. *See* CES REPORT, *supra* note 35, at 43–44 (recommending that contributions be based upon cumulative earnings to be phased in "1 percent in the first 5 years; 2 percent in the second 5 years; 3 percent in the third 5 years; 4 percent in the fourth 5 years and 5 percent thereafter").

58. Social Security Act Amendments of 1939 § 201, 53 Stat. at 1362–67.

59. H.R. REP. NO. 74-615, pt. 1, at 5–7.

60. For an exhaustive discussion of plan funding methods and elements, see DAN M. MCGILL ET AL., FUNDAMENTALS OF PRIVATE PENSIONS 201–333 (7th ed. 1996).

61. The Director of Old-Age and Survivors Insurance circulated a January 1940 memorandum reflecting the view and purpose of the 1939 amendments. *See* Memorandum from John J. Corson, Dir., Bureau of Old-Age and Survivors Ins., to Regional Representatives and Field Office Personnel, Old-Age and Survivors Insurance (Jan. 10, 1940), http://www.ssa.gov/history/reports/1939no3.html [hereinafter Corson Memorandum]. This memorandum explained that through a form of social insurance "we are endeavoring to protect society against the contingency that it will be called upon to support a large proportion of the people over sixty-five who can no longer support themselves." *Id.* The form of social insurance was meant to "replace a part of that wage income that made for the individual's own security and makes simultaneously for the protection of society against the neces[sity] of his support." *Id.*

The revision of the benefit formula reflects the change in the emphasis of the

clear from the legislative history that both the Roosevelt Administration and the Congress at the time viewed contributions by individual workers, as part of the financing of Social Security, to be an essential element of social insurance.[62]

Roosevelt's own insistence that workers who would ultimately benefit from the system should contribute to its costs is well known, as is his view that making direct payments into the system would create an unassailable "earned right" to retirement income that would not be subject to means testing.[63] The 1935 Committee on Economic Security (CES) Report to the

---

program. The original provisions offered primarily a plan for systematic savings for old age. The amendments, on the other hand, are designed to provide a minimum subsistence income for the retired worker and his dependents or for certain of his survivors, relating the amount of the benefit to his family responsibilities and, roughly, to the level of his former earnings as well as to the extent of his participation in the system. The primary monthly benefit, payable to a qualified worker at 65 or after, is based on his average monthly wage (as defined subsequently) according to the following formula: (1) a basic amount of 40 percent of the first $50 of the average monthly wage, plus 10 percent of the amount by which that average exceeds $50 and does not exceed $250, and (2) 1 percent of the amount calculated under (1) multiplied by the number of years in which the worker has received $200 or more in wages from covered employment. The minimum primary benefit is set at $10.

Lyle L. Schmitter & Betti C. Goldwasser, *The Revised Benefit Schedule Under Federal Old-Age Insurance*, SOC. SECURITY BULL., Sept. 1939, at 3, *available at* http://www.ssa.gov/history/reports/1939no2.html. "The average-wage formula in the amendments relates benefits not only to presumptive need, as indicated by the level of customary earnings, but also to the relative amount of time spent in covered employment." *Id.* at 7; *see also* Social Security Act Amendments of 1939 § 202(a); Social Security Act, 42 U.S.C. § 402 (2006) (describing the current benefit formula).

62. *An Act to Amend the Social Security Act of 1939 and for Other Purposes: Hearing on H.R. 6635 Before the S. Comm. on Finance.*, 76th Cong. 3, 5, 7–8, 16 (1939) (highlighting that contributions by individual workers were viewed as part of the financing).

63. Senator Moynihan (D-N.Y.) describes President Roosevelt's feelings as follows:

> We know one thing in particular: President Roosevelt was absolutely determined that the payments made into this system would be credited to the individual who had paid them. Each individual would have an account recording every nickle [sic] he and his employer put in, and a passbook in the form of a Social Security card with his or her name on it. In 1941, Luther Gulick, a very distinguished professor at Columbia University, and one of the founders of the profession of public administration in our country, was working temporarily in Washington. He went in to see President Roosevelt, who was not then surrounded by staff. . . . Professor Gulick suggested that perhaps the time had come to stop levying payroll taxes separately from income taxes. Gulick said that it is all really one set of finances. Should we not just have one rate and collect it at one time? It would be efficient. Why have two sets of books, two sets of rates of contribution, when one would do? Gulick went back and wrote a memorandum of the conversation. The President replied. He said: I guess you are right on the economics, but those taxes were never a problem of economics. *We put those payroll contributions in so as to give the contributors a legal, moral, and political right to collect their pension and their unemployment benefits with those taxes in there. No damned politician can ever scrap my Social Security Program.* Roosevelt

President echoed this conviction:

> Contributory annuities are unquestionably preferable to noncontributory pensions. They come to the workers as a right, whereas the noncontributory pensions must be conditioned upon a "means" test. Annuities, moreover, can be ample for a comfortable existence, bearing some relation to customary wage standards, while gratuitous pensions can provide only a decent subsistence.[64]

However, by the time the major expansion of Social Security took place four years later, the link between contributions and benefits was already being described in more ambiguous terms:

> The present old-age insurance system, while maintaining a reasonable relationship between past earnings and future benefits, provides proportionately greater protection for the low-wage earner and the short-time wage earner than for those more favorably situated. In other words, it recognizes *presumptive* need as an essential consideration in any socially adequate old-age insurance system. . . .
>
> But every worker, regardless of his level of earnings or of the length of time during which he has contributed, will receive more by way of protection than he could have purchased elsewhere at a cost equal to his own contributions. In other words, the system recognizes the principle of individual equity, as well as the principle of social adequacy.[65]

By 1939, it appears the redesigners of the original program viewed contributions as more of a political mechanism for assuring rights to adequate benefits in old age rather than as actual payment for a future retirement annuity or an investment yielding a return in the form of the retirement benefits. The expansion of the program to meet the needs of elderly spouses and survivors of covered workers and the acceleration of payment of benefits to meet more current needs, along with other changes that expanded the role of Social Security to resemble what it is today, required a recognition that worker contributions constituted partial, not complete, financing for the system as a whole, not direct payment for the benefits they would eventually receive.

In the original estimates for Social Security's financing, even in 1935, and to a greater extent in the 1939 revamping of the system, worker and employee contributions via the payroll tax were not seen as the sole, long-term source of financing.[66] General tax revenues were projected to begin

---

> wanted that money to be identified with the individuals who had contributed it. And that system worked very well indeed.

101 CONG. REC. 28,086 (1990) (emphasis added).

64. CES REPORT, *supra* note 35, at 39.

65. REPORT OF THE SOCIAL SECURITY BOARD, H.R. DOC. NO. 76-110, at 5 (1939), *available at* http://www.ssa.gov/history/reports/38ssbadvise.html.

66. "There can be no escape from the costs of old age; and since these costs must be

partially financing benefits at least by the 1960s on the assumption that contribution rates would not rise above 5%.[67] As will be discussed below, partial nonpayroll-tax financing is completely consistent with the social insurance model and belies any notion that Social Security benefits are tied in any direct way to their source of financing.[68]

So what conclusions can be drawn from examination of the fundamental principles associated with the Social Security contribution and benefit base that should guide any future changes? First, clearly the level of covered wages has traditionally been aimed at covering most wages in the national economy, but the highest wage earners have always had the top part of their salaries exempted from Social Security taxes and omitted from their earnings records for benefit computation purposes. It is not clear, however, that this design feature is necessarily an inviolable basic principle of social insurance generally or of the U.S. Social Security system in particular.

Second, the causative relationship between contributions and benefits that so many commentators, as well as members of the public, seem to perceive as a fundamental principle of Social Security simply does not exist, at least not in the sense of benefits resulting from or depending on taxes paid. Benefits are based on earnings recorded in the Social Security system, not on taxes paid, similar to the way workers covered by an employer-sponsored defined-benefit pension plan accrue benefits over a

---

met, an orderly system under which employers, employees, and the Government will all contribute appears to be the dignified and intelligent solution of the problem." CES REPORT, *supra* note 35, at 46. "Since the nation as a whole, independent of the beneficiaries of the system, will derive a benefit from the old-age security program, it is appropriate that there be Federal financial participation in the old-age insurance system by means of revenues derived from sources other than pay-roll taxes." ADVISORY COUNCIL ON SOC. SEC., FINAL REPORT, S. DOC. NO. 76-4, at 6 (1938), *available at* http://www.ssa.gov/history/reports/38advise.html. The Social Security Board, while differing in many respects from the Advisory Council's report, echoed this sentiment:

> The Board is of the opinion that it would be sound public policy to pay part of the eventual cost of the benefits proposed out of taxes other than pay-roll taxes, preferably taxes such as income and inheritance taxes levied according to ability to pay.
>
> The portion of the total costs to be met by taxes other than pay-roll taxes should depend upon the proportion of the general population covered by the insurance system. The wider the coverage, the more extensive this contribution from other tax sources might properly be.

REPORT OF THE SOCIAL SECURITY BOARD, H.R. DOC. NO. 76-110, at 12 (1939), *available at* http://www.ssa.gov/history/reports/38ssbadvise.html.

67. *See* CES REPORT, *supra* note 35, at 45 ("Benefit payments will be light in the early years but will increase steadily until, by 1965, they will exceed the annual receipts. It is at this stage that the Federal Government would begin to make contributions to the annuity system . . . .").

68. *See infra* text accompanying notes 70–71.

career with that employer.[69] There has never been any connection between the amount of FICA taxes paid by a worker and her employers on her behalf and her ultimate level of benefits.[70]

Of course, as a political matter, it may not matter that the tax–benefit connection does not exist if the public at large believes that it does, but that is likely a question of ignorance and a generational unfamiliarity with the principles of accrual of benefits under any kind of defined benefit pension plan, something with which today's workers have very little experience.[71] As a policy question, however, it is time to set aside popular mythology based on a misunderstanding of the actual fundamental principles of Social Security financing and benefits, one that clearly distorts both policy analysis and public understanding of the program.

Finally, the wage-base limit is really a by-product of the original limits on coverage for benefit purposes and in some ways is a relic of the original 1935 legislation that envisioned Social Security as a type of contributory

---

69. *See* 42 U.S.C. § 415 (2006) (describing the computational methods for primary insurance amounts). I previously described the Social Security system as follows:

> Payroll taxes are merely a method of financing the system, not the basis for benefits earned and paid out. Benefit calculations are made based on *earnings* recorded in the Social Security system, which is done as a record-keeping matter through withholding tax records filed with the Federal Reserve and forwarded to the Social Security Administration. But benefit calculations do not take into account the amount of *taxes* paid, and benefits cannot be reduced in the event of a failure to pay such taxes by the employer who is responsible for withholding FICA taxes from workers' paychecks. The system could as easily be financed through income tax revenues, like other government expenditures, without any impact on the earnings-based benefit structure.

Patricia E. Dilley, *Taking Public Rights Private: The Rhetoric and Reality of Social Security Privatization*, 41 B.C. L. REV. 975, 1000 (2000); *see also* Deborah A. Geier, *Integrating the Tax Burdens of the Federal Income and Payroll Taxes on Labor Income*, 22 VA. TAX REV. 1, 35 (2002).

70. *See* Dilley, *supra* note 69, at 1000.

71. Professor Stephen F. Befort chronicled this significant shift as follows:

> Through the 1970s, traditional defined benefit plans predominated. In 1975, for example, 87% of all workers covered by a pension plan participated in a defined benefit plan. Since 1980, however, there has been a significant shift toward defined contribution plans. While the number of employees covered by a defined benefit plan fell 25% between 1980 and 2000, the number participating in a defined contribution plan jumped 250%. As of 2005, twice as many American workers were covered by defined contribution plans as compared to defined benefit plans. Of those with pension coverage, only 19% of U.S. households are currently covered by a defined benefit plan, while 58% are covered solely by a defined contribution plan, and 23% participate in both types of plans.

Stephen F. Befort, *The Perfect Storm of Retirement Insecurity: Fixing the Three-Legged Stool of Social Security, Pensions, and Personal Savings*, 91 MINN. L. REV. 938, 948 (2007); *see also* Henry H. Drummonds, *The Aging of the Boomers and the Coming Crisis in America's Changing Retirement and Elder Care Systems*, 11 LEWIS & CLARK L. REV. 267, 281 (2007) ("In summary, the defined benefit traditional pension system suffers from a marked decline in its coverage of American workers and a funding crisis in sectors heretofore thought to be its strength.").

annuity to provide for a future limited floor of retirement security.[72] That version of Social Security was replaced, before it ever really took effect, by the more expansive social insurance model of the 1939 Amendments.[73] Social insurance has a much broader mission—protecting society by caring for individuals:

> Social insurance is one of the ways in which we endeavor to make society secure. . . . The basic purpose of all forms of social insurance is to replace a sufficient part of that wage income when it is lost as a result of any of these hazards—unemployment, accident, old age, or death of the wage earner—to insure not only that the individual may look forward to protection, but that society as well may be protected against the hazards which it faces.[74]

Clearly, Social Security was explicitly intended, and successfully functions, as a social stabilizer, protecting society from unrest by giving workers generally the promise of protection against destitution in old age.[75] From that perspective, high-wage workers are "purchasing" more with their Social Security contributions than just their future benefit entitlement, and the indirect economic benefit of social and economic stability for society generally, particularly for those most likely to reap the biggest economic benefit from society and a stable capitalist economy, was an intentional result.[76]

---

72. MULVEY & WHITMAN, *supra* note 41, at 1.

73. *See generally* Social Security Act Amendments of 1939, Pub. L. No. 76-379, 53 Stat. 1360.

74. Corson Memorandum, *supra* note 61. After passage of the 1939 Amendments, John Corson traveled around the country to explain the new law to employees of the agency. A Director's Bulletin put Corson's remarks into the written form from which this quotation is taken. These remarks, I believe, reflect the contemporary and later understanding of social insurance of most of the original designers of the 1939 bill.

75. President Roosevelt made this point explicitly in his statement to Congress even before the formulation of the Social Security Act:

> Among our objectives I place the security of the men, women and children of the Nation first. This security for the individual and for the family concerns itself primarily with three factors. People want decent homes to live in; they want to locate them where they can engage in productive work; and they want some safeguard against misfortunes which cannot be wholly eliminated in this man-made world of ours. . . . The third factor relates to security against the hazards and vicissitudes of life. Fear and worry based on unknown danger contribute to social unrest and economic demoralization. If, as our Constitution tells us, our Federal Government was established among other things, "to promote the general welfare," it is our plain duty to provide for that security upon which welfare depends.

President Franklin Delano Roosevelt, Message to Congress Reviewing the Broad Objectives and Accomplishments of the Administration (June 8, 1934), http://www.ssa.gov/history/fdrstmts.html#1939b.

76. This result, of course, is quite similar to one premise of progressive income taxation discussed below: higher income taxpayers benefit more from the institutions and protections of government and therefore should pay higher taxes. *See generally* Joseph M. Dodge, *Theories*

No fundamental principle of Social Security is necessarily at stake, therefore, in consideration of a wide range of proposals to raise, redesign, eliminate, or add to the base for tax purposes, benefit purposes, or both, or for adding other sources of financing for Social Security benefits. Even in 1939, the program's designers anticipated that the program's possible broader financing needs in the future would need to be met either by continuing increases in the tax base or infusions of general tax revenues.[77] Any discussion of financing changes that are "true to original principles of Social Security" clearly should include a broader menu of options than just increases in the base or rates of the current payroll-tax financing system. It is particularly appropriate to begin with an examination of the wage-base limit, as expansion of the base can serve two purposes: increased financial viability and increased fairness in the eyes of the taxpayers.

II. RAISING THE BASE TO TRADITIONAL TARGET LEVEL

The main purpose of the doughnut hole proposal is to increase payroll-tax contributions to Social Security only from higher wage workers in order to shore up the long-term financing of the system.[78] Setting aside for the moment the question of whether long-term financing needs to be addressed at all right now, the first issue should be the appropriate level for the wage base from a Social Security program perspective. Apart from political considerations, the base's two separate roles need to be reimagined from the perspective of the Social Security program's mission itself and not simply from a tax policy or political salability perspective.

Social Security was designed and has developed over several decades as a way to deal with certain problems that still exist: insuring working people and their families against the chance of destitution when they are no longer able to work (whether because of old age, disability, or death) and, as a result, promoting social stability throughout the life cycle.[79] The central

---

of Tax Justice: Ruminations on the Benefit, Partnership, and Ability-to-Pay Principles, 58 TAX L. REV. 399 (2005).

77. *See An Act to Amend the Social Security Act and for Other Purposes: Hearing on H.R. 6635 Before the S. Finance Comm.*, 76th Cong. 81, 83, 86, 92–93, 101, 250 (1939) (highlighting that increases in taxes as well as increased contribution rates were considered).

78. *See supra* notes 8–9 and accompanying text.

79. Frances Perkins, the "mother of Social Security" and a member of the original Committee on Economic Security that developed the proposal for the first Social Security legislation in 1935, described the origins of Social Security in a speech to Social Security Administration employees on the twenty-fifth anniversary of the program as follows:

> We were not yet out of the woods of the Great Depression and, of course, it was the Great Depression which we must never forget in this country, which was the proximate cause of this movement which was launched at that time—this movement to write under the lives of the American people a basis of security which came to

criterion to be applied to proposed modifications of the base (for benefits, contributions, or both) is whether the changes improve or detract from the program's ability to continue to meet that central purpose. This is not to ignore the impact of the dedicated payroll-tax base as part of the overall federal tax structure. Nonetheless, the programmatic purposes of the wage base should be balanced against the overall economic impact of base increases on Social Security taxpayers and beneficiaries.

One clear rationale for another round of ad hoc contribution and benefit base increases can be found in a comparison of wages covered by the FICA wage base at the onset of the program and now. In 1937, the $3,000 contribution and benefit base covered 92% of all wages in the national economy, but over the next fifty years, as the economy expanded in the post-World War II period, the percentage of wages covered declined, requiring periodic legislation to increase the base.[80] At several points during the history of the program, the disparity between wages actually covered and the traditional 90% standard was quite wide—for example, in 1965, only 71% of all wages were covered—and eight ad hoc increases in the wage base were put in place between 1939 and 1972.[81] Indexing the base to increases in average wages in 1972, combined with the last set of ad hoc increases that were eventually completed in 1992, was intended to eliminate the need for ad hoc increases.[82]

However, indexing the base to increases in average wages, rather than to some more comprehensive gauge, such as aggregate U.S. earnings, means that growing income and earnings inequality allows more compensation at the top of the earnings scale to escape the base. While the earnings of about 94% of all U.S. workers are completely covered by the 2010 base of

---

> them out of the orderly, substantial, and regular contributions to their future and to the future hazards.

Frances Perkins, Remarks at the Social Security Twenty-Fifth Anniversary Ceremonies (Aug. 15, 1960), http://www.ssa.gov/history/25annoasis.html.

80. *See generally* Patricia P. Martin & David A. Weaver, *Social Security: A Program and Policy History*, 66 SOC. SECURITY BULL. 1, 1 (2005), *available at* http://www.ssa.gov/policy/docs/ssb/v66n1/v66n1p1.pdf.

81. *See, e.g.*, MULVEY & WHITMAN, *supra* note 41, at 4.

82. In 1973, Commissioner of Social Security Robert M. Ball explained that major social security legislation enacted in July 1972 included

> [an increase] from $9,000 in 1972 to $10,800 in 1973 and to $12,000 in 1974 the maximum amount of a worker's annual earnings that may be counted in figuring his and his family's social security benefits (and on which he pays social security contributions) and provided in addition for keeping the amount up to date automatically in the future as average wages rise; and a revised contribution rate schedule . . . .

Robert M. Ball, *Social Security Amendments of 1972: Summary and Legislative History*, SOC. SECURITY BULL., March 1973, at 3, 3, *available at* http://www.ssa.gov/history/1972amend.html.

$106,800, the share of total wages subject to the payroll tax has been steadily falling—from about 90% in 1982, at the time of the last big Social Security financing bill, to 85% in 2005—mainly because of the great disparity between average wage increases and compensation increases for the highest wage earners.[83] If no additional changes are made to the wage base, it is projected to cover only 83% of wages by 2014.[84] However, increasing the base to cover 90% of wages in the national economy would not be a trivial change. For example, in 2005, meeting that goal would have required a jump from that year's base limit of $90,000 to $150,000.[85]

Even so, no matter what measure is used, increasing the contribution and benefit base at least back up to the 90% level seems to be the easiest case to make from the perspective of fidelity to the basic design and purpose of Social Security, for several reasons. From a simple financing perspective, it could be argued that coverage of almost all, if not 100% of, wages earned is one way to insure adequate financing of current benefits without having to increase the FICA rate to levels that would be extremely burdensome for lower and middle wage workers. Nonetheless, there is more to the base than financing concerns, which may make raising it a programmatic necessity regardless of financing concerns. It makes sense, then, to unpack the base into contribution purposes and benefit purposes and to examine the rationales for each type of limit independently.

A. Benefit Base

The traditional goal of including at least 90% of wages for both benefit and tax purposes satisfies important political and philosophical objectives of social insurance—if almost all workers and most of their wages are covered, they have a commitment to the program for themselves as well as for society in general. However, the importance of universal benefit coverage—both of workers and of their wages—goes deeper than creation of stakeholders.

Coverage of most earnings for benefit purposes is the underpinning of social insurance's role of encouraging social stability—tying together economic interests of working, middle, and upper economic classes. It is not surprising that the designers of Social Security thought it important to cover at least earnings up to the top 10% of wage earners, given the dramatic downturn of economic fortunes during the Great Depression that saw those even at or near the top lose assets and the capacity to earn

83. KATHLEEN ROMIG & JANEMARIE MULVEY, CONG. RESEARCH SERV., SOCIAL SECURITY: RAISING OR ELIMINATING THE TAXABLE EARNINGS BASE 1 (2009).

84. *Id.*

85. *Id.*

through no fault of their own.[86] Melding the interests of top and bottom earners served not only as political glue but as a stabilizer of economic expectations for society generally—wage earners covered by Social Security might fall down the economic ladder, but in old age they would not need to appeal for government aid to the poor.[87]

While some workers manage to earn at or above the base for their entire careers, many more have earnings histories that fluctuate, sometimes dramatically, over thirty-five to forty years of work.[88] If the benefit base is compressed and fails to capture earnings near the top, benefits could be dramatically affected for workers who have some good years accompanied by many average or below average ones. The whole point of providing an entitlement in the future to benefits based on past earnings is to allow for the possibility of misfortune along the way to old age, to provide income

86. *See generally* DAVID M. KENNEDY, FREEDOM FROM FEAR: THE AMERICAN PEOPLE IN DEPRESSION AND WAR, 1929–1945, at 58–59 (1999) (providing examples of the poor economic climate during the Depression, such as the 26,355 business failures by the end of 1930).

87. *See supra* notes 63–67 and accompanying text.

88. The University of Michigan Retirement Research Center studied this phenomenon and concluded,

> In practice, when we compare the hypothetical profiles with actual earnings, we find that the scaled profiles do not compare well to actual earnings paths. One reason for this divergence is that the assumption of steady work does not track the experience of actual Health and Retirement Study (HRS) workers. Well over one-third of all men and women in our sample did not have covered earnings in their 20s, and many women had zero earning years after that. All the hypothetical profiles are higher and flatter than the typical HRS workers in our sample. We also find that the Average Wage Index, intended to reflect a weighted average of actual earnings at any given time, does not match the average earnings of any given cohort. In addition, the AWI exceeds average actual earnings during working cohorts' early years, and, using measures unaffected by high outlier earners, it is still higher than HRS actual cohort earnings in all years. Further, median HRS actual earnings were more similar to the low versus the medium scaled profile. Even after restricting the HRS sample to respondents with substantial work histories, the medium scaled profile is 28% above HRS actual median earnings, implying a lifetime difference of more than $150,000.

Andrew Au, Olivia S. Mitchell & John W.R. Phillips, *Modeling Lifetime Earnings Paths: Hypothetical Versus Actual Workers* 19 (Univ. Mich. Ret. Research Ctr., Working Paper No. 2004-074, 2004), http://www.mrrc.isr.umich.edu/publications/papers/pdf/wp074.pdf. In addition, the Center for Retirement Research Center at Boston College found,

> Few workers have level career earnings, so the traditional approach to policy simulation represents a serious distortion of actual labor market experience. Moreover, differences in the pattern of career earnings can produce wide disparities in pension entitlements, even for workers with the same average earnings, under individual account and other retirement plans.

Barry Bosworth, Gary Burtless & C. Eugene Steuerle, *Lifetime Earnings Patterns, the Distribution of Future Social Security Benefits, and the Impact of Pension Reform* (Ctr. for Ret. Research at Boston Coll., Working Paper No. 1999-06, 2000), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=252052.

assurance in advance and without proof of need at the time benefits are paid.

Correcting the benefit base to cover at least 90% of wages (not just workers) is thus completely consistent with fundamental principles of Social Security and indeed is necessary to fulfill program benefit goals of adequate benefit replacement and program participation. Raising the base higher than the traditional 90% coverage standard raises additional questions, however. One rationale that has always been given for limiting benefit coverage to earnings up to, but not above, the benefit base is that replacement of any part of wages above the base is unnecessary in a public-benefit program.[89] Given the increasing concentration of wealth and earnings levels at the top end of American incomes and the growing inequality of those incomes, it might be appropriate to raise the benefit base above 90%, but there is no clear line indicating how high is too high.

B. Contribution Base

The programmatic connection between the base and benefits is clear—the base serves as a limit to the amount of earnings that can be used to calculate the ultimate benefit entitlement, and the level of the base is important in capturing a complete picture of a worker's earnings history. The connection between the base and wages subject to FICA, however, is less clearly connected to Social Security program goals beyond simply raising sufficient funds to pay benefits. The wage aspects of the base, therefore, unlike the benefit aspects, implicate both Social Security programmatic principles and general tax justice and policy issues.

1. The Contribution Base from a Programmatic Perspective

As described earlier, the original base was set at $250 per month, or $3,000 per year for benefit-accrual reasons, but it is not clear that the House drafters of the 1935 Act had anything more in mind than simple symmetry when they used the same limit for FICA purposes.[90] This symmetry is consistent with the original vision of the program as a compulsory, federally sponsored annuity program.[91] This original design

89. *See* ROMIG & MULVEY, *supra* note 83, at 1 (explaining that the Social Security program was originally intended only to "provide a 'core' benefit as a floor of protection against poverty").

90. *See supra* note 43 (describing how the Social Security Administration calculated benefits in 1935).

91. From the descriptions in the 1935 CES Report and other documents, the original design resembled an employer-provided pension plan more than social insurance:

Pensions sufficient for a decent subsistence for all of the aged who are dependent

explains the insistence in the original bill on categorizing payments into the system as "contributions" rather than taxes—employees were described as making contributions to social insurance, while employers were said to be paying an "excise tax" on their employees' wages.[92]

However, while workers clearly felt they were contributing to their own future retirement through the FICA tax withheld from their wages, it is not clear that policymakers viewed those withheld amounts as anything other than taxes, albeit taxes dedicated to the financing of those future benefits. The 1938 Advisory Council's report to the President on the upcoming rewrite of the Social Security Act describes both the employee and employer shares of FICA as a tax and further recommends requiring employers to show employees the amount of taxes deducted from their wages under the old-age insurance system.[93] The expansion of Social Security in the 1939 Amendments into a true, broad-based social insurance program protecting workers and families from future need in retirement or because of early death made it harder than ever to view the program as an individual annuity system.[94] The public entitlement required public financing, even if most of the needed funds came through the dedicated payroll tax.

As discussed earlier, it is important for the benefit base to cover at least 90% of earnings in the U.S. economy if the Social Security program is to meet its goal of paying benefits that provide adequate, but not excessive, earnings replacement for workers whose earnings may greatly fluctuate throughout their working lifetime.[95] The question is whether there is a

---

> upon the public for support are approved by the overwhelming majority of the people of this country. In order to reduce the pension costs and also to more adequately provide for the needs of those not yet old but who will become old in time, we recommend a contributory annuity system on a compulsory basis, to be conducted by the Federal Government.

CES REPORT, *supra* note 35, at 42.

92. *Id.* at 22.

93. ADVISORY COUNCIL ON SOC. SEC., FINAL REPORT, H.R. DOC. NO. 76-4, at 6–7 (1938), *available at* http://www.ssa.gov/history/reports/38advise.html.

94. GEOFFREY KOLLMANN, CONG. RESEARCH SERV., SOCIAL SECURITY: SUMMARY OF MAJOR CHANGES IN THE CASH BENEFIT PROGRAM: 1935–1996, at 2 (1996), *available at* http://www.ssa.gov/history/pdf/crs9436.pdf.

95. This concern was expressed very early in the switch from cumulative to average earnings as the basis for calculating the primary insurance amount (PIA):

> The Board recommends that benefits be calculated upon the basis of average wages, rather than, as at present, upon total accumulated wages. This change would make it possible to increase early benefits and to relate benefits more closely to the previous normal wage income of the individual. It would also eliminate, as the years go by, the large bonus which present provisions would afford those who have had only a brief period of participation prior to the date of retirement. . . .
>
> While the Board believes that benefits should be related to the average wage, it

similar programmatic imperative for setting the wage base at 90% or more of average earnings. The contributory function of the payroll tax is its most direct connection to Social Security program objectives, as it was envisioned originally as a way to establish an unbreakable political entitlement to eventual benefit receipt.[96] It is not clear, however, that the absolute level of contribution is particularly significant in the creation of that entitlement.

For one thing, as discussed above, there is no—and never has been any—statutory connection between entitlement to benefit payments and payroll taxes paid. Benefits are required to be paid to each worker who has satisfied the basic requirements for coverage—based on earnings reported or demonstrated to have been earned for the requisite number of quarters—whether or not her employer actually withheld the proper amounts or paid over to the Treasury amounts owed.[97] Moreover, in a public program governed by statutes that can be amended at any time by Congress, the entitlement of any worker to any particular benefit is theoretically subject to change any time Congress is in session, regardless of any contributions made over her working lifetime.

In reality, in the entire history of Social Security, no Congress has ever acted to reduce the current benefits of workers already receiving them, and only once has Congress reduced a scheduled increase in benefits for beneficiaries already in pay status: the 1983 Social Security Amendments delayed the annual cost-of-living increase from June to December on a permanent basis, beginning with the 1983 Cost-of-Living Adjustments (COLA).[98] This action was only taken during an extreme financial crisis in

---

> recognizes that benefits should also be related to the number of years the individual has been in covered employment and has made contributions. The Board therefore recommends that an insured individual, upon retirement, receive a basic benefit related to his average wages; and that, for every year he has earned more than some small specified amount of wages in covered employment, his basic monthly benefit be increased by a specified percentage. Conversely it recommends that for every year a person does not earn this specified amount of wages, the basic monthly benefit be reduced by the same percentage.

REPORT OF THE SOCIAL SECURITY BOARD, H.R. DOC. NO. 76-110, at 6–7 (1939), http://www.ssa.gov/history/reports/38ssbadvise.html. The latter part of this recommendation, to increase or reduce benefits based on years of earnings above or below a set limit, was not included in the final 1939 legislation. However, other aspects of the benefit formula, such as dropping out the lowest five years of earnings before calculating the average indexed monthly earnings (AIME), may be said to be addressing some of the same concerns.

96. *See* MULVEY & WHITMAN, *supra* note 41, at 1.

97. *See supra* note 23 and accompanying text; *see also* Social Security Act of 1935, Pub. L. No. 74-271, § 202, 49 Stat. 620, 623; Social Security Amendments of 1939, Pub. L. No. 76-379, § 202(a), 53 Stat. 1360, 1363–67.

98. *See* Social Security Amendments of 1983, Pub. L. No. 98-21, 97 Stat. 65.

which the system's reserves were scheduled to be exhausted in the middle of 1983, making it impossible to pay the full amount of benefits due out of either current or accumulated payroll taxes, a situation not currently predicted to occur again until around 2040.[99]

The same cannot be said for benefits scheduled to be paid but not yet in payment status. Congress has acted several times to amend the program to reduce or eliminate benefits promised to future beneficiaries. One major example of such a cutback occurred in 1977 when Congress corrected an error in calculating automatic wage indexing of initial benefits, enacted in 1972, which had been discovered to be increasing initial benefits at roughly twice the intended rate.[100] The correction was made for beneficiaries first entitled to benefits on January 1, 1979, and the ensuing furor over the "notch" between benefits for those becoming entitled in 1978 and earlier and those becoming entitled in 1979 and later made Congress wary of such abrupt changes in benefits even for future beneficiaries.[101] The 1983 Amendments included possibly the largest cutback ever enacted in benefits for future beneficiaries in the form of the "increase in the retirement age."[102] This change is still in the process of phasing in over a twenty-year period (Congress having learned its lesson with the "notch baby" furor), but when completely in effect, it will raise the age for full benefits from 65 (in 1983) to 67 by 2022—a change that is not really an increase in the retirement age, but rather is a benefit decrease of up to 30% for those taking benefits before age 67 in 2022 and later.[103]

Critics of Social Security point to these kinds of reductions in future benefits and the ability of Congress to reduce or eliminate any benefits at any point as proof that there is no such thing as real entitlement to Social Security benefits regardless of the contributory FICA structure.[104] But it is

---

99. *See id.*; *see also* H.R. REP. NO. 98-25, pt. 1, at 1–2 (1983), *reprinted in* 1983 U.S.C.C.A.N. 143, 219–20 (describing the economic situation that made the Amendments necessary); 129 CONG. REC. 7392–94 (1983), *reprinted in* 1983 U.S.C.C.A.N. 143, 498–99; 2008 BD. OF TRUSTEES REPORT, *supra* note 8, at 10 (projecting the years in which Social Security trust funds will be exhausted).

100. *See* Social Security Amendments of 1977, Pub. L. No. 95-216, § 201, 91 Stat. 1509, 1514; *see also* Soc. Sec. Admin., History of SSA-Related Legislation, http://www.ssa.gov/legislation/history/95.htm.

101. For a brief summary of the controversy, see Op-Ed, *The Greed of the Notch Babies,* N.Y. TIMES, Jan. 13, 1998, at A22, *available at* http://www.nytimes.com/1988/01/13/opinion/the-greed-of-the-notch-babies.html.

102. *See* Social Security Act Amendments of 1983, § 201(a), 97 Stat. 107 (amending 42 U.S.C. § 416).

103. *Id.*

104. For an extended discussion of why the Social Security benefit entitlement is just as secure, and perhaps more secure, than private entitlement in savings and investments, see Patricia E. Dilley, *The Evolution of Entitlement: Retirement Income and the Problem of Integrating*

naïve to think any sort of private entitlement or ownership right to future income, whether in private pension plans or investment accounts designed to produce old-age income, is any more secure than the Social Security entitlement.[105] Private investment accounts are of course subject to the vagaries of investment markets, not to mention other sorts of losses both in value and sometimes in title to which private property is also subjected (bank failures and eminent domain exercise are two examples).[106] As for employer-sponsored private pensions, over the past fifty years, long after the economic catastrophe of the Great Depression revealed the weaknesses of pension funding and despite the funding requirements put in place by ERISA in 1974 and later, employers have frequently underfunded the trusts financing those plans and subsequently gone bankrupt, leaving the plan without sufficient funds to pay benefits.[107]

Private Pensions and Social Security, 30 LOY. L.A. L. REV. 1063 (1996–1997).

105. For a discussion of the public–private determination of property rights, see Amnon Lehavi, *The Property Puzzle*, 96 GEO. L.J. 1987 (2008).

106. For a report on the losses to 401(k) account balances from the last two years of market downturns, see, for example, Jack VanDerhei, *The Impact of the Recent Financial Crisis on 401(k) Account Balances*, EMP. BENEFIT RES. INST. ISSUE BRIEF, Feb. 2009, at 1, 10, http://www.ebri.org/pdf/briefspdf/EBRI_IB_2-2009_Crisis-Impct.pdf (charting the time needed to recover from 401(k) losses based on certain equity and non-equity return assumptions at figures 6 and 7). The collapse of banks is far from a historic footnote, and their takeover by the FDIC is now a weekly occurrence. One of the first and most extensive examples was the collapse of IndyMac Bank in California, which was taken over by the FDIC on July 11, 2008, and resulted in deposit losses in excess of the $100,000 insured limit for thousands of depositors. *See, e.g.*, William Heisel, *IndyMac's Shuffle Ran Over Depositors*, L.A. TIMES, Feb. 28, 2009, at C1 (stating that an estimated 10,000 IndyMac depositors lost $270 million in deposits); Damian Paletta, Lingling Wei & Ruth Simon, *IndyMac Reopens, Halts Foreclosures on Its Loans*, WALL ST. J., July 15, 2008, at C1 (stating that while the FDIC normally insures up to $100,000 per depositor, nearly $1 billion of IndyMac's deposits were uninsured). At this writing, there are large numbers of banks on the brink of similar collapses, thanks largely to the home-mortgage-loan and ensuing foreclosure debacle of the last several years. *See* Damian Paletta & David Enrich, *Banks on Sick List Top 400: Industry's Health Slides as Bad Loans Pile Up; Deposit-Insurance Fund Shrinks*, WALL ST. J., Aug. 28, 2009, at A1 (stating that 416 banks were on the FDIC's "problem list"); Press Release, FDIC, Statement by FDIC Chairman Sheila Bair at the Quarterly Banking Profile Press Conference (Aug. 27, 2009), http://www.fdic.gov/news/news/press/2009/pr09_qbp.html (stating that the number of "problem" institutions is at a fifteen-year high).

107. For example, an employer's or plan administrator's failure to fiscally respond to the fluctuation of interest rates can lead (and has led) to the underfunding and termination of employer-sponsored pension plans. In particular, when interest rates fall, employers sponsoring private pension plans assume a lower rate of return on the money invested in the plan. Consequently, the employer must invest more money into the plan to make up for lower expected returns. Without doing so, it is unlikely that a plan will meet its prescribed level of expected future funding. However, many employers have failed to conform their investments with the potential losses attributable to lowered interest rates, resulting in failing or underfunded pensions. To protect against the risk of underfunded or failed pension plans, the Employment Retirement Income Security Act (ERISA) provides a government-

Social Security, on the other hand, is backed by the power to tax as well as by the political will of American workers and retirees who have insisted, through the political process, on Congress respecting their basic entitlement to benefits, even in the face of the dire financial emergency of 1983 and the strong efforts of the second President Bush in 2005.[108] There are also two additional critical differences between the public and private entitlements. First, Social Security guarantees a level of income in old age, whereas private investment assets can guarantee only equity ownership, not actual income in retirement—it is the value and liquidity of the asset when income is needed that matters, not the security of one's right to the asset.[109] Employer-sponsored defined-benefit plans also guarantee a level of income, but only if they are adequately funded, which is far from a sure thing.[110] In any event, defined-benefit pension plans have been slowly dying over the last thirty years so that few American workers will be able to count on them

---

run insurance scheme—the Pension Benefit Guaranty Corporation (PBGC)—which insures nonforfeitable retirement pension benefits for single-employer, defined pension plans. Nevertheless, the amount of coverage that PBGC can guarantee is limited under ERISA § 4022. The principle limitation involves a cap on the maximum benefits that PBGC will pay. Under § 4022, benefits payable to a participant under a plan are guaranteed only to the extent that they do not exceed the statutory maximum. Under the single-employer program, the limit is adjusted annually based on changes in the Social Security contribution and benefit base. In addition, PBGC does not guarantee benefit payments that exceed the amount of a participant's accrued plan benefit payable at normal retirement age. *See* Employee Retirement Income Security Act of 1974, 29 U.S.C. § 1322 (2006). For an analysis of the current probability of pension-fund failures from a practitioner's perspective, see Alex D. Moglia, *Underfunded Pension Funds: A Ticking Time Bomb for Companies and Taxpayers*, AM. BANKR. INST. J., Oct. 2009, at 46.

108. In a poll conducted in May 1999 by Princeton Survey Research Associates,

> [fifty-eight] percent of Americans favored a system that would include private accounts. The idea was opposed by 33 percent. But the results were almost precisely the opposite when the same survey asked workers to choose between a program that guaranteed a monthly benefit based on lifetime earnings, as under the current system, and a program that would allow individual investment in the market without a guarantee. Given that choice, 59 percent favored the guaranteed payment, while 33 percent backed private investment.

Richard W. Stevenson, *Bush to Advocate Private Accounts in Social Security*, N.Y. TIMES, May 1, 2000, at A1, *available at* http://www.nytimes.com/2000/05/01/us/2000-campaign-issues-bush-advocate-private-accounts-social-security.html?pagewanted=all. An Associated Press/Ipsos poll on Social Security showed consistent near 60% disapproval of President Bush's handling of Social Security, and a CNN/Opinion Research Corporation poll showed that by October 2008, there was over 60% disapproval of private investment of Social Security taxes. *See* PollingReport.com, Social Security, http://www.pollingreport.com/social.htm (last visited Apr. 20, 2010).

109. I have discussed this point thoroughly in my previous work. *See* Dilley, *supra* note 104; *see also* Befort, *supra* note 71, at 963–65.

110. *See* Befort, *supra* note 71, at 950–51 (stating that "[a]n increasing number of defined benefit plan sponsors fail to fulfill their pension promises").

in the future.[111] Second, the costs of the public-entitlement promise are spread across the entire working population and guaranteed to be collected through the public taxing power, while the private-entitlement promise depends on the economic solvency and well-being of individual workers, of their employers, or on the market's valuation of assets at the time of the worker's retirement, none of which are either guaranteed or predictable.[112]

Thus, the individual worker's contribution to Social Security, which most Americans today appear to understand goes to pay for current benefits rather than being saved for their future retirement, establishes a politically formidable, if not a legally unchangeable, entitlement to benefits. The function of the payroll tax in forging a strong connection between workers and the system is therefore, as Roosevelt predicted, an essential element of the program.[113] The extent to which the dedicated FICA tax contribution cements worker support for Social Security is hard to measure, but most polls of taxpayers over the past several decades have shown that the best tolerated tax has always been the payroll tax.[114] The assumption may be that workers feel they know what they pay the Social Security the tax for, unlike the income tax which funds the more amorphous

111. For a discussion of the declining importance of defined benefit plans, see JOHN H. LANGBEIN, SUSAN J. STABILE & BRUCE A. WOLK, PENSION AND EMPLOYEE BENEFIT LAW 58 (4th ed. 2006).

112. Yale economist Robert Shiller continues to be the leading and most correct analyst of both housing prices and stock market cycles. *See, e.g.*, STANDARD & POOR'S, S&P/CASE–SHILLER HOME PRICE INDICES 2008, A YEAR IN REVIEW, http://www2.standardandpoors.com/spf/pdf/index/Case-Shiller_Housing_Whitepaper_YearinReview.pdf (offering an analytical description of the housing market recession); *see also* Robert Shiller, Online Data, http://www.econ.yale.edu/~shiller/data.htm (last visited Apr. 20, 2010) (describing available data sets on consumer price indices from 1871 to the present).

113. This is not to say that the payroll tax must be the only source of financing—that question is outside the scope of this Article, but a future article currently in progress entitled *Dedicated to the Ones We Love* will explore the broader question of the payroll tax principle in general.

114. For instance, when asked about their income taxes in a USA TODAY/CNN/Gallup poll taken September 10–14, 1999, "68% of respondents said their income taxes are too high, compared with 29% who said they are about right or too low. When asked about their Social Security taxes, 43% said they are too high, while 49% said they are about right or too low." Owen Ullmann, *Payroll Tax Relief Isn't High Priority*, USA TODAY, Sept. 24, 1999, http://www.usatoday.com/money/wealth/saving/msw126.htm. Similarly, according to the Tax Foundation's 2006 Annual Survey, only 14% of those surveyed (15% in 2005) found the Social Security payroll tax to be the least fair, while 25% (26% in 2005) found the federal income tax to be least fair and 31% (30% in 2005) found the federal estate tax to be least fair. *See* TAX FOUND., 2006 ANNUAL SURVEY OF U.S. ATTITUDES ON TAX AND WEALTH (2006), http://www.taxfoundation.org/files/survey_topline-20060405.pdf (similarly demonstrating that Americans view federal income taxes as less fair than Social Security taxes).

"government."

However, even if the idea of contributing payments based on wages is a fundamental element of Social Security, the contribution principle alone still provides little guidance on the appropriate level of the wage base—programmatic imperatives essentially grounded in benefit-coverage goals do not necessarily dictate the flat taxation of exactly 90% of all wages in the economy. Instead, an examination of tax justice and tax policy principles may provide some adequate guidelines to help properly assess where to draw the line between appropriate and excess wage-base levels.

2. Contribution Base from a Tax Perspective

Discussion of Social Security financing has generally been relegated to the province of economists, who apply almost exclusively efficiency-based critiques to tax systems and who have mainly analyzed the payroll tax in the context of its possible wage-depression effects as well as its perceived inability to assure the long-term financing of Social Security benefits, with a view to building support for privatizing the system.[115] In contrast, my focus in this Article is the optimal level for the contribution and benefit base from the perspective of the Social Security benefit program, as well as from the perspective of the tax system, leaving the comparatively simple questions about adequate financing for Social Security until the end. Disregarding for the moment the question of whether the payroll tax itself is the best financing option for Social Security at all, we can begin a tax policy analysis of the FICA wage-base limit by examining a couple of issues that frequently crop up in economists' discussions of the payroll tax—distributional effects (i.e., the regressive nature of the tax and the base) and the somewhat amorphous notion of "tax fairness."

a. Distributional Considerations

The payroll tax is commonly criticized as the most regressive aspect of the U.S. tax system because it taxes all workers at the same rate—6.2% of wages for the Old-Age, Survivors, and Disability Insurance (OASDI) program and 1.45% of wages for Health Insurance, the Medicare portion —regardless of their level of earnings and, for the OASDI portion, only up

115. *See* Lewis D. Solomon & Geoffrey A. Barrow, *Privatization of Social Security: A Legal and Policy Analysis*, 5 KAN. J.L. & PUB. POL'Y 9, 13–15 (1995); Martin Feldstein, *The Missing Piece in Policy Analysis: Social Security Reform* 24–29 (Nat'l Bureau of Econ. Research, Working Paper No. 5413, 1996); *see also* PETER J. FERRARA, SOCIAL SECURITY: THE INHERENT CONTRADICTION 311 (1980); Martin Feldstein, *Toward a Reform of Social Security*, 40 PUB. INT. 75 (1975) (serving as one of the earliest salvos); Sylvester J. Schieber & John B. Shoven, *Social Security Reform: Around the World in 80 Ways*, 86 AM. ECON. REV. 373, 376 (1996).

to the wage base limit.[116] This "flat tax" contrasts with the progressive income tax which taxes higher income taxpayers at a higher rate on their top marginal income.[117] As a result, while workers making $50,000 per year and $150,000 per year, respectively, will pay the same nominal payroll tax rate, their income tax rates will vary considerably, with the first worker paying a top marginal rate of 25% on the last $16,050 of her income, while the second will pay a top marginal rate of 28% on the last bracket of income.[118] Under progressive rate theory, the FICA tax result is perverse—the $50,000 a year worker, with fewer dollars remaining after paying for necessities than the $150,000 a year worker, ends up paying the same nominal rate on those excess dollars.[119]

As one would expect, the base and rate intersect to aggressively increase the regressivity of the tax. Since the rate on wages above the base is zero, the higher the wages, the lower the effective payroll tax rate on the taxpayer's entire earnings.[120] As a result, for example, someone reporting $300,000 per year in wages is paying an effective OASDI payroll-tax rate of just over 2% on all her wages, while someone earning $106,800 (the base for 2010) or less, is paying the statutory—and effective—rate of 6.2%. This is clearly no longer a flat tax rate; it is a pyramid, the complete inverse of the progressive income tax in that the more someone earns, the less payroll tax he pays.

Of course, earnings above the wage base are not counted for benefit purposes, so the regressive effect of the base could be rationalized as being essentially irrelevant from a Social Security perspective. It might be argued that those wages above the base are outside the closed-contribution and benefit system, and that including them for purposes of calculating an effective payroll tax rate distorts the contribution–benefit relationship. Moreover, because the benefit structure is mildly progressive in that lower wage workers receive benefits that are higher than a strictly proportional

116. *See* Soc. Sec. Admin., Trust Fund Data, Social Security & Medicare Tax Rates, http://www.socialsecurity.gov/OACT/ProgData/taxRates.html (last visited Apr. 20, 2010).

117. *See* I.R.C. § 1 (2006) (detailing the different tax brackets under the progressive income tax); Rev. Proc. 08-66, 2008-45 I.R.B. 1107 (providing the 2009 cost-of-living adjustments for the progressive tax).

118. Rev. Proc. 08-66, 2008-45 I.R.B. 1110, tbl.3. This rate example, of course, does not include the effects of the rate differential between ordinary income and capital gains rates.

119. *See, e.g.*, Jason Bordoff & Jason Furman, *Progressive Tax Reform in the Era of Globalization: Building Consensus for More Broadly Shared Prosperity*, 2 HARV. L. & POL'Y REV. 327, 331–41 (2008); Geier, *supra* note 10, at 821; Vada W. Lindsey, *The Widening Gap Under the Internal Revenue Code: The Need for Renewed Progressivity*, 5 FLA. TAX REV. 1, 13–14 (2001).

120. *See generally* Deborah A. Geier, *The Payroll Tax Liabilities of Low- and Middle-Income Taxpayers*, 106 TAX NOTES 711 (2005) (discussing the burden of payroll taxes on low-income taxpayers).

benefit formula would produce, it can be argued that the progressively structured benefits are a trade-off for a regressive tax structure.[121]

One problem with the progressive benefit–regressive contribution trade-off rationale, however, is that it relies on viewing the payroll tax as not simply a premium-paying mechanism but as an actual payment for benefits. The trade-off rationale posits that while low-wage workers pay a higher percentage of their wages than do workers with wages above the wage base, they get more for those tax payments in the form of disproportionately higher benefits.[122] However, as previously discussed, the amount of payroll taxes paid by or on behalf of any worker has no connection to the benefits she eventually receives; benefit entitlement is earned, not purchased with taxes.[123] While contributory financing is a fundamental principle of social insurance in the United States, contributions to cement basic entitlement to ultimate benefits on the one hand and payments as a quid pro quo for specific benefits on the other are not the same thing. If benefits are not functionally related to taxes, then the progressive benefit structure must be—and is—justified on Social Security program grounds, while the tax system has to be viewed as a financing mechanism, one of many possible ones, whose distributional effects should be critiqued based on consistent tax principles as well as Social Security program needs.

b. Tax Fairness—Who Benefits? Who Pays?

From a distributional perspective, the FICA wage base would be most equitable if it covered all wages, resulting in an effective wage tax identical to the statutory payroll-tax rates. Would this be a fair outcome, though, from a theoretical tax perspective? Standard theories of tax justice have seldom been discussed in connection with payroll taxes generally, or the wage base specifically, except as an example of a somewhat crude version of the "benefit" theory of tax equity:

> The prevailing modern view is that the quasi-exchange version of the benefit principle should be cabined off to government activities that involve citizen use of government property, facilities, and services. . . . Arguably,

121. *See, e.g.*, Posting of the Economist to Free Exchange, http://www.economist.com/blogs/freeexchange/2009/04/are_payroll_taxes_regressive (Apr. 14, 2009, 18:00 EST) (discussing whether social security taxes are regressive or progressive); 2009 BD. OF TRUSTEES REPORT, *supra* note 8; *see also Investing in the Private Market: Hearing Before the Subcomm. on Social Security of the H. Comm. on Ways and Means*, 106th Cong. 95 (1999) (statement of John Mueller, Senior Vice President and Chief Economist, Lehrman Bell Mueller Cannon, Inc.).

122. *See* Posting of the Economist to Free Exchange, *supra* note 121.

123. *See* 42 U.S.C. §§ 402(a), 413(a)(2)(A), 414(a) (2006).

> government insurance, such as Social Security, Medicare, workers' compensation, and unemployment compensation, also might be included within the benefit principle to the extent that beneficiaries are limited to those who (directly or indirectly) pay appropriate amounts of "premiums" into the system relative to anticipated benefits. On the other hand, these programs are mandatory (and therefore operate to override preferences to self-insure and avoid risks) and often entail redistribution because of the premium and/or benefit structure.[124]

This view of Social Security taxes as a direct quid pro quo for Social Security benefit payments when (and if) they are eventually received by the individual taxpayer is the basis for much of the modern conservative "money's worth" critique of Social Security.[125] Raising the wage base to mitigate the harsh regressive distributional effects of the payroll tax would only exacerbate the perceived inequity of high-wage workers paying FICA taxes "in exchange" for future benefits that might be less than the equivalent amount that investments would have earned for them in the private markets.[126]

However, there are broader articulations of the benefit theory of taxation that provide an alternate view of Social Security taxes—a view that is certainly more consistent with Social Security's founding principles and provides some support for the current tax and benefit structure. While the "new benefit theory," as Professor Dodge has labeled it,[127] does not provide specific, measured results in terms of benefits for individual taxpayers, the lack of a direct quid pro quo may be the whole point, as others have suggested:

> To me, the bottom-line question is: *How should the costs of maintaining a regulated capitalist economy, with its laws of supply and demand that create wealth, be allocated among the members of the population?* In my view, essentially all tax revenue goes toward paying the costs of maintaining a regulated capitalist economy . . . . Therefore, the costs of paying for that system should be allocated across the population at least in proportion to the money benefits extracted under that

---

124. Dodge, *supra* note 76, at 405–06.

125. For a succinct statement of Feldstein's view that Social Security is a bad investment, see MARTIN FELDSTEIN, CATO INST., PRIVATIZING SOCIAL SECURITY: THE $10 TRILLION OPPORTUNITY (1997), http://www.cato.org/pubs/ssps/ssp7.html.

126. *See generally id.* I have previously criticized this "rate of return" argument, which of course has not stopped anyone from making it. *See* Dilley, *supra* note 69.

127. Professor Dodge described the "new benefit theory" as follows:

> The new and expanded version of the benefit principle purports to be a norm of tax fairness that avoids the measurement problem inherent in the quasi-exchange version of the benefit principle by postulating that the measure of a person's benefit from government is none other than his or her financial (as opposed to psychic) well-being.

Dodge, *supra* note 76, at 406 (footnote omitted).

> system. The person earning $500,000 per year is able to do so *only because* he or she lives in a regulated capitalist system and can exploit the market to sell products or services.[128]

Professor Geier's version of the benefit theory of taxation echoes the mission statement, as it were, of Social Security as restructured in the 1939 Amendments, which was, as discussed above, to provide a benefit for society as a whole, not simply to pay benefits to ensure that individual workers were spared an indigent old age.[129] The case for raising the wage base to cover some or all of the top 10% of earnings is more persuasive if those earners are seen as paying their FICA taxes not just for their individual benefits but also for prevention of the social disintegration that would very possibly flow from allowing a generation of elderly people to slide even further into poverty than is already the case.[130] The "money's

---

128. Deborah A. Geier, *Incremental Versus Fundamental Tax Reform and the Top One Percent*, 56 SMU L. REV. 99, 119–20 (2003) (footnotes omitted). Geier points out, in another setting, that

> [f]our commonly invoked "tax justice" fairness norms are: (1) the equal-sacrifice principle; (2) the principle that persons should sacrifice to government according to the benefits received from government; (3) the principle that persons should sacrifice to government according to their standard of living or well-being (what economists call 'utility'); and (4) the principle that persons should sacrifice to government according to their respective abilities to pay (meaning material wealth).
>
> . . . .
>
> In short, current tax policy "fairness" debates typically involve only arguments concerning whether the "standard-of-living" norm or the "ability-to-pay" norm is more persuasive. The "benefit" norm is typically viewed as a relic of a simpler time.
>
> . . . .
>
> It seems to me that the mere exploitation of our economic system to earn income (including foreign income, which can be earned only because of the U.S. legal and economic environment that allows and supports foreign direct and indirect investment)—whether investment income or active business income—is sufficient under a reconstituted benefit theory to justify (in general) income taxation.

Deborah A. Geier, *Time to Bring Back the 'Benefit' Norm?*, 102 TAX NOTES 1155, 1155, 1157 (2004) (internal quotation marks omitted) (footnote omitted). The same assessment seems to me to clearly apply to the Social Security tax as well.

129. *See supra* Part II.B.

130. According to the standard U.S. poverty level measures, the U.S. elderly as a group are at about the same level of poverty as the population at large; however, this measure has been criticized as being out of date, ignoring the much higher costs of medicine and medical care that the elderly disproportionately face. Recently, New York City Mayor Bloomberg's staff developed an alternative measure under which about one-third of New York's elderly would be considered to be living in poverty. *See* Cara Buckley, *City Refines Formula to Measure Poverty Rate*, N.Y. TIMES, July 14, 2008, at B2, *available at* http://www.nytimes.com/2008/07/14/nyregion/14poverty.html; *see also* CARMEN

worth" argument, from this perspective, becomes largely irrelevant in determining the appropriate level of Social Security taxes, including the wage base.

c. Optimal Base Level

What then would be the optimal level for the wage base, given that neither program principles nor tax policy principles provide any specific guidance? As long as the contribution and benefit base is unified, clearly the minimum level should be the traditional 90% of wages in the national economy. The current indexation mechanism is not effectively maintaining that level, as pointed out above, because the increasing disparity between high-income earners and everyone else has increased dramatically over the last two decades.[131]

Raising the contribution and benefit base for both tax and benefit calculation purposes to $150,000, which in 2006 would have covered 90% of all earnings, would at that point have eliminated about 40% of the long-term revenue shortfall for the system.[132] An ad hoc increase of this nature, however, would mean an immediate tax increase for what might be called middle-upper income wage earners and would undoubtedly be wildly unpopular with that group of likely voters, making any such proposal difficult to enact in the absence of an immediate financing crisis along the lines of the 1983 financing situation.[133]

DENAVAS-WALT, BERNADETTE D. PROCTOR & JESSICA C. SMITH, U.S. DEP'T OF COMMERCE, INCOME, POVERTY, AND HEALTH INSURANCE COVERAGE IN THE UNITED STATES: 2008, at 2, 18 (2009), http://www.census.gov/prod/2009pubs/p60-236.pdf (noting that in 2007 the number of poor elderly would be 13 million higher if social security payments were excluded from income, more than quadrupling the number of poor elderly, and that the elderly represent 12.6% of the total U.S. population and 9.2% of the poor population).

131. *See* MULVEY & WHITMAN, *supra* note 41, at 1.

132. *Id.* at 13. The authors put forth a few options for raising or eliminating the taxable earnings base. For example,

> [o]ne proposal would slowly raise the taxable wage base for both employers and employees to cover 90% of all earnings and credit these taxes to allow individuals to receive correspondingly higher benefits. In 2006, it was estimated that a cap of $171,600 would roughly cover 90% of wages. Under this option, benefits at retirement for high earners would also rise. These changes would have a net positive impact on the Social Security Trust Funds. . . . Raising the wage base to 90% would eliminate 43% of the long-range financial shortfall—extending the Trust Funds' exhaustion date to 2044. To achieve solvency for the full 75-year projection period under this option, the total payroll tax rate would have to be raised by an additional 1.09 percentage points (from 12.40% to 13.49%) or other policy changes would have to be made to cover the shortfall.

Id. (footnotes omitted).

133. *Id.* For a discussion of the importance of the 1983 Social Security financial crisis to

One widely circulated proposal for gradually raising the base to the 90% target level was made by the late Robert Ball, former Commissioner of Social Security and a widely respected voice in social insurance analysis for fifty years.[134] The Ball proposal was designed to resolve any long-term financing shortfalls in Social Security revenues by addressing the wage-base issue as well as adding additional dedicated financing sources to the current mix of payroll taxes and income tax revenues from taxation of Social Security benefits. Ball proposed a 2% per year increase in the base limit, in addition to already-scheduled automatic-indexing increases over several years, to eventually reach 90% of covered payroll.[135] This is an example of one possible way to gradually phase in an increase while minimizing the immediate effect on workers who now have at least some wages above the current base limit, a clear political advantage. By itself, this change would do very little to address any near-term financing concerns, but the base would eventually be brought up to its traditional target level with only a very small yearly impact in increased taxes for each worker.

However, even with such an ad hoc increase, whether all at once or over a long period of time, the base would still be playing catch-up in the future, as the current measure for automatic increases in the wage base, which uses average earnings to develop the index figure, misses wages at the top.[136] A further suggestion, therefore, has been made to change the base for the index calculation from average wages to aggregate earnings, which would stabilize the 90% level, preventing further leakage resulting from increasing earnings disparity.[137] Such a technical change in the definition of the wage base would clearly be a correction within the scope and intent of current law, although the 90% of aggregate earnings target would have to be reached by one or more ad hoc increases in the current base before any new indexing measure would be effective. The effect such a change would have on Social Security financing would depend on how quickly the new measure was phased in, but modifying the base calculation to use a more expansive measure would directly address at least some fairness issues of the current base by making future automatic increases more likely to capture wages at the top. Such a change would clearly improve the function of the contribution and benefit base in capturing most earnings for both benefit

---

the enactment of rescue legislation, see ALTMAN, *supra* note 36.

134. *See* Century Found., Social Security Reform Check List #1: The Robert M. Ball Plan (1999), http://www.socsec.org/facts/Check_Lists/checklist1.pdf.

135. *Id.*

136. Soc. Sec. Admin., National Average Wage Index, http://www.ssa.gov/OACT/COLA/AWI.html (last visited Apr. 20, 2010).

137. *See* MULVEY & WHITMAN, *supra* note 41, at 8.

and tax purposes. Still, however, from the perspective of the ordinary taxpayer, any limit on earnings subject to FICA is likely to seem unfair.

III. THROUGH THE DOUGHNUT HOLE?

Since 1935, the concept of the contribution and benefit base has remained essentially unchanged except for regular increases in amount. Changing the measure of contribution and benefit base increases from average to aggregate wages could be viewed as one way to redesign the base, but it would essentially be a correction in the measure used in the current indexing system, not a fundamental redesign. The innovative aspect of the doughnut hole suggestion is that it would actually redesign the base by leaving the current indexed base as it is and instituting a second-tier base applicable to workers with yearly earnings at or above $250,000 (presumably indexed), thus creating a gap in which earnings above $106,800 (the limit for 2010) but below $250,000 would not be subject to FICA at all.[138]

President Obama's economic advisors said during the 2008 campaign that the $250,000 level for his wage-base proposal is not necessarily set, and that the increase in the wage base on which FICA taxes are assessed might or might not be linked to an increase in benefit base on which benefits are calculated.[139] Obviously, more net revenue would be raised if the wage base were raised while the benefit base remained set and increased only as average wages increase. Even if additional benefits were to be paid as a result of this redesigned wage base, however, the proposal would resolve a substantial part of the currently projected long-range deficit for the OASDI part of Social Security.[140]

---

138. *See supra* note 9 and accompanying text for discussion of the doughnut proposal.

139. The formula for determining the "contribution and benefit base" is determined each year by applying a formula set forth in § 430. 42 U.S.C. § 430 (2006). For 2008, the contribution base was $102,000 with a tax rate of 6.2%. In 2009, the contribution base was $106,800. Obama's advisors have also suggested that the tax rate above the doughnut hole would not be the full FICA rate, but rather a 3–4% tax. *See supra* notes 8–9 and accompanying text. This kind of "surtax" model is discussed below.

140. As specialists in this area have discussed,

> Raising or eliminating the cap on wages that are subject to taxes could reduce the long-range deficit in the Social Security Trust Funds. For example, if the maximum taxable earnings amount had been raised in 2005 from $90,000 to $150,000—roughly the level needed to cover 90% of all earnings—it would have eliminated roughly 40% of the long-range shortfall in Social Security. If all earnings were subject to the payroll tax, but the base was retained for benefit calculations, the Social Security Trust Funds would remain solvent for the next 75 years. However, having different bases for contributions and benefits would weaken the traditional link between the taxes workers pay into the system and the benefits they receive.

MULVEY & WHITMAN, *supra* note 41. In effect, a decoupling of the wage base from the

The key feature of this proposal, the gap in imposition of the payroll tax for those making over the current base but less than $250,000 per year, bears a strong resemblance to what is popularly known as the doughnut hole in the Medicare prescription-drug program, a gap in coverage that was built into the reimbursement structure in the legislation that created the program in 2004. The Medicare drug program covers drug expenses up to $2,810 a year and then does not cover drug expenses above that until the individual's expenses for the year reach $4,550 (these are the limits for 2010), after which reimbursement begins again.[141] While this doughnut hole, as it came to be known, may be effective in limiting program costs while still providing additional coverage for those with very high drug expenses, it has proven to be confusing and worrying for elderly Medicare beneficiaries.[142] A tax doughnut hole is unlikely to create that level of anxiety, but other problems might make it equally unpalatable.

Clearly the main reason for the proposed Social Security base gap is political palatability, as it would exempt a large group of upper-middle-class voters (or at least a large number of people who think of themselves as middle class) from a tax increase.[143] It is unclear whether there is any real rationale for the doughnut hole beyond the politics of distributional effect. But if we take this proposal seriously from a policy perspective, despite its many flaws, perhaps there are elements that could be incorporated into

---

benefit base was accomplished for the Medicare program in 1993 when the wage base was eliminated for the Health Insurance (HI) portion of FICA but not for the cash benefit retirement, survivor, and disability programs. The base elimination in Medicare raised few programmatic concerns because Medicare benefits are in the form of payment of medical expenses and are not connected to preentitlement earnings or taxes paid.

141. *See* Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Pub. L. No. 108-173, 117 Stat. 2066 (codified in scattered sections of 42 U.S.C. (2006)*; see also* CTRS. FOR MEDICAID AND MEDICARE SERVS., CMS LEGISLATIVE SUMMARY: MEDICARE PRESCRIPTION DRUG, IMPROVEMENT, AND MODERNIZATION ACT OF 2003, PUB. L. NO. 108-173 (2004), http://www.cms.hhs.gov/MMAUpdate/downloads/PL108-173summary.pdf; David Pratt, *The New Medicare Part D Prescription Drug Benefit*, 17 ALB. L.J. SCI. & TECH. 337 (2007). For long-range estimates of the plan, see 2009 BD. OF TRUSTEES REPORT, *supra* note 8. "Under the intermediate assumptions, the annual balance is positive for eight years (through 2016) and is negative thereafter. This annual deficit rises rapidly, reaching 2 percent of taxable payroll by 2024, and continues rising generally thereafter, to a level of 3.87 percent of taxable payroll for 2083." *Id.*

142. *See* Posting of Victoria E. Knight to Wall Street Journal Health Blog, http://blogs.wsj.com/health/2008/11/18/seniors-still-mystified-by-medicares-doughnut-hole/tab/article/ (Nov. 18, 2008, 16:39 EST).

143. *See, e.g.*, Posting of David Leonhardt to Economix, http://economix.blogs.nytimes.com/2009/08/21/what-about-the-upper-middle-class (Aug. 21, 2009, 11:55 EST) (discussing views on the effect Obama's Social Security proposal would have on the upper-middle class); GERALD PRANTE, TAX FOUND., NEW CENSUS DATA ON INCOME GIVES A WELCOME DOSE OF FACT CHECKING TO "MIDDLE-CLASS" RHETORIC, Sept. 11, 2007, http://www.taxfoundation.org/files/ff102.pdf.

other possible redesigns for the base that achieve some of the goals of the doughnut hole without creating the same problems. The common thread of much of the substantive critique of the proposal, however, is the transformation it would produce in the character of the payroll-tax system, making it in many ways, mostly unfortunate, more similar to the income tax.

Sadly, many commentators criticizing the proposal appear to have little understanding of Social Security's actual underlying principles. Therefore, it is necessary on the one hand to analyze and critique the doughnut hole idea from a genuine Social Security and tax policy perspective, while on the other to debunk at least some of the commentary that cynically or mistakenly uses distorted descriptions of Social Security core principles to score political or ideological points. Much of the rhetoric in this commentary amounts to thinly disguised attacks on Social Security itself, taking advantage of the doughnut hole proposal to renew a decades-old challenge to the concept of social insurance.[144]

A. Substantive Critique

The principal purpose of the doughnut hole proposal is to increase Social Security revenues while making the payroll tax less regressive in impact, all in a politically palatable manner. As discussed above, raising the wage base to cover at least 90% of wages in the economy is not only consistent with program purposes, it is probably a necessity given the great disparity in wages and income in the American economy that has caused the wage base to lag behind that target.[145] But the Obama proposal would import distributional equity into the payroll tax to an unprecedented extent while still holding harmless a large segment of politically influential wage earners.[146]

The underlying problem is, of course, that the payroll tax is essentially a flat tax imposing a proportional tax burden up to the wage base, with a regressive effect resulting from the lack of taxes on wages above the base. All other factors being even (same base, same definition of wages, etc.), a

144. *See infra* notes 160–75 and accompanying text.

145. For an analysis of the growing gap in income between rich and poor, see ARLOC SHERMAN, CTR. ON BUDGET & POLICY PRIORITIES, INCOME GAPS HIT RECORD LEVELS IN 2006, NEW DATA SHOW: RICH-POOR GAP TRIPLED BETWEEN 1979 AND 2006 (2009), http://www.cbpp.org/cms/index.cfm?fa=view&id=2789.

146. *See, e.g.*, Posting of Glenn Kessler to Trail, http://blog.washingtonpost.com/44/2008/06/13/obama_clarifies_social_securit.html?hpid=topnews (June 13, 2008, 11:14 EST) (noting that Obama asserted that his plan "'can extend the promise of Social Security without shifting the burden on to seniors' while leaving 'absolutely no change' in taxes for 97 percent of Americans").

flat-rate payroll tax simply cannot produce a progressive-rate tax result, even though raising the base would mitigate the regressive effect of the current tax. The doughnut hole proposal is an attempt to go a step further by essentially imposing a new tax on earnings above $250,000 with no additional limit; however, while it would increase the overall progressivity of the payroll tax, it would also create somewhat perverse effects and incentives for earnings below $250,000.[147]

First, the open question of whether the proposal would apply to the base for benefits as well as for taxes has no good or obvious answer. The overall issue of splitting the two bases as a general proposition will be discussed below, but in looking at the question strictly in the context of the doughnut hole proposal, obvious equity issues would arise if earnings above $250,000 were counted for benefit purposes but those between the current base and $250,000 were not. For one thing, including those top earnings in the benefit calculation would fly in the face of one principal rationale for the limit on the benefit base—that workers at the very top of the earnings scale should not get additional benefits based on those earnings from a publicly funded social insurance program.[148] That objection might be overcome by the programmatic purpose discussed earlier, that the wage base should be as inclusive as possible in order to properly reflect the ups and downs of a worker's record who might be at the top for a few years and near the bottom for others.[149]

However, there is no real answer for the complaints of those with earnings in the gap about getting less in the way of benefit accrual than those with much higher earnings. This result, while probably not increasing significantly the overall replacement rates for the highest earners, would be difficult to explain and very likely would be perceived as inequitable by those with earnings in the gap and probably by those with earnings below the gap as well. The political power of the earnings-based benefit principle would certainly be diminished by a benefit-accrual structure that to most people would appear as capricious and unfair as the current Medicare prescription-drug coverage gap does.[150]

Second, the gap would complicate the payroll tax for both workers and employers, negating one great advantage of any flat tax—simplicity and

147. *See* discussion *supra* Part II.B.2.

148. *See supra* note 89 and accompanying text (noting that the benefits base reflects the concept of limiting upper-income benefits out of public necessity).

149. It might also be possible to amend the benefit formula, currently a three-bracket structure, to add a fourth tier to provide 10% or 5% of AIME above a certain level, instead of the 15% provided currently, so as to minimize the effect of the highest earnings on the ultimate benefit.

150. *See* Pear, *supra* note 29 (describing disdain toward the current Medicare doughnut hole policy).

ease of collection. Since the payroll tax is essentially administered and collected by employers, the employers would bear most of the burden of determining when workers would hit the first contribution base limit so as to stop withholding at that point, only to have to begin withholding again later in the year after the $250,000 floor for the next tier of the base was reached.[151] From the employee side, the complication would be more a matter of understanding and acceptance of a more complex structure that would be difficult to explain. Again, the example of the Medicare prescription is instructive—regardless of the desirability of the distributional consequences of the program's coverage gap, it is very difficult to explain, let alone justify, to those directly affected.

Granted, the elderly population affected by the drug plan may be less able to absorb the nuances of the changes in Medicare than workers who are younger and who have less immediately at stake than eighty-five-year-old widows terrified of being unable to pay for their medications. Nonetheless, complexity is one of the most widely perceived negatives of the income-tax system in contrast to the simplicity and relative ease of the payroll tax, and the more the latter is modified to look more like the former, the more resistance to any change in the payroll tax is likely to be created among taxpayers generally.[152]

Third, a gap in the wage base would in all likelihood exacerbate already-existing perverse incentives to game the payroll-tax system by keeping compensation out of the payroll-tax box. The current contribution and benefit base limit for the OASDI portion of FICA generally limits its effects to employees who have few opportunities to change the way they are compensated in order to avoid the payroll tax.[153] However, when the wage

151. The task would be made even more difficult by gaming opportunities, which are discussed below. *See infra* notes 153–57 and accompanying text.

152. "The complexity of our tax code breeds a perception of unfairness and creates opportunities for manipulation of the rules to reduce tax. The profound lack of transparency means that individuals and businesses cannot easily understand their own tax obligations or be confident that others are paying their fair share." PRESIDENT'S ADVISORY PANEL ON FED. TAX REFORM, SIMPLE, FAIR & PRO-GROWTH: PROPOSALS TO FIX AMERICA'S TAX SYSTEM 1 (2005), http://www.taxfoundation.org/UserFiles/Image/Blog/Executive_Summary.pdf. *See generally* Stanley S. Surrey, *Complexity and the Internal Revenue Code: The Problem of the Management of Tax Detail*, 34 LAW & CONTEMP. PROBS. 673 (1969) (discussing the structural complexity of the income-tax collection procedure).

153. Most employees do not control the timing of their wages and are governed by their employer's decisions on how often to pay them (weekly, bi-weekly, monthly, etc.) for services performed. In many states, in fact, it is illegal for employers to agree to defer wages or salaries for any period. *See, e.g.*, Stanton v. Lighthouse Fin. Servs., Inc., 621 F. Supp. 2d 5, 14–16 (D. Mass. 2009) (ruling that an agreement to defer the payment of salary violated the Massachusetts Weekly Wages Act and was therefore void). Executives signing contracts for deferral of a portion of future compensation under nonqualified arrangements are not

base was eliminated for Medicare's hospital-insurance portion of the tax in 1993, highly compensated executives and professionals suddenly faced an additional 1.45% tax on all their earnings over the regular base.[154] This relatively sudden tax increase on upper level wage compensation was likely a factor encouraging the development of a plethora of deferred compensation and equity compensation techniques that allow highly paid employees to defer (and sometimes permanently evade) income and payroll taxation on substantial portions of their income from work.[155]

Even though Congress has acted recently to try to rein in the most egregious of abusive deferred-compensation techniques,[156] the ability of highly paid executives to structure their compensation arrangements seems not to have been substantially limited.[157] These techniques would therefore be available for evasive possibilities for these taxpayers in the event of any large increase in the payroll tax. The introduction of a gap in the earnings to which an additional 6.2% of payroll taxes would apply adds more wrinkles to the possibilities and also increases the universe of workers with more incentive to distort the character of their compensation. While workers making less than $250,000 generally have less control over their compensation than executives and professionals at higher compensation levels, compliant employers, anxious to avoid the employer share of the FICA tax, might well be helpful in keeping compensation characterized as wages under the $250,000 limit. This kind of gaming strategy is more akin to the kind of manipulations that have long plagued the income tax and might cost the Social Security system much of the limited acceptance the payroll tax has enjoyed over the years.

---

affected by such statutes. For a discussion of the wage realities for lower paid Americans, see HEATHER BOUSHEY ET AL., CTR. FOR ECON. POLICY & RESEARCH, UNDERSTANDING LOW-WAGE WORK IN THE UNITED STATES (2007), http://www.inclusionist.org/files/lowwagework.pdf.

154. *See* MULVEY & WHITMAN, *supra* note 41, at 2–3 (discussing the outcome of the Medicare wage-base elimination).

155. The general topic of the ability of highly compensated employees to manipulate the payroll tax system will be the subject of another article. *See* Patricia E. Dilley, *Are [Payroll] Taxes Really Just for the Little People* (forthcoming).

156. *See* I.R.C. § 409A (2006) (enacted by the American Jobs Creation Act of 2004, Pub. L. No. 108-357, § 885, 118 Stat. 1418, 1634–41 (2004)). *But see* Michael Doran, *Time to Start Over on Deferred Compensation*, 28 VA. TAX REV. 223 (2008) (arguing that § 409A of the Internal Revenue Code and the ensuing regulations implementing it did little to restrain game playing with executive compensation arrangements).

157. *See, e.g.*, Victor Fleischer, *Two and Twenty: Taxing Partnership Profits in Private Equity Funds*, 83 N.Y.U. L. REV. 1 (2008) (analyzing private equity funds employing the partnership form to provide a profit interest to manager partners, thereby transforming services income into investment income taxed at the capital gains rate, currently 15%, as opposed to ordinary income, which is taxed at a top rate of 35%).

In sum, the doughnut hole proposal might achieve marginally better distributional results in tax burdens than the current FICA wage base does, but at the high cost of importing to the Social Security tax system some of the most unpopular characteristics of the income tax—seeming capriciousness, unreasonable complexity, and increased incentives to distort economic realities in order to avoid a tax liability.[158] The only real advantage of the gap from a tax fairness perspective is that only taxpayers at the highest income levels would have increased tax liability, so that those taxpayers would finally be bearing what many workers who never have earnings in excess of the wage base would consider a more equitable tax burden.

Indeed, much of the criticism of the proposal simply points out (with considerable horror, it must be said) the new, higher marginal tax rates to which the highest earning taxpayers would be subjected once FICA applies to wages above $250,000.[159] But those increased rates would be the product of almost any proposal to increase revenues for Social Security, with or without the doughnut hole. The question is whether the increase in what many would consider "tax fairness" is worth the considerable negatives that accompany the gap, and whether it might not be worth taking the political heat and simply raising the wage base without a gap.

B. The Illusory Critiques

Clearly the doughnut hole proposal has a number of serious problems when examined from a Social Security program as well as a tax equity and administration perspective. However, the proposal has also been attacked by a number of commentators claiming to base their opposition to the proposal on its betrayal of what they portray as fundamental principles of the Social Security system. On closer examination, though, many of the most ferocious critiques are grounded in a complete misunderstanding or misrepresentation of basic principles of Social Security. The end result,

158. Some of the unpleasant characteristics of the income-tax system have been described as follows:

> Substantively, the income tax is a mess. Taxpayers at every income level confront extraordinary complexity. . . . Many feel like "chumps" if they pay the taxes they legally owe. Young people, especially, admit that they feel no compunction about filling out their tax forms dishonestly. And the Internet has facilitated growth of the "tax deniers" movement—people who spread their rejection of the legitimacy of any income tax requirements, including the requirement for employers to withhold taxes on their employees' wages.

Michael J. Graetz, *Taxes that Work: A Simple American Plan*, 58 FLA. L. REV. 1043, 1045–46 (2006).

159. *See, e.g.*, Andrew G. Biggs, *Barack Obama's Social Security Donut Hole*, AEI.org, March 2008, at 1, http://www.aei.org/issue/27704.

whether intentional or accidental, is the promotion of notions not only antithetical to a real understanding of Social Security, but also designed to paint a false picture about the program's financial future and the relationship between benefits and financing alternatives.

For example, one commentator, in discussing the President's wage-base proposal, has said that the link between the benefit base and contribution base means Social Security was modeled on a "Contributory Model," under which "you pay in part of your paycheck until you've paid enough to 'cover' your benefits, then (if you keep earning) you don't have to contribute any more."[160] It is unclear what this commentator had in mind, but this description bears little resemblance to either the original Social Security plan of 1935 or the revamped version of social insurance enacted in the Social Security Amendments of 1939 or, indeed, even to any private pension model in existence either in the 1930s or later.[161] Some employer-sponsored defined-benefit pension plans can be described as "contributory" in that they allow or require employees as well as employers to make contributions.[162] However, such plans were highly unlikely to have served as a model for planners in the 1930s as they were quite rare until after the enactment of Social Security, and even under such plans (they are still to be found in state and local government plans), the employee's contribution does not limit the amount of benefits paid.[163]

A persistent element of many of these critical assessments is the charge that a gap in the base would fundamentally change Social Security or somehow be in complete contradiction to basic principles of the program.

---

160. Posting of Mickey Kaus to Slate, http://www.slate.com/id/2193674/#bigdonut (June 17, 2008, 12:47 EST).

161. *See supra* notes 38–45 (discussing the original Social Security Act of 1935 and the Amendments of 1939).

162. STAFF OF J. COMM. ON TAXATION, 109TH CONG., PRESENT LAW AND BACKGROUND RELATING TO EMPLOYER-SPONSORED DEFINED BENEFIT PENSION PLANS AND THE PENSION BENEFIT GUARANTY CORP. ("PBCG") 13 (Comm. Print 2005), *available at* http://www.house.gov/jct/x-3-05.pdf (discussing the present rules regarding qualified retirement plans).

163. Three-quarters of the pension plans established between 1874 and 1929 were wholly employer financed. The other quarter of plans either required employee contributions or, less commonly, allowed voluntary employee contributions to supplement employer-provided sums. The reverberating economic effects of the Great Depression of the 1930s, coupled with the United States government's creation of Social Security, led employers to begin largely trending toward implementation of contributory pension plans in place of defined benefit plans. *See* Patrick W. Seburn, *Evolution of Employer-Provided Defined Benefit Plans*, MONTHLY LAB. REV., Dec. 1991, at 16, 19, *available at* http://www.bls.gov/opub/mlr/1991/12/art3full.pdf. It is not really clear what Kaus means by "contributory" pension plan—he might well be thinking of a defined contribution plan along the likes of a § 401(k) plan, which did not exist until the late 1970s, making it unlikely as a model for Social Security. *See* Posting of Mickey Kaus, *supra* note 160.

"Social Security is structured so that the more you pay in, the more you get back. That's what supposedly makes it a compact among the generations and not a welfare program. Actually, what it does is make it an inefficient, disguised welfare program."[164] In his comment, Ramesh Ponnuru seems to be assuming that the gap proposal would apply to the wage base alone, which is probably reasonable given the difficulties discussed earlier with omitting a large range of earnings records from benefit calculations. Nonetheless, that feature is not yet actually a part of the proposal according to Obama's advisors.[165]

Ponnuru also completely misstates and oversimplifies the structure of Social Security benefit accrual, which is, as described earlier, based on earnings, not on taxes paid, and of the Social Security benefit structure, which provides redistributive benefits based on the weighted-benefit formula but not paid based on proof of need, which is the essence of welfare.[166] The critique itself, however, reveals the commentator's essential hostility to the notion of redistribution, something that is hardly a direct product of the gap but is rather essential to the mission of social insurance.

Another commentator, Nicholas Kaster, echoes the same theme:

> Moreover, under current rules, Social Security caps both benefits and earnings. Thus, unless Obama also favors paying more Social Security benefits to the wealthier earners—highly unlikely—then his plan undermines Social Security's historic role as a basic social safety net rather than a program that redistributes income. This realization has triggered criticism even from Democrats, including Henry Aaron of the liberal Brookings Institution and former Rep. Charles Stenholm of Texas. When you say you're going to begin means-testing the program," Stenholm noted, "you begin to convert Social Security from an insurance program to a welfare program."[167]

Kaster is of course mistaken about Social Security's historic role—Social Security has always functioned as the baseline of income support in old age, part of the "safety net," but it has done so by redistributing income, through the progressive-benefit formula, with benefits financed by all

---

164. Posting of Ramesh Ponnuru to National Review Online, http://corner.nationalreview.com/post/?q=ZjYxMTIwZTIxMTBkYTlhNzhjZDFjNjgwMWNiZWEzZDc= (June 13, 2008, 17:17 EST).

165. *See* Furman & Goolsbee, *supra* note 8; Obama Remarks, *supra* note 8; Rhee, *supra* note 8.

166. *See* 42 U.S.C. § 415 (2006) (requiring that an individual's basic Social Security benefit be calculated based on percentages of the individual's average indexed monthly earnings). For a discussion of the nature of welfare, see FRANCES FOX PIVEN & RICHARD A. CLOWARD, REGULATING THE POOR: THE FUNCTIONS OF PUBLIC WELFARE (1971).

167. Posting of Nicholas J. Kaster to American Thinker Blog, http://www.americanthinker.com/blog/2008/06/obamas_latest_proposal_to_incr.html (June 19, 2008, 11:01 EST).

workers regardless of their chances of ultimately collecting benefits, in a manner similar to private-insurance risk sharing.[168]

It is hard to imagine, moreover, how a public program could function as a safety net without some degree of redistribution. As discussed earlier, the decision to pay higher than strictly proportional benefits to the lowest earners was made right from the start of Social Security and is certainly not a product of President Obama's proposal.[169] It is true that means testing Social Security benefits would fundamentally change the program's character, but that is not what the wage-base-gap proposal would do.[170] It is a substantial and unsupported leap from a proposal to tax higher wage workers, with or without allowing them to accrue additional benefits on their highest wages, to "means testing," which requires actual "testing" of "means," i.e., proof of inadequate income and assets, and demonstration of current need.[171] The core principle of social insurance is the right to a stream of income in old age or disability based on presumed rather than demonstrated need, providing future security while avoiding disincentives to accumulate and save income and resources.[172] This is the diametric opposite of the welfare-program dynamic, the essence of which is to support those who are in current (and in the U.S. program design, dire) need.[173]

The common characteristic of these critiques of the gap proposal is their

---

168. Private insurance program distribution methods have been described as follows:

> The characteristic of risk distribution sets insurance contracts apart from other kinds of contracts. It can be said, then, that a contract of insurance is an agreement in which one party (the insurer), in exchange for a consideration provided by the other party (the insured), assumes the other party's risk and distributes it across a group of similarly situated persons, each of whose risk has been assumed in a similar transaction.

ROBERT H. JERRY, II & DOUGLAS R. RICHMOND, UNDERSTANDING INSURANCE LAW 14 (4th ed. 2007) (emphasis removed).

169. See *supra* notes 36–76 and accompanying text (exploring the origins of Social Security as a redistributive system).

170. *See* Posting of Nicholas J. Kaster, *supra* note 167 (arguing that "unless Obama also favors paying more Social Security benefits to the wealthier earner . . . his plan undermines Social Security's historic role as a basic social safety net").

171. *See generally* PIVEN & CLOWARD, *supra* note 166 (discussing the welfare program dynamic).

172.

> Although the definition of social insurance can vary considerably in its particulars, it basic features are: the insurance principle under which a group of persons are 'insured' in some way against a defined risk, and a social element which usually means that the program is shaped in part by broader social objectives, rather than being shaped solely by the self-interest of the individual participants.

Soc. Sec. Admin., Historical Background and Development of Social Security, http://www.socialsecurity.gov/history/briefhistory3.html (last visited May 13, 2010) (discussing the general principles of social insurance).

173. *See generally* PIVEN & CLOWARD, *supra* note 166.

mischaracterization of Social Security's basic principles, a distortion that mainly seems to stem from an underlying opposition to income redistribution, which is an actual basic principle of social insurance but not something peculiarly characteristic of the gap proposal itself. The notion that redistribution from higher wage workers to lower wage workers would somehow abrogate the "compact between generations" or taint the earnings-based foundation of the benefit structure ignores the fact that redistribution from higher income to lower income is at the heart of not just Social Security, but public financing of government functions generally.[174]

These kinds of criticisms of the doughnut hole proposal should therefore be seen less as actual critiques of the gap and more as rhetorical tactics aimed at contributing to an overall misapprehension of Social Security—part of the continuing conservative resistance to Social Security's redistributive income security, a campaign that has persisted through the entire seventy-five-year history of the program.[175] This strategy essentially creates a "straw man" Social Security, one based on individual equity and return on tax payments that ignores the reality of the actual program of redistributive social insurance. This falsely reconfigured "Social Security" is then used to oppose suggested changes that might actually be consistent

---

174. *See supra* notes 125–30 and accompanying text (discussing Social Security as a function of redistribution within the larger ambit of the social safety net).

175. Conservative Senator Barry Goldwater published *Conscience of a Conservative*, a work criticizing both Democrats and Republicans in 1960; Nancy J. Altman describes his argument as maintaining that

> Social Security and welfare should be provided by the private sector or, if government involvement was absolutely necessary, by state and local governments, but under no circumstances by the federal government. Rather, Social Security and programs like it inevitably, according to Goldwater, lead to "unlimited political and economic power . . . as absolute . . . as any oriental despot." The recipient of these programs, according to Goldwater, is transformed by them "into a dependent animal creature."

ALTMAN, *supra* note 36, at 199. Republican President Ronald Reagan proposed a Social Security reform package that included a recommendation to reduce early retirement benefits among other cuts:

> The Reagan proposal would have reduced the benefits for people who retired early more than the actuarial reduction warranted. Specifically, the law provides that people who retire at age 62 receive the actuarially equivalent 80 percent of the monthly amount received by people who retire at age 65. The administration proposed to reduce the percentage to 55 percent of the age 65 benefit.

Id. at 231. Ultimately, President Reagan dropped the most controversial aspects of his plan but formed the bipartisan National Commission on Social Security Reform to make recommendations regarding the Social Security System. *Id.* at 234, 237. The recommendations of the Commission are reflected in the Social Security Amendments of 1983. *Id.* at 253. The next attack on Social Security came from Republican President George W. Bush, who established a presidential commission to study Social Security and stipulated "that the commission's recommendations 'must include individually controlled voluntary personal accounts.'" *Id.* at 265. President Bush continued his campaign to privatize social security throughout his two terms. *Id.* at 272.

with the principles and purposes of the real Social Security program, ultimately driving the debate into a discussion of false choices about the program's financial future and benefit structure.

C. Other Options Within the Current Base Paradigm

The doughnut hole proposal is not the only option for revamping the current wage-base limit with a view to increasing future program revenues. There are a variety of possible redesigns, which include eliminating the limit altogether, restructuring the base into a series of progressive brackets, and expanding the definition of wages to include items such as certain types of deferred compensation when earned rather than when received and compensation for services currently characterized as return on equity. A brief survey of other options for using the base to raise additional revenue gives an indication of why it is critical at this point to rethink the distinction between earning benefits and paying for them.

1. Eliminating the Limit on the Contribution and Benefit Base

The rationale for limiting the base at any particular point below 100% of earnings and wages is somewhat fuzzy and may well be more easily defended for benefits than for taxes. Rather than raising or redesigning the base, eliminating it altogether might resolve several issues at once, especially in spreading the impact of the FICA tax more equitably across income lines and in raising considerable revenue that would go a long way toward resolving the possible future financing difficulties of Social Security.[176] Of course, additional questions would be raised from the absence of a limit specific to FICA, particularly in connection with the additional benefits that would accrue to very highly paid workers. The fundamental question, however, is whether the limited wage base is really an essential element of either the Social Security benefit or tax structure.

The benefit base serves as a limit on yearly benefit accruals under Social Security, a somewhat different approach from the benefit-accrual systems permitted for use in private defined-benefit pension systems. In private defined-benefit systems, benefits accrue over the working career, usually ratably for each year of service, and the final benefit is usually based on the average of several of the highest years of wages under the plan—highest five or highest three, for example—multiplied by a percentage formula and the numbers of years of service under the plan.[177] There is no real

176. *See* MULVEY & WHITMAN, *supra* note 41, at 17–18 (examining the potential impact of eliminating the wage base).

177. *See generally* MCGILL ET AL., *supra* note 60.

counterpart to the Social Security benefit-base limit in private plans, whose benefits to highly paid participants are now limited primarily by the tax nondiscrimination rules which require (theoretically, if not always in practice) roughly comparable benefits to be paid to highly compensated and non-highly-compensated employees.[178]

The primary difference between benefit accrual under private systems and under Social Security is that the distributional goals of private plans are generally the mirror opposite of the public social insurance program. Private plans generally strive to place as much of the total benefits as possible into the hands of the most highly paid participants, and inclusion of the highest levels of compensation in benefit calculations helps to serve that end.[179] The Social Security benefit structure, in contrast, has a function peculiar to an earnings-based social insurance system in that it is designed to provide proportionally higher benefits, as a percentage of lifetime earnings, to lower earners.[180] The benefit formula determines the level of income replacement for beneficiaries with earnings records at all levels, but the base limit implicitly sets the target for the highest income level we are willing to insure with public funds, since earnings above the base are excluded from the earnings record to which the benefit formula is applied.

The issue of limiting the public subsidy for high-income beneficiaries is important, particularly since one of the only areas of publicly perceived unfairness in the Social Security benefit structure itself is the receipt of benefits by people who apparently do not need them.[181] Yet there is no

178. *Id.*

179. *Id.*; *see also* Michael J. Graetz, *The Troubled Marriage of Retirement Security and Tax Policies*, 135 U. PA. L. REV. 851, 876 (1987) ("The revenue loss attributable to private pensions has been estimated to benefit high-income workers disproportionately, and the distribution of benefits from private pension plans is skewed in the same direction.").

180. The decreasing percentage applied to increasing levels of wages in the benefit formula produces this result. *See* Social Security Act, 42 U.S.C. § 402 (2006) (stating the current benefit formula).

181. *See, e.g.*, Posting of Casey B. Mulligan to Economix, http://economix.blogs.nytimes.com/2010/02/24/are-we-overpaying-grandpa/ (Feb. 24, 2010, 6:00 EST). For a response to this suggestion, see Posting of Dean Baker to Beat the Press, http://www.prospect.org/csnc/blogs/beat_the_press_archive?month=02&year=2010&base_name=the_government_pays_more_money (Feb 25, 2010; 05:10 EST) ("Of course, it would be foolish to compare the money that rich investment bankers get in interest payments on money they have lent to the government with the pure transfer payments that the government makes to ensure that poor children have a decent chance in life. But, it is also foolish to compare the retirement benefits that seniors have largely paid for during their working life, through Social Security and Medicare taxes, with the pure transfer payments that the government makes to ensure that poor children have a decent chance in life."). For a discussion of what means testing of benefits might mean for the program, see AMER. ACAD. OF ACTUARIES, MEANS TESTING FOR SOCIAL SECURITY (2004),

persuasive programmatic reason (as opposed to political) why the benefit formula could not achieve most of the benefit-limitation goals without any limit on earnings recorded for benefit accrual. As a mechanical question, it would be quite feasible to amend the benefit formula to reduce the benefit accrual on those wages by adding gradually smaller brackets for higher levels—for example, adding decreasing accrual rates on top of the current top 15% rate, to be applied to wages above the current base at brackets designed to produce minimal increments in ultimate benefit amounts. If the objection to inclusion of all earnings in the base is the prospect of excessive publicly funded benefits, a revised benefit formula could insure diminishing replacement rates for wages at the top level and minimize the resulting increase in benefits.

Moreover, there is a reasonable argument for including even earnings at very high levels over a worker's thirty-five or forty-year working career. As discussed earlier, few workers consistently earn at an extremely high level throughout their careers, and it could be argued that the ultimate benefit amount for an earner with a volatile earnings record would more fairly reflect her lifetime average record if the years with very high earnings were included in benefit calculations.[182] Nonetheless, the question remains whether allowing the benefit structure to reflect lifetime earnings for all workers, regardless of earnings levels, is consistent with the targeted earnings-replacement rates on which the benefit structure is based and with the principles of social insurance. As discussed above, the limit on the amount of wages subject to FICA taxes was essentially a by-product of the program designers' determination to cover industrial employees for benefit purposes and has been maintained largely as a matter of symmetry with the benefit base.[183] From a programmatic perspective, then, there appears to be no real reason to exclude any wages from the wage base if all earnings are included (at least to some extent) in the benefit base. From a tax perspective, elimination of the base would essentially result in an increase of the top marginal tax rates on wage income alone, which would exacerbate an already-existing issue—the problem of top levels of compensation escaping FICA taxes altogether.

Wages would appear to be a less malleable base for taxation than income—they are recorded and reported by a third party, the employer, and for the most part the amount and timing of an employee's wages are not under the recipient's control. But highly compensated workers whose wages now largely escape the OASDI portion of FICA because they are

---

http://www.actuary.org/pdf/socialsecurity/means_0104.pdf.

182. *See supra* note 95 and accompanying text (noting that earnings often fluctuate over the course of a worker's career).

183. *See generally supra* notes 41–42.

above the wage base are far more likely to control the form and conditions of their compensation than the vast majority of salaried or hourly workers. Highly paid executives are better able to recharacterize compensation as non-wage income or to delay receipt in order to manipulate the timing and amount of tax liability.[184] As discussed earlier, these mechanisms became more popular and important to highly compensated executives when the limit on the base for purposes of the Medicare hospital-insurance portion (1.45%) of the FICA tax was eliminated in 1993.[185] Any attempt to impose an additional 6.2% OASDI tax on top of earnings will inevitably encourage even more strategies to avoid compensation in the form of wages. In the absence of measures to capture compensation in disguise, eliminating the limit on the wage base for tax purposes may result in far less revenue than might be anticipated.

2. Making the Base Progressive

Simply raising or eliminating the current base limit is not the only option for making the base more fair and effective in financing and calculating benefits. So long as most wages of most workers are included in the main contribution and benefit base, a tiered approach for wages above the currently applicable base might be explored. This would be similar to the doughnut hole proposal in some ways, except that instead of a gap where no tax is imposed followed by full taxation above the gap, intermediate base levels and taxes could be added to the current base. For example, half the current tax rate could apply to wages between the current base and $250,000, and then a quarter to 30% of the rate to wages above that, with proportionately smaller benefit accruals at each level.[186] There are many possible variations on the theme of a progressive wage base, but the essential goal is a compromise, allowing the benefit-accrual system to partially reflect earnings above the current base level while imposing the wage tax on all earned income.

Such a tiered system would clearly be the inverse of a progressive rate

184. *See supra* notes 153–58 and accompanying text (discussing the ability of higher paid persons to engage in gaming of the Social Security benefit system).

185. *See* MULVEY & WHITMAN, *supra* note 41, at 8 (noting the elimination of the Medicare wage base and discussing potential effects of elimination of the Social Security wage base in comparison).

186. This tiered approach is similar to legislation introduced in the 109th Congress by Representative Wexler. His bill, H.R. 2472, would impose an additional Social Security contribution of 3% of wages above the current wage base on workers and employers. In his bill, however, earnings above the base would not be included in the benefit computation base. *See* Social Security Forever Act of 2005, H.R. 2472, 109th Cong. (2005), *available at* http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=109_cong_bills&docid=f:h2472ih.txt.pdf.

structure, and would import some of the complexity of the tiered income-tax design into the payroll tax. On the other hand, it would reduce some of the incentives for gamesmanship at the top wage levels because the tax rate would be less than the full 6.2%, which would provide some benefit accrual even at the very top of the earnings scale. Most importantly, perhaps, such a base structure would restore some fairness to the payroll tax in the mind of most taxpayers by imposing at least some FICA tax on all wages. It is unclear, however, that the complexity resulting from mirroring the progressivity of the income tax (limited as it is) in the payroll-tax base is worth the effort, particularly when there is a more direct, simpler, and ultimately more flexible way to accomplish the same goal. That solution, however, requires some fundamental rethinking of the basis of entitlement to Social Security benefits.

IV. RETHINKING THE BASE LIMIT PREMISE

While the doughnut hole proposal may lack credibility and substance as a serious policy proposal at this point, even suggesting such a change opens the debate about FICA taxes to a new level of questions about what other design changes might be considered that would improve the future finances of Social Security while also promoting its programmatic goals. What is necessary is a simple but fundamental shift in thinking, albeit one grounded in the program's original and enduring premise that work itself creates the right to security and that could lead to more creative approaches to future financing shortfalls in Social Security, based on decoupling the base for earning benefits from the base for payroll taxes. Such an approach would make it easier to consider alternative revenue options, to add to the current nonpayroll-tax sources of financing Social Security, that would achieve greater parity in tax burdens and additional revenue (if and when that is necessary) without violating any real basic principles of social insurance.

This reframing is essentially focused on the legal basis of entitlement under Social Security: both Medicare and Social Security cash benefits are earned through working, not through paying payroll taxes, and benefits are financed through many sources, not just payroll taxes.[187] Payroll taxes finance not the individual benefits of the taxpayer but the overall Social Security system and the furtherance of its goals, ultimately social stability and individual financial security.[188] The failure to recognize and structure public policy around the distinction between taxes paid and benefits earned has made the payroll tax into a straightjacket on analysis and development

187. *See supra* notes 121–30 and accompanying text (discussing the erroneous notion of Social Security as a system that returns benefits based on the amount of tax contributed).

188. *Id.*

of financing options. Once the distinction is recognized, however, reimagining the contribution and benefit base (and in particular the limit) becomes a real possibility and more than an academic exercise.

The traditional defined-benefit pension model, particularly the concepts of benefit accrual combined with program-funding requirements, provides an analogy that may make this distinction easier to recognize. Employer-sponsored pension plans base the ultimate benefit of participants in the plan on, in part, a method of accruing benefits over the participant's years of service with the employer, similar to the years of earnings credits tracked by the Social Security Administration for each worker covered by the program.[189] Financing of private pension benefits, on the other hand, is a completely separate issue, with funding requirements based on funding needs for whatever benefits have been promised under the terms of the plan.[190] While there are many dissimilarities between Social Security and employer-provided pensions, in this respect the comparison is an apt one—workers earn their private pension benefits, and plan sponsors, normally employers, fund the trust that eventually pays those benefits. In the same way, it is work that creates the entitlement to Social Security benefits, and those benefits are funded by the "plan sponsor," the taxpayers themselves, with the federal government as the manager of the plan and its trust.

If we separate the tax function of the base from the benefit-accrual function, immediately the policy options for increasing revenues to the Social Security system are freed from the inherent regressivity and possible employment impact of payroll-tax increases. The possibilities for additional financing range from a dedicated income surtax on taxpayers with more than $250,000 in earned income in a year, which would mimic the Obama payroll-tax proposal, to simply supplementing payroll-tax revenues with general tax revenues in any year in which the trust fund reserves are inadequate to fully fund benefit payments. Even the President's doughnut hole proposal might be more feasible if it produced additional revenues from the highest wage workers with no additional benefit entitlement, although all the aforementioned problems with tax avoidance would still be a formidable obstacle.

If additional revenues are the only goal, it might be more consistent with Social Security principles to simply impose a surtax on income—not wages—above a certain adjusted gross income level, whether $250,000 or $200,000, and earmark the results for financing Social Security. This approach would have many signal advantages—the tax rate and affected

189. *See supra* note 177 and accompanying text (explaining how benefits accrue during an employee's working career).

190. *See id.* (discussing financing of private pension benefits).

income level could be adjusted simply based on financing needs rather than on juggling issues of benefit accrual and wage definitions. A surtax could also be readily justified based on the last two decades of changes in the nature of compensation from wages to other types of compensation which escape the wage base, as well as the skewing of the American income structure to an imbalanced share of national income and wealth at the top.[191] Such a tax could also be said to make up for the dramatic loss of employer-provided pensions over the last twenty years at the expense of working people's retirement security while the highest corporate earners steered a higher and higher proportion of company earnings into their own compensation packages.[192]

The major objection to these approaches would of course be that it is somehow a breach of principle to look to extra-payroll-tax revenue since the system has always been financed by contributions out of the wages of those accruing benefits on those same earnings. Yet, as discussed earlier, one of the longest standing, yet almost totally ignored, principles of Social Security is the assumption of the drafters that general revenues would of course be part of the long-term financing of the program.[193] In many ways, that is the case already, given that income taxes imposed by the 1983 Amendments on Social Security benefits received by higher income beneficiaries are earmarked for the Social Security trust funds to be used for financing the program into the future.[194]

More importantly, most objections to using general revenues for Social Security financing are based on a misapprehension that Social Security benefits are tied to taxes or contributions so that infusing general revenues into the program's financing would be some sort of violation of a "more you pay, more you get" principle. On the contrary, as stated earlier, benefit accrual under Social Security is based on the individual earnings history, while payroll taxes are a financing source unconnected to benefits actually paid.[195] The contributory principle is important and is more than simply a symbol—it helps to create a direct commitment, a participatory connection between worker and program in a way that a completely general revenue financing system could not. If additional financing is needed, however, it does not seem unreasonable to require additional

191. *See generally* McMahon, *supra* note 3 (discussing the increasing concentration of income and wealth in the top 1% of the American population).

192. *Id.*

193. *See supra* notes 66–68 and accompanying text (showing the presumption by the original drafters that general revenues would be needed).

194. Social Security Amendments of 1983, Pub. L. No. 98-21, § 121, 97 Stat. 65, 80–84.

195. *See supra* notes 121–30 and accompanying text (discussing the erroneous notion of Social Security as a system that returns benefits based on the amount of tax contributed).

contributions from those at the top income levels who now reap a significantly higher portion of the benefits of a relatively stable social system for an aging population based largely on Social Security.

A more serious problem with decoupling the base for accrual of benefits from the base for taxes is whether the contributory principle would be so degraded that public support for Social Security—and tolerance of FICA taxes—would substantially erode. Currently the public is largely unaware of the degree to which general tax revenues already finance Social Security and Medicare, and the strong feeling of ownership toward both programs, exhibited as recently as in the protests against health care reform that declared hostility to government-provided health care while at the same time demanding protection for Medicare, may well be grounded primarily in the belief that payroll taxes are the source of entitlement.

Of course, decoupling the tax base from the benefit-accrual base in no way implies a change in the earnings basis of entitlement, and additional financing from nonpayroll-tax sources would simply be an extension of current nonpayroll-tax financing. The deeper issue is public perception, which usually trumps reality and fact. If public support for Social Security rests on the perception that each worker is contributing to her own individual savings account that is drawn on in retirement, it might be difficult to persuade the public that adding other revenues to payroll-tax financing is consistent with an earnings-based entitlement. On the other hand, there is very strong public support for increasing the wage-base limit to address any revenue shortfall the system might experience,[196] indicating that American workers generally think higher income taxpayers are not contributing their fair share to support Social Security.

Political considerations aside, a compelling argument can be made for increasing the contributions of higher wage workers without commensurate increases in benefits, and for either a wage tax or an income tax from the perspective of the same fundamental notions of fairness and social benefit that underlie the justification for progressive income taxation. It can be argued that it is fair to tax higher income taxpayers at a higher marginal rate on their top brackets of income, in part because they benefit to a greater degree from social and economic structures and institutions. For example, a surgeon earning $1 million in a year is able to make and keep those earnings because of a host of public goods: police and fire protection, courts to enforce property rights, public infrastructure, the military, and not least, the Medicare program which makes it possible for many of her patients to pay for her services. Because of her higher level of benefit from

196. *See supra* note 4 and accompanying text (providing polling data on citizen support for an elimination of the tax cap).

those public goods, it is argued that it is not unreasonable for our surgeon to pay a higher marginal rate on the top part of her income as recognition of her greater degree of benefit from the protection of society.

This reasoning applies equally well to the benefits of Social Security which extend far beyond individual benefits—Social Security's role in stabilizing society and the economy by insuring a steady stream of income at all income levels, maintaining demand, and providing retirees a stable set of income expectations cannot be overstated. The "money's worth" argument frequently used by critics of the program to attack its value to high-wage workers ignores the value to the upper class of the enormous social good of economic and social stability through the life cycle that Social Security provides. It is therefore not unreasonable to ask them to pay what might be thought of as a "stability premium" in the form of higher payroll or income taxes to be used to insure Social Security's benefit payments.

Ultimately, the issue of separating the benefit base from the tax base is a question of philosophy rather than economics or tax theory. Americans are unaccustomed to the notion that is widely accepted in most industrialized countries: that people have a fundamental right, based not on work or taxes or fees but on existence in the nation's jurisdiction, to basic welfare in the form of at least minimal income support and access to health care. Most Americans recognize the civic rights laid out in the Constitution to free speech, voting, etc. as being inalienable for American citizens, but economic and social welfare rights have always in the American system had to be earned, paid for, or both. Social Security is based on that very American notion that economic security in old age must be earned and cannot be a "gift" from the government.

The conflation of the benefit and tax base limits, however, has allowed that "earned right" to be portrayed, particularly in recent years, as not so much earned as "paid for"—paving the way for the last few decades of red herring arguments about the "return on investment" that compared payroll taxes paid in to benefits received. Correcting that erroneous framing to restore the "earned" part of the earned-right principle is an enormous task. Americans have become inured to having every public issue framed as a question of narrow cost–benefit analysis and to looking at all public goods, including the general welfare provided by government programs like Social Security, as a question of narrow self-interest. When the issue of Social Security financing is framed this way, it becomes extremely difficult to increase the payroll-tax base limit without invoking cries of protest that higher wage workers will get less than they ought to in benefits in exchange for their tax payments. When the system is looked at correctly, however, as a system in which benefits are earned, and the system as a whole is paid for

with direct contributions from workers and with other tax revenues as well, the question of raising additional revenue immediately has multiple answers, of which raising the taxable-wage-base limit is only one.

CONCLUSION

Tax policy analysts and commentators are reluctant to follow President Obama through the new doughnut hole he has proposed for the Social Security wage base for a variety of reasons both legitimate (tax avoidance and gaming possibilities) and misplaced (violation of some basic Social Security principle). I suggest in this Article that a clear analysis of the proposal and alternatives should be grounded in a different way of looking at the base and at the fundamentals of entitlement to benefits that are in fact, rather than in fantasy, consistent with Social Security's core principles. If we reframe the entitlement notion itself as one of earnings and work, rather than payments and taxes, the entire question of Social Security financing becomes infinitely more open to multiple answers, ranging from changes in the tax base limit without changes in the benefit base to income-tax surtaxes on higher income taxpayers.

Of course, there is no immediate need to do anything about Social Security financing at all; on a trust fund reserve basis, it will most likely be unnecessary to raise revenue for the OASDI cash-benefit system beyond what is currently projected for at least two decades and possibly longer.[197] A detailed discussion of financing projections and the relationship between dedicated tax financing and the trust-fund concept must wait for a follow-up article, but clearly it makes little sense to increase the dedicated tax base limit now to raise payroll-tax revenues that are not yet needed to pay for current payments. That approach has been tried before—the 1983 Social Security legislation put in place benefit reductions (including the increase in the age for receipt of full benefits) and base and FICA tax rate changes that were purposely designed to build up a large reserve through the 1990s and the first decade of the 21st century, to be drawn on if and when yearly revenues became insufficient to pay yearly benefits in the second quarter of the new century.[198] Now that we may be approaching the point of drawing on those reserves by 2018 or earlier to make up yearly revenue shortfalls, the trust-fund reserves are characterized by critics of Social Security as "imaginary" and alarming cries that the system is "insolvent" pervade the public commentary.

Despite a concerted campaign from conservative political and economic analysts to assert that the trust fund does not exist, however, it is undeniably

197. *See generally* 2009 BD. OF TRUSTEES REPORT, *supra* note 8.

198. *See generally* Social Security Amendments of 1983, Pub. L. No. 98-21, 97 Stat. 65.

true that American workers paid higher payroll taxes than necessary over the last fifteen years in order to fund in advance the retirement of the baby boom and later generations.[199] Because the trust funds are held as special-issue obligations of the U.S. Treasury, however, the practical effect of surplus payroll-tax revenues over this period has been to finance the general revenue deficit with regressive payroll-tax collections. Once the trust-fund-bond reserves begin to be called upon to make up for yearly revenue shortfalls, clearly income and other tax revenues will be under more pressure to meet other government obligations. Nonetheless, the political–economic obligation to pay Social Security benefits that are earned over a working lifetime, regardless of the source of revenue, remains in place, and the political reality of the trust-fund reserves made up of those excess tax payments largely from baby-boom workers cannot be wished away by those who would rather reduce benefit levels than increase income taxes to repay general fund obligations to the Social Security trust funds.

The real utility of discussing changes to the wage-base limit now, including the President's doughnut hole notion, is to begin to reframe public and policymaker understanding of the real basis for Social Security entitlement as I have laid out in this Article. By reimagining the Social Security contribution and benefit base limit, we can free the analysis and the policymakers from the straightjacket of regressive taxation as well as from irrelevant arguments about rates of return on tax payments into the system. This is not really either an economic or tax policy issue per se—it is rather a question of properly understanding the peculiarly American premise that those who work are entitled to dignified and meaningful economic security. That premise has provided a sound, if not generous, basis for economic and social stability: The future stability of the Social Security benefit entitlement and of the revenue stream that funds those benefits must rest on an accurate understanding and application of that premise so that earned benefits can continue to be paid for by whatever means necessary.

199. *See, e.g.*, COMM. ON WAYS AND MEANS, 98TH CONG., ACTUARIAL COST ESTIMATES OF THE EFFECTS OF PUBLIC LAW 98-21 ON THE OLD-AGE, SURVIVORS AND DISABILITY INSURANCE AND HOSPITAL PROGRAMS 22–23 (Comm. Print 1983) (showing the buildup in trust-fund reserves because of the excess of payroll taxes collected over benefits being paid out from the mid-1980s through 2020, with the corresponding decline in reserves thereafter).

DEFINING DEFERENCE DOWN, AGAIN: INDEPENDENT AGENCIES, *CHEVRON* DEFERENCE, AND *FOX*

RANDOLPH J. MAY*

TABLE OF CONTENTS

INTRODUCTION

In 2006, I published an article in this law review entitled *Defining Deference Down: Independent Agencies and* Chevron *Deference*.[1] In that article, I posed the question, "Should the statutory interpretations of independent regulatory agencies, such as the FCC's determination at issue in *Brand X*, be accorded a lesser degree of judicial deference than those accorded to executive branch agencies?"[2] In response, I suggested that "a reading of *Chevron* that

* Randolph J. May is President of the Free State Foundation, an independent non-partisan think tank located in Rockville, Maryland.

1. Randolph J. May, *Defining Deference Down: Independent Agencies and* Chevron *Deference*, 58 ADMIN. L. REV. 429 (2006).

2. *Id.* at 432. *Brand X* refers to the Supreme Court's decision in *National Cable & Telecommunications Ass'n v. Brand X Internet Services*, 545 U.S. 967 (2005). My previous article discusses the *Brand X* decision in considerable detail. For present purposes, it suffices to note that the Supreme Court, in reviewing a decision of the Federal Communications Commission (FCC) interpreting a provision of the Communications Act, held that, when in conflict, *Chevron* deference trumps the doctrine of stare decisis. *Chevron* deference refers to the standard of deference to be accorded actions of administrative agencies when they interpret

accords less deference to independent agencies' decisions than to those of executive branch agencies would be more consistent with our constitutional system and its values."[3]

Chevron's central holding is that when a statutory provision is ambiguous,[4] if the agency's interpretation is "based on a permissible construction of the statute,"[5] the agency's interpretation is to be given "controlling weight."[6] The literature on *Chevron* is vast, and my earlier article explains *Chevron*'s basic principles and contains citations to many other sources which discuss the case, so I am not going to rehash *Chevron* here. Rather, in order to provide the context for my contention in *Defining Deference Down* that independent agencies should receive less deference than executive branch agencies, I wish only to quote here what I regard as the key passage setting forth the *Chevron* Court's rationale:

> Judges are not experts in the field, and are not part of either political branch of the Government. . . . [A]n agency to which Congress has delegated policymaking responsibilities may, within the limits of that delegation, properly rely upon the incumbent administration's views of wise policy to inform its judgments. While agencies are not directly accountable to the people, the Chief Executive is, and it is entirely appropriate for this political branch of the Government to make such policy choices—resolving the competing interests which Congress itself either inadvertently did not resolve,

---

ambiguous statutory provisions. *See* Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837, 842–45 (1984).

3. May, *supra* note 1, at 453.

4. Of course, "If the intent of Congress is clear, that is the end of the matter; for the court, as well as the agency, must give effect to the unambiguously expressed intent of Congress." *Chevron*, 467 U.S. at 842–43. Determining whether the intent of Congress is clear is step one of *Chevron*.

5. *Id.* at 843. When the intent of Congress is not clear, what constitutes a "permissible" construction of a statute at the step two inquiry naturally may not be self-evident. For most scholars, permissibility equates with the same type of reasonableness analysis that courts undertake in deciding whether an agency decision is arbitrary or capricious under § 706(2)(A) of the Administrative Procedure Act (APA), 5 U.S.C. § 706(2)(A) (2006). *See* Ronald M. Levin, *A Blackletter Statement of Federal Administrative Law*, 54 ADMIN. L. REV. 1, 37–38 (2002) ("If the statutory meaning on the precise issue before the court is not clear, or if the statute is silent on that issue, the court is required to defer to the agency's interpretation of the statute if that interpretation is 'reasonable' or 'permissible' ('step two' of *Chevron*). . . . Courts may look, for example, to whether the interpretation is supported by a reasonable explanation and is logically coherent. In this regard, the step two inquiry tends to merge with review under the arbitrary and capricious standard . . . .").

6. *Chevron*, 467 U.S. at 844. The key point here is that, apart from the vagaries of defining permissibility or reasonableness in any given case, when *Chevron* applies, it requires a highly deferential review that generally is outcome-determinative. As Jeffrey Lubbers points out in his authoritative text, "The Supreme Court has only rarely set aside an agency action under step two." JEFFREY S. LUBBERS, A GUIDE TO FEDERAL AGENCY RULEMAKING 499 (4th ed. 2006).

> or intentionally left to be resolved by the agency charged with the administration of the statute in light of everyday realities.[7]

Relying heavily on the obvious import of this passage, I argued in *Defining Deference Down* that the *Chevron* doctrine is rooted in the Supreme Court's understanding of fundamental separation-of-powers principles, which dictate that when Congress leaves gaps in a statute, it is for the politically accountable branches, not unelected judges, to make policy by doing the gap filling.[8] That being so, because the independent agencies, such as the Federal Communications Commission (FCC), are less politically accountable than the executive branch agencies with respect to their policymaking actions, I suggested that it should follow that courts reviewing independent agencies' statutory interpretations should accord them less *Chevron* deference.[9] For good measure, I added that "it is odd in a

---

7. *Chevron*, 467 U.S. at 865–66. To reinforce the political accountability rationale, the Court added that "federal judges—who have no constituency—have a duty to respect the legitimate policy choices made by those who do." *Id.* at 866.

8. *See* SECTION OF ADMIN. LAW & REGULATORY PRACTICE, AM. BAR ASS'N, A GUIDE TO JUDICIAL AND POLITICAL REVIEW OF FEDERAL AGENCIES 56 (John F. Duffy & Michael Herz eds., 2005) ("Thus, *Chevron* has significant institutional implications, shaping the relationship among the branches of government and serving as a kind of 'counter-*Marbury*' for the regulatory state."). The reference to *Chevron* as a kind of counter-*Marbury* is from Cass R. Sunstein, *Law and Administration After* Chevron, 90 COLUM. L. REV. 2071, 2075 (1990) ("*Chevron* promises to be a pillar in administrative law for many years to come. It has become a kind of *Marbury,* or counter-*Marbury*, for the administrative state."). By counter-*Marbury*, Professor Sunstein meant to contrast *Marbury's* oft-repeated dictate that it is for the judges to "say what the law is," Marbury v. Madison, 5 U.S. (1 Cranch) 137, 177 (1803), with the highly deferential *Chevron* review standard that tilts toward allowing the agencies to say what the law is.

9. See May, *supra* note 1, at 442–45, where I discussed the nature of the independent regulatory agencies, describing the features, such as staggered fixed terms and bipartisanship requirements, which are intended to make them independent. A significant feature is the provision, found in several of the independent agency statutes, that prevents the President from removing commissioners except upon "good cause." Such a removal limitation was upheld against constitutional attack in *Humphrey's Executor v. United States*, 295 U.S. 602, 631–32 (1935). The Supreme Court said that, in light of the removal limitation and other features discussed in my *Defining Deference Down* article, Federal Trade Commission (FTC) commissioners were intended to be "free from executive control." *Id.* at 628. The FCC and FTC share many of the same institutional features that lead them to be considered "independent" agencies, including having a bipartisan mix of commissioners that serve for staggered fixed terms. The Securities and Exchange Commission (SEC) and the Commodities Futures Trading Commission (CFTC) also share these features and are considered independent agencies. I discuss the nature of independent regulatory agencies, and especially the FCC, in more detail in Randolph J. May, *The FCC's Tumultuous Year 2003: An Essay on an Opportunity for Institutional Agency Reform*, 56 ADMIN. L. REV. 1307, 1310–12 (2004). For a very useful comprehensive study of independent agencies, see Marshall J. Breger & Gary J. Edles, *Established by Practice: The Theory and Operation of Independent Federal Agencies*, 52 ADMIN. L. REV. 1111, 1112–14 (2000).

constitutional system with three defined branches for courts to give controlling deference to agencies that, not without reason, are commonly referred to as 'the headless fourth branch.'"[10]

In *Defining Deference Down*, I observed that the question whether independent agencies should receive a lesser degree of *Chevron* deference had been subjected to little examination. I could find no court opinion addressing the question and only sparse commentary in the academic literature. To this same point, David Gossett commented in 1997 that *Chevron*'s political accountability rationale "would imply that independent agencies might not deserve *Chevron* deference, though no [commentary] seems to have explored this idea."[11]

While the commentary was very sparse, there nevertheless had been some hints here and there that others might share my view that independent agencies should receive less deference. Notably, Elena Kagan—now Solicitor General of the United States—suggested linking "deference in some way to presidential involvement" in her magisterial article, *Presidential Administration*.[12] She proposed a "more refined version" of *Chevron*, one in which deference for an agency interpretation would be tied to the level of presidential involvement in the decisionmaking process.[13] Solicitor General Kagan suggested this refined *Chevron* doctrine "would begin by distinguishing between actions taken by executive branch agencies and those taken by independent commissions."[14] After discussing the factors that give independent agencies considerably greater freedom from presidential control than executive agencies, including especially the lack of presidential removal power with respect to independent agencies,[15] she explicitly suggested that a revised *Chevron* doctrine "attuned to the role of the President would respond to this disparity by giving greater deference to executive than to independent agencies."[16]

---

10. May, *supra* note 1, at 451.

11. David M. Gossett, Chevron*, Take Two: Deference to Revised Agency Interpretations of Statutes*, 64 U. CHI. L. REV. 681, 689 n.40 (1997).

12. Elena Kagan, *Presidential Administration*, 114 HARV. L. REV. 2245, 2376 (2001). At the time she wrote *Presidential Administration*, Kagan was a Visiting Professor at Harvard Law School. Later she served as Dean of the law school before becoming Solicitor General of the United States. In the Clinton Administration, she served as Deputy Assistant to the President for Domestic Policy and Deputy Director of the Domestic Policy Council.

13. *Id.* at 2377.

14. *Id.* at 2376.

15. Significantly, Kagan refers to the lack of presidential removal power with respect to the commissioners of independent agencies as "the core legal difference between these entities." *Id.*

16. *Id.* at 2377. Another of the few isolated references to the question of *Chevron* deference to independent agency deference came from John Duffy. He stated that "[i]f the courts really followed the common law logic of *Chevron*, they should have balked at extending

In the Supreme Court's *Brand X* decision, where *Chevron* deference played a determinative role in affirming an FCC order interpreting a statutory provision, neither the majority nor concurring or dissenting opinions questioned whether the independent agency should receive a lesser degree of deference than executive agencies. The Court assumed no difference in treatment between executive departments and independent agencies. Indeed, there was no discussion even intimating the question ought to be examined. This past Term, however, in *FCC v. Fox Television Stations, Inc.*,[17] the beginnings of such a discussion did emerge, albeit not directly in the context of the application of *Chevron*. The Supreme Court's opinions in the *Fox* case are well worth examining not only for what they say more broadly about judicial review of changes in agency policy, an important administrative law issue which will be discussed here only briefly, but also for what the opinions may portend concerning the question of a differential standard of review for executive branch and independent agencies. That question, first examined in *Defining Deference Down*, remains my project here.

I. *FCC V. FOX TELEVISION STATIONS, INC.*: THE FCC CHANGES ITS BROADCAST INDECENCY ENFORCEMENT POLICY

In *Fox*, the Supreme Court, reversing the Second Circuit, affirmed a change of FCC policy to the effect that even isolated, nonrepetitive incidents of indecent speech could be sanctioned.[18] The FCC had gradually expanded its enforcement of the statutory indecency prohibition since the Supreme Court, in the 1978 landmark *Pacifica* case,[19] sustained the agency's initial indecency enforcement activity against statutory and constitutional attack. In the FCC enforcement actions ultimately reviewed in the *Fox* case, the agency articulated a new policy to the effect that it could sanction a "non-literal (expletive) use of the 'F- and 'S-Words' even when the word was used only once."[20] The court of appeals held the FCC's actions unlawful on the basis that the agency's new "fleeting expletives" policy was inadequately explained and, therefore, arbitrary and capricious

---

Chevron to [independent] agencies, which have less democratic accountability than agencies like the EPA, whose heads serve at the pleasure of the President." John F. Duffy, *Administrative Common Law in Judicial Review*, 77 TEX. L. REV. 113, 203 n.456 (1998).

17. 129 S. Ct. 1800 (2009), *rev'g and remanding* 489 F.3d 444 (2d Cir. 2007).

18. A federal statute prohibits broadcasting of "any . . . indecent . . . language." 18 U.S.C. § 1464 (2006).

19. *See* FCC v. Pacifica Found., 438 U.S. 726, 729, 750–51 (1978) (affirming that the FCC's determination that a radio station's daytime broadcast of George Carlin's "Filthy Words" monologue was sanctionable under the indecency prohibition).

20. 129 S. Ct. at 1807.

under the Administrative Procedure Act's (APA's) review standard.[21]

I do not want to focus much of my attention on the aspect of the Supreme Court's decision that addresses an important general administrative law issue which is likely to spawn much commentary among administrative law professors and practitioners—that is, the Court's holding that there is no basis in the APA for subjecting an agency *change* of policy to a "more searching standard of review" than that applied to the adoption of the *existing* policy.[22] As Justice Scalia put it for the *Fox* majority, the agency "need not demonstrate to a court's satisfaction that the reasons for the new policy are *better* than the reasons for the old one; it suffices that the new policy is permissible under the statute, that there are good reasons for it, and that the agency *believes* it to be better."[23] According to Justice Scalia, the APA makes no distinction "between initial agency action and subsequent agency action undoing or revising that action."[24] In short, contrary to the practical import of the Second Circuit's decision, the Court held that when adopting a new policy "the agency need not always provide a more detailed justification than what would suffice for a new policy created on a blank slate."[25]

Another aspect of the *Fox* case that I will not address in-depth warrants at least brief mention. Fox and other broadcast networks claimed before the agency and in court that the FCC's new "fleeting expletives" policy violates their First Amendment rights because the vagueness of the agency's new policy chilled free speech.[26] Because the Second Circuit held the FCC's "fleeting expletives" policy unlawful as arbitrary and capricious, it did not decide the First Amendment question.[27] Nevertheless, in dicta, it proceeded to question whether the FCC's new policy "can survive First Amendment scrutiny."[28] Based on its examination of the relevant First Amendment jurisprudence, the court of appeals majority concluded, "[W]e are sympathetic to the Networks' contention that the FCC's indecency test is undefined, indiscernible, inconsistent, and consequently,

21. See 5 U.S.C. § 706(2)(A) (2006) ("The reviewing court shall . . . hold unlawful and set aside agency action . . . found to be arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law . . . .").

22. *Fox*, 129 S. Ct. at 1810.

23. *Id.* at 1811.

24. *Id.*

25. *Id.*

26. Fox Television Stations, Inc. v. FCC, 489 F.3d 444, 463 (2d Cir. 2007), *rev'd*, 129 S. Ct. 1800 (2009).

27. *See* 489 F.3d at 462 (holding that the agency failed to provide a "reasoned analysis justifying its departure from the agency's established practice").

28. *Id.* at 463.

unconstitutionally vague."[29] After holding that the Second Circuit erred in finding the agency's action unlawful under the APA's arbitrary and capricious standard, the Supreme Court's *Fox* majority observed that "[i]t is conceivable that the Commission's orders may cause some broadcasters to avoid certain language that is beyond the Commission's reach under the Constitution."[30] But it too declined to address the constitutional issue, with Justice Scalia declaring, "Whether that is so, and, if so, whether it is unconstitutional, will be determined soon enough, perhaps in this very case."[31]

II. DIFFERENT REVIEW STANDARDS FOR INDEPENDENT AND EXECUTIVE BRANCH AGENCIES: A DEBATE EMERGES IN THE *FOX* CASE

In the context of deciding whether the FCC's change of policy regarding the indecency prohibition was lawful, a debate emerged in the *Fox* case, albeit not altogether sharply, as to whether the actions of the independent agencies such as the FCC should be subject to a heightened standard of judicial review.[32] Apparently because the FCC's change of policy was not based on an interpretation of a statutory term as was the case in *Chevron* itself,[33] the discussion in *Fox* concerning whether more or less deference is due independent agencies did not refer directly to *Chevron*. Nevertheless, as will be seen, there were certainly *Chevron*-like echoes as the Justices debated the relevance of the FCC's political accountability (or lack thereof) to determine whether the proper standard of review should be more or less searching. It would not be at all surprising to see these echoes reverberate in a way that leads, sooner or later, to a more robust dialogue concerning the differential review issue I raised in *Defining Deference Down*.

---

29. *Id.*

30. *Fox*, 129 S. Ct. at 1819.

31. *Id.* Justice Thomas wrote a concurring opinion "to note the questionable viability of the two precedents that support the FCC's assertion of constitutional authority to regulate the programming at issue in this case." *Id.* at 1819–20 (Thomas, J., concurring). He asserted that *Red Lion Broadcasting Co. v. FCC*, 395 U.S. 367 (1969) and *FCC v. Pacifica Foundation*, 438 U.S. 726 (1978) "were unconvincing when they were issued, and the passage of time has only increased doubt regarding their continued validity." *Id.* at 1820. I have been a proponent of this view. In support of his assertion, Justice Thomas, *id.* at 1822, cited my recent article, Randolph J. May, *Charting a New Constitutional Jurisprudence for the Digital Age*, 3 CHARLESTON L. REV. 373 (2009).

32. I should note here that I understand Justice Breyer objected to Justice Scalia's characterization of his position as advocating a "heightened standard" of review. *Fox*, 129 S. Ct. at 1831 (Breyer, J., dissenting). Apart from the semantics, for my purposes the point, to employ *Chevron*-speak, is that it is clear that Justice Breyer advocated a less deferential standard of review than did Justice Scalia, and the difference is in some material way related to the status of the FCC as an independent regulatory agency.

33. Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837, 840 (1984).

Before offering my own thoughts concerning the Justices' statements in *Fox*, it is useful to set forth their statements relating to the question of a differential deference standard. It is best to begin with the dissents to which Justice Scalia, this time in a plurality opinion, is so clearly responding. Justice Breyer began his dissent by pointing to the characteristics of the FCC, such as fixed terms of office for commissioners and the fact that they are not directly responsible to the voters, that he says give the agency its independence.[34] He declared that despite the fact that the law grants those in charge of independent agencies broad authority to determine policy, "it does not permit them to make policy choices for purely political reasons nor to rest them primarily upon unexplained policy preferences."[35] According to Justice Breyer, an independent agency's "comparative freedom from ballot-box control makes it all the more important that courts review its decisionmaking to assure compliance with applicable provisions of law—including law requiring that major policy decisions be based on articulable reasons."[36] He emphasized the important role agency expertise plays in producing reasoned decisions.[37] Suffice it to say, with no purpose served by detailing all his points here,[38] Justice Breyer found the FCC's change of policy to be inadequately explained. In his view, it was not reasoned decisionmaking but rather was arbitrary and capricious decisionmaking. Notably, although he distinguished at the outset between policy choices made for purely political reasons and policy choices based on reasoned decisionmaking, Justice Breyer did not explicitly identify the source of any claimed "purely political" reasons for the FCC's policy change. He just identified what he saw as the defects in the agency's reasoning.

In his dissent, Justice Stevens suggested that independent agencies like the FCC should be considered much more as arms of Congress than of the Executive Branch. He observed that in *Humphrey's Executor*,[39] the Supreme Court "made clear, however, [that] when Congress grants rulemaking and adjudicative authority to an expert agency composed of commissioners selected through a bipartisan procedure and appointed for fixed terms, it substantially insulates the agency from executive control."[40] Having in

34. *Fox*, 129 S. Ct. at 1829 (Breyer, J., dissenting).

35. *Id.*

36. *Id.* at 1830.

37. *Id.*

38. *See id.* at 1832–41. Justice Breyer discussed the FCC's failure, in his view, to sufficiently address the First Amendment implications of the change in its "fleeting expletives" policy and also the adverse financial impact on local broadcasters stemming from the requirements imposed by the new policy, such as the need to purchase time delay equipment. *Id.*

39. Humphrey's Ex'r v. United States, 295 U.S. 602 (1935).

40. *Fox*, 129 S. Ct. at 1825 (Stevens, J., dissenting). In *Humphrey's Executor*, the Court

mind these institutional agency characteristics, Justice Stevens declared that independent agencies are better viewed as agents of Congress, quoting *Humphrey's Executor* to the effect that these agencies are established "to carry into effect legislative policies embodied in the statute in accordance with the legislative standard therein prescribed, and to perform other specified duties as a legislative . . . aid."[41]

The upshot for purposes of reviewing agency action, according to Justice Stevens, is that "[t]here should be a strong presumption that the FCC's initial views, reflecting the informed judgment of independent commissioners with expertise in the regulated area, also reflect the views of the Congress that delegated the Commission authority to flesh out details not fully defined in the enacting statute."[42] In this instance, this strong presumption that the FCC's initial views properly reflected congressional intent meant that it "makes eminent sense to require the Commission to justify why its prior policy is no longer sound before allowing it to change course."[43] Of course, Justice Stevens probably did not mean to imply that the FCC did not offer any justification, just not one that, in his view, was sufficient.

In the face of these dissents, Justice Scalia, in the portion of his opinion commanding only a plurality,[44] characteristically gave no ground. According to Justice Scalia, "the independent agencies are sheltered not from politics but from the President, and it has often been observed that their freedom from presidential oversight (and protection) has simply been replaced by increased subservience to congressional direction."[45] Justice Scalia asserted that the change in policy at issue in *Fox* "was spurred by significant political pressure from Congress."[46] He characterized Justice Stevens's view of the relationship between Congress and the FCC as a "principal–agency relationship,"[47] which he suggested might be unconstitutional on separation-of-powers grounds if one were to take this

---

had said that Congress intended to create "a body which shall be independent of executive authority, *except in its selection*, and free to exercise its judgment without leave or hindrance of any other official or any department of the government." 295 U.S. at 625–26.

41. *Fox*, 129 S. Ct. at 1825 (quoting *Humprey's Executor*, 295 U.S. at 628).

42. *Id.* at 1826.

43. *Id.* Justice Stevens also suggested that the Communications Act, 47 U.S.C. §§ 151–614 (2006), the APA's judicial review provision, 5 U.S.C. § 706(2)(A) (2006), and the rule of law "all favor stability over administrative whim." *Id.*

44. Chief Justice Roberts and Justices Thomas and Alito joined this portion of Justice Scalia's opinion. *See id.* at 1815–19 (plurality opinion).

45. *Id.* at 1815.

46. *Id.* at 1815–16.

47. *Id.* at 1816.

principal–agency relationship seriously.[48] Despite such intimation of unconstitutionality, Justice Scalia, referring merely to statements made by representatives at two congressional committee hearings,[49] concluded, "If the FCC is indeed an agent of Congress, it would seem an adequate explanation of its change of position that Congress made clear its wishes for stricter enforcement."[50]

In any event, apart from the degree of congressional (or presidential) control exerted, Justice Scalia found no "applicable law" in the APA, or otherwise, requiring that rulemaking by independent agencies be subject to "heightened scrutiny."[51] Curiously, Justice Scalia stated that "it is hard to imagine any closer scrutiny than that we have given to the Environmental Protection Agency, which is not an independent agency."[52] And, just as curiously, Justice Scalia concluded this portion of his opinion by stating, "There is no reason to magnify the separation-of-powers dilemma posed by the Headless Fourth Branch by letting Article III judges—like jackals stealing the lion's kill—expropriate some of the power that Congress has wrested from the unitary Executive."[53]

As I will explain in the next section, I believe Justice Scalia's view—that for purposes of applying deference independent and executive branch agencies should be treated alike—not only magnifies the separation-of-powers dilemmas inherent in the nature of independent regulatory agencies

---

48. The intimation came in the form of Justice Scalia's throwaway line, "Leaving aside the unconstitutionality of a scheme giving the power to enforce laws to agents of Congress . . . ." *Id.* The line was followed immediately by Justice Scalia's statement that "[*i*]*f* the FCC is indeed an agent of Congress," then the fact that Congress made clear its wishes for stricter indecency enforcement should suffice for an adequate reason for changing the agency's policy. *Id.* (emphasis added).

49. *See id.* at 1816 n.4.

50. *Id.* at 1816.

51. *Id.* at 1817.

52. *Id.* This is a curious statement because *Chevron* itself involved a statutory interpretation by the Environmental Protection Agency (EPA). The whole point of *Chevron* is that in light of the political accountability of executive branch agencies such as EPA, their rulings are owed deference as long as they are reasonable. It is somewhat jarring, then, to see Justice Scalia declaring that it is difficult to imagine any closer scrutiny than the Court has given EPA actions.

53. *Id.* (citation omitted). Justice Scalia supposes that subjecting decisions of independent agencies to closer scrutiny than those of executive branch agencies magnifies separation-of-powers problems, perhaps by calling further attention to these constitutional anomalies in which executive, legislative, and judicial functions are exercised by the same entity. As I made clear in *Defining Deference Down*, May, *supra* note 1, at 451, my view is that by giving less deference to independent agencies' decisions, courts might at least mitigate to some extent separation-of-powers concerns. Thus, I stated, "[I]t is odd in a constitutional system with three defined branches for courts to give controlling deference to agencies that, not without reason, are commonly referred to as 'the headless fourth branch.'" *Id.*

but is inconsistent with the principal political accountability rationale of *Chevron*.

III. "THE HEADLESS FOURTH BRANCH" WARRANTS LESS JUDICIAL DEFERENCE

Before addressing the way in which Justice Scalia dealt in *Fox* with the separation-of-powers concerns that he acknowledged existed and which, after all, are central to the political accountability rationale upon which *Chevron* principally rests, I will acknowledge that Justice Scalia is correct that, on its face, the APA does not distinguish between executive and independent regulatory agencies for purposes of review of agency action.[54] But, of course, it does not preclude such differentiation either. *Chevron* itself did not even refer to the APA review provision, even as the Court established a new deference requirement relating to judicial review that governs large amounts of agency action.

While this omission in *Chevron* may seem somewhat odd, it is not illogical to the extent the question whether an agency action is "arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law,"[55] is not necessarily coincident with the question of how much deference should be given the agency in deciding whether the action complies with the § 706 standard.

Thus, when the *Chevron* Court said that "controlling weight" should be given to the agency's statutory interpretation when Congress has left a gap to be filled,[56] it was not necessarily purporting to change the substantive meaning of the arbitrary and capricious test. Rather, its action can be viewed as an effort to tip the scale decidedly in the agency's direction by the weight accorded to the agency's interpretation. Formulating different degrees of deference in reviewing agency actions—such as according "controlling weight" or not—is not unlike the Supreme Court formulating different degrees of scrutiny—"strict," "intermediate," or "rational basis"—in assessing the constitutionality of laws, or common law courts or legislatures devising different evidentiary standards, such as "preponderance of the evidence" or "substantial evidence." In short, I do not see the APA as a bar to applying a less deferential standard of review to the actions of independent agencies.

While Justice Scalia did not make much of the point, it is also true that the *Chevron* Court referred to Environmental Protection Agency's (EPA's) "expertise" in implementing a regulatory regime that is "technical and

54. *Fox*, 129 S. Ct. at 1817.
55. 5 U.S.C. § 706(2)(A) (2006).
56. Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837, 844 (1984).

complex" as a basis for giving deference to EPA's determination.[57] And the Court observed that EPA "considered the matter in a detailed and reasoned fashion."[58] No doubt recognition of agency expertise is a factor supporting deference to an agency's determination, all the more so in areas that are especially technical and complex. Both executive branch agencies, such as EPA, and independent agencies, such as FCC, possess such institutional expertise, and they are both often called on to make decisions on technical and complex matters. Nevertheless, the fact remains that *Chevron* deference was not premised principally upon agency expertise, but rather upon the notion that there should be political accountability for policy choices that Congress did not make itself.[59] As the Court put it, "federal judges—who have no constituency—have a duty to respect the legitimate policy choices made by those who do."[60] Because executive and independent agencies are not politically accountable for making policy in the same way, agency expertise, while not irrelevant, is not a reason in and of itself to require that both types of agencies be treated alike for purposes of fashioning a deference standard.

In response to the dissents, Justice Scalia does not argue that the APA by its terms requires that independent and executive branch agencies be treated alike for purposes of judicial review. Nor does he argue that the fact that both types of agencies possess expertise relevant to their institutional tasks requires like treatment. Rather, he ultimately places the most weight upon the notion that not subjecting the independent agencies to more searching judicial scrutiny avoids magnifying the separation-of-powers dilemmas posed by the "Headless Fourth Branch."[61] Justice Scalia's approach not so much avoids magnifying separation-of-powers problems as, with some sleight of hand, it downplays them. He accomplishes this by exaggerating the extent to which the independent agencies are politically accountable to Congress, while at the same time fully acknowledging that they are not accountable to the President. Specifically, Justice Scalia states, "The independent agencies are sheltered not from politics but from the

57. *Id.* at 865.

58. *Id.*

59. In contrast, in *Humphrey's Executor*, in the course of highlighting the FTC's freedom from executive control, the Court touted the agency's "body of experts who shall gain experience by length of service" as a distinguishing characteristic of the agency's independence. 295 U.S. 602, 625 (1935).

60. *Chevron*, 467 U.S. at 866.

61. The term *headless fourth branch* was used to describe the independent agencies by a presidential management commission in 1937 studying the organization and management of the federal government. PRESIDENT'S COMM. ON ADMIN. MGMT., REPORT OF THE COMMITTEE WITH STUDIES OF ADMINISTRATIVE MANAGEMENT IN THE FEDERAL GOVERNMENT 40 (1937).

President, and it has often been observed that their freedom from presidential oversight (and protection) has simply been replaced by increased subservience to congressional direction."[62] In my view, he magnifies congressional control too much.

In support of his assertion concerning subservience to Congress, Justice Scalia cites a footnote in Elena Kagan's *Presidential Administration* article to the effect that "[a]s a practical matter, successful insulation of administration from the President—even if accomplished in the name of 'independence'—will tend to enhance Congress's own authority over the insulated activities."[63] It may be that Solicitor General Kagan believes that successful insulation of independent agencies from presidential control has the effect of enhancing Congress's own authority. But that is a far cry from concluding that congressional control is such that it puts independent agencies under Congress's thumb (and certainly not under its thumb based on a few statements by representatives at congressional hearings, which was the factual context of the alleged congressional influence in *Fox*).[64] After all, the pertinent question for separation-of-powers purposes really is not whether Congress's authority might be somewhat *enhanced* by the lack of presidential control, but rather whether the *extent* of congressional control rises to the level of ensuring the meaningful political accountability which separation of powers is designed to ensure.

Indeed, Solicitor General Kagan's view appears to be distinctly different from that which Justice Scalia assumed when he cited her article for support. Further along in *Presidential Administration*, when she explicitly advocates giving less *Chevron* deference to the decisions of independent agencies than to those of executive agencies, Kagan makes quite clear that she views the independent agencies as not sufficiently accountable to *either* the President or Congress to justify according them the same deference accorded to the more politically accountable executive agencies. Apart from what she calls "the institutional characteristics that make Congress a less reliable overseer of agency action than the President,"[65] Kagan emphasizes that "the constitutional limits on Congress's ability to establish a hierarchical relationship with the independent agencies (most notably, by retaining removal power over their heads) preclude equating the two kinds of control."[66]

---

62. FCC v. Fox Television Stations, Inc., 129 S. Ct. 1800, 1815 (2009) (plurality opinion).

63. *See id.* (citing Kagan, *supra* note 12, at 2271 n.93).

64. *Id.* at 1816 n.4.

65. Kagan, *supra* note 12, at 2377 n.506.

66. *Id.* (citation omitted).

In other words, as I asserted in *Defining Deference Down*,[67] the presidential removal power is the key to the executive agencies' political accountability. Surely both the President and Congress each have various means of exercising influence over the independent agencies. But they are both alike in lacking the critical ability to remove agency commissioners without cause. In *Humphrey's Executor*, the Supreme Court put the point plainly: "For it is quite evident that one who holds his office only during the pleasure of another, cannot be depended upon to maintain an attitude of independence against the latter's will."[68] This lack of removal power, what Kagan calls the "core legal difference" between independent and executive branch agencies, is what, combined with their unique organizational structure, gives the independent agencies their independence.[69]

In his famous dissent in *Morrison v. Olson*,[70] Justice Scalia argued (persuasively in my opinion) that the then-existing "independent counsel" statute was unconstitutional as a violation of separation of powers because of limitations placed on the President's authority to remove the counsel except upon good cause. Referring to *Humphrey's Executor*, he pointed out, with respect to the independent counsel, that the limitation on the President's removal power constituted an effective impediment to presidential control.[71] Indeed, Justice Scalia argued that the removal limitation constituted such an impediment to presidential control that this diminishment of executive authority violated the separation of powers that he described as so central to the preservation of our liberties.[72]

The relevance of Justice Scalia's *Morrison* dissent to *Fox* is this: In *Morrison* he recognized, as had the Court years earlier in *Humphrey's Executor*, the centrality of the removal power to the independence vel non of government officials. After all, the effect of this "power to fire," or "coercive influence" as *Humphrey's Executor*[73] put it, is only common sense logic. But in *Fox*, Justice Scalia ignored the fact that Congress lacks the authority to remove independent agency commissioners, absent impeachment proceedings.[74]

---

67. May, *supra* note 1, at 447–48; *see also* 1 KENNETH CULP DAVIS & RICHARD PIERCE, JR., ADMINISTRATIVE LAW TREATISE § 2.5 (3d ed. 1994) ("The characteristic that most sharply distinguishes independent agencies is the existence of a statutory limit on the President's power to remove the head (or members) of an agency.").

68. Humphrey's Ex'r v. United States, 295 U.S. 602, 629 (1935).

69. Kagan, *supra* note 12, at 2376.

70. 487 U.S. 654, 706 (1988) (Scalia, J., dissenting).

71. *Id.* at 706.

72. *See id.* at 706–07.

73. 295 U.S. at 630.

74. *Cf.* Bowsher v. Synar, 478 U.S. 714, 726 (1986) ("[W]e conclude that Congress cannot reserve for itself the power of removal of an officer charged with the execution of the laws except by impeachment."). Impeachment proceedings rarely, if ever, have been

Thus, while no one doubts Congress has the means to influence agency actions through investigatory or oversight hearings, such as those to which Justice Scalia referred in his *Fox* opinion, or through other means such as the confirmation and appropriations processes,[75] these means, absent the removal power, do not give rise to the same degree of political accountability for policymaking upon which the *Chevron* rationale primarily rests. Once again, recall that in *Chevron* the Court pointed to the political accountability of EPA as part of the "incumbent administration."[76] By deliberate design, and by virtue of the Supreme Court's decision in *Humphrey's Executor*, the independent agencies are not considered part of the incumbent administration and they do not enjoy—or suffer, as the case may be—the same degree of political accountability.

CONCLUSION

Certainly, there is a respectable body of opinion that *Humphrey's Executor*, in insulating the independent agencies from presidential control, is constitutionally suspect on separation-of-powers grounds.[77] Professors Lessig and Sunstein, for example, have stated "the case was a bizarre and unfounded exercise in constitutional innovation,"[78] an innovation that threatens "the core constitutional commitments to political accountability, expedition in office, and coordinated policymaking."[79] However bizarre and unfounded *Humphrey's Executor* may be, the constitutional sanction it gave to independent agencies like the FCC seems now embedded in our constitutional culture, despite nonfrivolous separation-of-powers concerns.

But the fact that the constitutional status of the independent regulatory agencies does not appear to be threatened per se does not mean that, in reviewing their actions, courts should not strive to act consistently with, or at least to not diminish, "the core constitutional commitments to political accountability."[80] In the main, *Chevron* deference is primarily all about this constitutional commitment to political accountability. And the debate that

---

instituted against independent agency officials. I am unaware of such a case.

75. There is a vast literature on the myriad ways that Congress can exercise influence on agency actions. For a good source, with citation to many authorities, see Jack M. Beermann, *Congressional Administration*, 43 SAN DIEGO L. REV. 61 (2006).

76. Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837, 865–66 (1984).

77. For further analysis on this issue and additional sources, see May, *supra* note 1, at 450 nn.117–19.

78. Lawrence Lessig & Cass R. Sunstein, *The President and the Administration*, 94 COLUM. L. REV. 1, 101 (1994).

79. *Id.* at 114.

80. *Id.*

emerged in *Fox* between Justices Scalia, Breyer, and Stevens concerning review of the FCC's changed indecency policy, with the back-and-forth exchange concerning the extent to which the agency was subject to congressional control, revolves around the constitutional commitment to political accountability. Although the *Fox* opinions did not directly invoke *Chevron*, they definitely sounded in *Chevron* in their invocations of the relevance of political accountability to a more or less deferential standard of review of an independent agency's actions.

At the end of the day in *Fox*, Justice Scalia's view, embodied in his plurality opinion, prevailed—that is, the actions of independent agencies are not subject to any form of heightened scrutiny on review as a result of the agencies' status. I think Justice Scalia's view is based on an exaggerated notion of congressional control of the independent agencies' actions that assumes a greater degree of agency political accountability than is warranted. He accepts, rather uncritically, the notion that the independent agencies are insulated from presidential control. But in *Fox* he does not confront the reality that it is the limitation on presidential removal power of agency heads which is at the heart of such insulation and that the absence of the removal power similarly limits congressional control of the independent agencies. While Justice Scalia professed a desire to avoid magnifying separation-of-powers problems, in my view his approach achieves just the opposite by, in effect, derogating the core commitment to political accountability that constitutional separation of powers embodies.

With *Defining Deference Down*, based on what I see as the principal political accountability rationale underpinning *Chevron*, my project was to begin a more robust dialogue concerning whether a less deferential judicial review standard of independent agency actions would be more consistent with core separation-of-powers values. While I expect that *Fox* will be seen first and foremost through the lens of a more conventional administrative law "change of agency policy" case, I have hopes that it will also be an impetus for the dialogue that I aim to further with this follow-on article. For regardless of the outcome, a discussion relating to the impact of judicial review doctrines on separation of powers and political accountability is never out of place in our democratic republic. Indeed, it is to be welcomed.

* * *

A SURVEY OF FEDERAL AGENCY RULEMAKERS' ATTITUDES ABOUT E-RULEMAKING

JEFFREY S. LUBBERS*

TABLE OF CONTENTS

I. BACKGROUND ON RULEMAKING IN THE UNITED STATES

In the United States, the 1946 Administrative Procedure Act (APA) contains the general requirements for federal agency promulgation of regulations. This procedure is often called notice-and-comment rulemaking, deriving from the fact that the operative APA section requires

* Professor of Practice in Administrative Law, American University Washington College of Law. This Article was prepared originally to appear as a chapter in the forthcoming book CONNECTING DEMOCRACY: ONLINE CONSULTATION AND THE FUTURE OF DEMOCRATIC DISCOURSE (Stephen Coleman & Peter M. Shane eds., anticipated 2010).

(1) publication of a notice of proposed rulemaking, (2) opportunity for public participation in the rulemaking by submission of written comments, and (3) publication of a final rule and accompanying statement of basis and purpose not less than thirty days before the rule's effective date.

These requirements may be exceeded by agencies voluntarily or pursuant to other programmatic statutes that provide more elaborate public procedures. However, even this procedural floor does not apply to all rulemaking. Certain types of rules are exempted from some of these requirements, and entire classes of rules are totally exempted from APA notice-and-comment requirements. These exemptions reflect the APA drafters' cautious approach to imposing procedural requirements on a myriad of agency functions, as well as their willingness, in some situations, to permit agencies a measure of discretion in fashioning procedures appropriate to the particular rulemaking involved. This basic APA model has proved successful and is being emulated around the world.[1]

A. Electronic Rulemaking (e-Rulemaking)

With the technological revolution wrought by the Internet, the character of rulemaking is changing. What once was an all-paper process—with paper notices published in a paper *Federal Register*, paper comments submitted by hand or by post to the agency and filed in a filing cabinet in a room in the bowels of an agency—has been largely replaced by an electronic process with electronic notices, comments, and dockets available for anyone around the world to access with a click of a computer mouse. The U.S. Government has established a government-wide web portal that allows the public to file comments on any pending rule.[2]

Much has been written about this "rulemaking revolution," even though it is clearly in its early stages.[3] The main touted benefits from e-

1. *See, e.g.*, Jeffrey S. Lubbers, *Notice-and-Comment Rulemaking Comes to China*, ADMIN. & REG. L. NEWS, Fall 2006, at 5, 5–6 (describing recent experiments by large Chinese municipalities with public comment procedures); Katsuya Uga, *Development of the Concepts of "Transparency" and "Accountability" in Japanese Administrative Law*, 1 U. TOKYO J.L. & POL. 25, 36–38 (2004) (describing the public comment procedures in Japan). However, for a lamentation about how the "basic model" has become overly laden with other review and analysis requirements in the United States, see Jeffrey S. Lubbers, *The Transformation of the U.S. Rulemaking Process—For Better or Worse*, 34 OHIO N.U. L. REV. 469, 473–78 (2008).

2. The website address is http://www.regulations.gov (last visited Mar. 22, 2010). For a comprehensive discussion of the history, goals, and remaining challenges of this effort, see COMMITTEE ON THE STATUS AND FUTURE OF FEDERAL E-RULEMAKING, ACHIEVING THE POTENTIAL: THE FUTURE OF FEDERAL E-RULEMAKING (2008), http://resource.org/change.gov/ceri-report-web-version.fixed.pdf.

3. Much of the following discussion is derived from JEFFREY S. LUBBERS, A GUIDE TO FEDERAL AGENCY RULEMAKING 217–39 (4th ed. 2006). For a succinct history of the "rise of

rulemaking, of course, are increased opportunities for information dissemination, public participation, and governmental transparency, along with better outcomes and greater trust in government. Commenters can now e-mail their comments to the agency with just a keystroke and agencies can post all comments on their websites for everyone in cyberspace to read and react to. The days of having to travel to Washington to physically visit a dusty records repository are over. Possibilities abound for enhancing the entire notice-and-comment process.[4]

In e-rulemaking, notices can be improved and more widely disseminated.[5] Automatic notices can be generated by request to individuals who have requested them. Notices can be made word-searchable, and alternative or revised drafts can be posted with the changes clearly designated. Moreover, related studies, required draft regulatory analyses, and other information can be linked to the notices to provide easier public access. The comment process can also be made much more "user-friendly" and responsive to agency needs through the use of request-for-comments forms, the segmentation of proposed rules for comments, and opportunities to file reply comments[6]—even producing "threads" of comments on particular issues. And the final stage of rulemaking can be

---

e-rulemaking," see Cary Coglianese, *E-Rulemaking: Information Technology and the Regulatory Process*, 56 ADMIN. L. REV. 353, 363–66 (2004). *See also* Stuart W. Shulman, *E-Rulemaking: Issues in Current Research and Practice*, 28 INT'L J. PUB. ADMIN. 621 (2005); Beth Simone Noveck, *The Electronic Revolution in Rulemaking*, 53 EMORY L.J. 433 (2004); Barbara H. Brandon & Robert D. Carlitz, *Online Rulemaking and Other Tools for Strengthening Our Civil Infrastructure*, 54 ADMIN. L. REV. 1421 (2002); Stephen Zavestoski & Stuart W. Shulman, *The Internet and Environmental Decision Making: An Introduction*, 15 ORG. & ENV'T 323, 326 (2002). Links to some of these and many other related papers and studies are available on the website of the Harvard University's John F. Kennedy School of Government's Regulatory Policy Program. John F. Kennedy School of Government, E-Rulemaking Papers & Reports, http://www.hks.harvard.edu/m-rcbg/rpp/erulemaking/papers&reports.htm (last visited Mar. 22, 2010).

4. Note, however, that the Administrative Procedure Act's (APA's) notice requirement is not met when an agency gives notice of a proposed rule only on the Internet instead of in the *Federal Register*. Util. Solid Waste Activities Group v. EPA, 236 F.3d 749, 754 (D.C. Cir. 2001).

5. Many of the ideas in this paragraph for enhanced citizen participation through e-rulemaking are discussed more fully in Noveck, *supra* note 3, at 471–94.

6. As one agency expert described it,

> [W]e can say the comment period ends on November 1st. From November 1st, for example, to December 1st, we're going to allow anybody to come back and reply to what someone else has said. Not say something new, but reply to what others said. It will help the agency, at least theoretically, [to] more efficiently address the comments that they've received.

Neil Eisner, Dep't of Transp., Comments at American University's Center for Rulemaking's E-Rulemaking Conference 77 (Jan. 8, 2004), http://www.american.edu/academic.depts/provost/rulemaking/transcripts.pdf.

enhanced through new publication techniques, such as linking all other related regulatory documents and final regulatory analyses, and grouping comments and the agency's response.

Others have focused on the possibilities of using these electronic tools for more *interactive* rulemaking.[7] Suggestions for "deliberative dialogue[s],"[8] online chat rooms,[9] or electronic negotiated rulemaking concerning proposed regulations have proliferated, but so far their potential is untapped.[10]

It remains to be seen whether e-rulemaking will revolutionize public participation. As one leading commentator has concluded, "Electronic rulemaking may transform the process fundamentally or it may simply digitize established paper-based processes."[11] The route that e-rulemaking takes in the future may depend on how well a series of legal and technical questions can be answered.[12]

But if the process is to be transformative, this transformation of the rulemaking (and docketing) process should be viewed as having two main purposes. The first is an *informational* one of providing a global, seamless view of each rulemaking, and the second is a *participatory* one.

Achieving the informational goal means providing access to every meaningful step in the generation of a rule, from the statute enacted by Congress that authorizes the rule to the earliest agency action (perhaps an "advance notice of proposed rulemaking") to the last step in the process—whether it be the final rule, a decision in a court challenge, or later agency amendments, interpretations, guidelines, or enforcement actions.[13] It also means that the public should be provided a "vertical" view of pending or final rules—what might be called "drilling down" into the meaningful

7. *See, e.g.*, Thomas C. Beierle, *Discussing the Rules: Electronic Rulemaking and Democratic Deliberation* 8–13 (2003) (Res. for the Future, Discussion Paper No. 03-22, 2003), http://www.rff.org/rff/Documents/RFF-DP-03-2.pdf.

8. Noveck, *supra* note 3, at 499.

9. Stephen M. Johnson, *The Internet Changes Everything: Revolutionizing Public Participation and Access to Government Information Through the Internet*, 50 ADMIN. L. REV. 277, 321–24 (1998) (discussing early experiments by the Nuclear Regulatory Commission).

10. *See, e.g.*, Beierle, *supra* note 7, at 8 (discussing some agency attempts to use dialogues in rulemaking).

11. STUART W. SHULMAN, THE INTERNET STILL MIGHT (BUT PROBABLY WON'T) CHANGE EVERYTHING: STAKEHOLDER VIEWS ON THE FUTURE OF ELECTRONIC RULEMAKING 35 (2004), http://erulemaking.ucsur.pitt.edu/doc/reports/e-rulemaking_final.pdf.

12. The following discussion is adapted from Jeffrey S. Lubbers, *The Future of Electronic Rulemaking: A Research Agenda* (John F. Kennedy Sch. of Gov't, Harvard Univ., Regulatory Policy Program, Working Paper No. RPP-2002-04, 2002), http://www.hks.harvard.edu/m-rcbg/research/rpp/RPP-2002-04.pdf, *reprinted in* ADMIN. & REG. L. NEWS, Summer 2002, at 6.

13. I am indebted to Professor Cary Coglianese for this insight.

agency and outside studies and analyses that are now found in the docket, along with the public comments, for any significant proposed and final rule—and, where possible, through links into those secondary studies and analyses referenced in the primary studies.

The participatory goal of the transformation of rulemaking is ultimately to make it possible for participants to participate in real time with other stakeholders in a rulemaking process (an idealized "chat room") that will allow a more rational, interactive, and less adversarial path to an optimum final rule. And as information-filtering technologies (à la Google) become more sophisticated and allow more tailoring for individualized needs, commenters will also be able to zero in on their particular interests and contribute more targeted comments.[14]

Both the informational and participatory goals raise issues which require further research and experimentation. Informational issues include: the ways to best integrate existing sources of information and docketing concerns, such as those related to scanning, archiving, handling of attachments, copyright, authentication, security, and privacy. Participatory issues include: how to best reach the goal of better, more targeted notices; the possibility of providing easier, more convenient comment opportunities; what rules should govern rulemaking "chatrooms"; and the broad question of electronic "negotiated rulemaking."[15]

B. Impact of e-Rulemaking on the Agencies

The flip side of increased public participation, of course, is increased responsibilities on agencies to digest and react to a higher volume of comments. Blizzards of comments have become increasingly common in controversial rulemakings, and e-rulemaking can only further this trend. Professor Strauss has warned of some of the problems this might cause:

14. Professor Stuart Shulman persuasively illustrated this last point in a presentation at the Fall 2005 meeting of the American Bar Association Section of Administrative Law and Regulatory Practice.

15. For more on these issues, see LUBBERS, *supra* note 3, at 226–36.

> I think we're going to see an enormous explosion in the volume of rulemaking comments, and some of them will be quite manipulative. And it will be a challenge for the agencies receiving these comments to tell the one from the other, the valid from the invalid. And then, once they have received hundreds of thousands, tens of thousands of comments, the impulse to treat them as a reflection of e-democracy—we're hearing from the people, and what we do ought to reflect the people, rather than we are collecting information and what we ought to do ought to reflect the outcome of that information—is going to be quite strong.[16]

Professor Herz concurs that this may be a problem:

> What can realistically be expected of an agency dealing with a million comments, thousands of which duplicate one another? The old model of careful individual consideration is inapplicable. Unavoidably, the agency will start to do what, for example, members of Congress do: avoid the subtleties and keep a running tally with the grossest sort of division—basically "for" or "against."[17]

This, he cautions, may not only lead to "information overload"[18] (although technology may also make it possible for agencies to efficiently sort and categorize voluminous comments),[19] it might lead to a general politicization of the rulemaking process, moving away from the technocratic model of rulemaking, where the substance of the comment is

---

16. Peter Strauss, Comments at American University's Center for Rulemaking's E-Rulemaking Conference, *supra* note 6, at 28.

17. Michael Herz, *Rulemaking*, *in* DEVELOPMENTS IN ADMINISTRATIVE LAW AND REGULATORY PRACTICE 2002–2003, at 129, 148–49 (Jeffrey S. Lubbers ed., 2004). He also points out, "There is one important caveat, however. To the extent that the comments are duplicative, the burden of responding is not increased." *Id.* at 149 n.78.

18. *Id.* at 149; *see also* Randolph J. May, *Under Pressure: Campaign-Style Tactics Are the Wrong Way to Influence Agency Decisions*, LEGAL TIMES, July 7, 2003, at 44 (referring to a mass e-mail, post card, and call-in campaign which resulted in the Federal Communications Commission receiving 750,000 e-mails in response to a deregulatory initiative and rulemaking); Jim Rossi, *Participation Run Amok: The Costs of Mass Participation for Deliberative Agency Decisionmaking*, 92 NW. U. L. REV. 173, 224–28 (1997) (maintaining that although increased participation can result in greater amounts of information available to decisionmakers and participants, this may lead agency decisionmakers to "miss the forest for the trees").

19. *See* Professor Stuart Shulman, Univ. of Pittsburgh, Comments at American University's Center for the Study of Rulemaking, Panel 4: Participation in Rulemaking 15 (Mar. 16, 2005), http://www.american.edu/academic.depts/provost/rulemaking/transcripts.pdf ("Part of what we're doing with the computer scientists is developing tools for dealing with this information flood, and we're making some progress . . . where we'll be able to deliver a tool to agency personnel who want to identify [as] quickly as possible those clusters of duplicate and near-duplicate e-mails."). For a technical paper describing these promising techniques for sorting comments, see Hui Yang & Jamie Callan, *Near-Duplicate Detection for eRulemaking*, *in* PROCEEDINGS OF THE SIXTH NATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH (2005), http://erulemaking.ucsur.pitt.edu/doc/papers/dgo05-huiyang.pdf.

more important than who submitted it or how many times it was repeated, to a type of referendum.[20] "In short," he notes, rather disquietingly, "the new technology is forcing agencies toward a particular model of the process and function of rulemaking, as opposed to enabling agencies to better function under the model chosen independent of that technology."[21] Other researchers have found a proliferation of "form comments,"[22] making Professor Noveck's concern about the use of robot programs to generate "notice and spam" all the more disquieting.[23]

II. THE SURVEY

To find out how the advent of e-rulemaking is perceived among federal rulemakers, I designed and distributed a survey to rulemakers, using an electronic survey program.[24] After designing the questions with helpful constructive criticism from Professor Peter Strauss and a very experienced rulemaking supervisor from the U.S. Department of Transportation, Neil Eisner, I circulated it to Mr. Eisner and other such supervisors and asked that they encourage their rulemaking staffers to take this survey.

---

20. Professor Herz points to the example of the "roadless rule," a heavily litigated rule issued in the waning days of the Clinton Administration, which attempted to restrict road construction in large parts of Forest Service land:

> The rule has generated a number of legal challenges, with several district judges finding defects in the process, and the Bush Administration is considering diluting its protections in Alaska. Comments on the proposed rule and/or the Draft EIS, and on the current Alaska proposals, numbered in the millions and have been overwhelmingly in favor of stringent protections. Press coverage has overwhelmingly treated the comment process as a sort of vote. This conception can also be seen in an amicus brief submitted to the Ninth Circuit in *Kootenai Tribe* by the Montana Attorney General. The brief's basic point had nothing to do with legality, but came down to this: "Hey, Montanans overwhelmingly support this rule, as shown by tabulating our comments during the process." Emphasizing that 67 percent of commenters in Montana (and 96 percent nationwide) favored stronger protections than were anticipated in the Draft EIS, and that the Forest Service responded by strengthening protections, the brief concludes that the rule is "the product of public rulemaking at its most effective." What's more, the Ninth Circuit placed some weight on this argument.

Herz, *supra* note 17, at 150–51 (footnotes omitted).

21. *Id.* at 151.

22. *See* David Schlosberg, Stephen Zavetoski & Stuart Shulman, To Submit a Form or Not to Submit a Form, That is the (Real) Question: Deliberation and Mass Participation in U.S. Regulatory Rulemaking (May 5, 2005) (unpublished manuscript), http://erulemaking.ucsur.pitt.edu/doc/papers/SDEST_stanford_precon.pdf (finding significant differences between respondents who submitted original comments and those who submitted form letters). For more such research, visit the website of the e-rulemaking group at the University of Pittsburgh, http://erulemaking.ucsur.pitt.edu.

23. Noveck, *supra* note 3, at 441.

24. I used Survey Monkey (professional subscription), www.surveymonkey.com.

The survey is intended to be exploratory. As such, I used a combination of convenience and snowball sampling because the desired sample characteristics (in this case federal rulemakers who use e-rulemaking) are not that numerous or identifiable and not easy to access. I relied primarily on referrals from the federal rulemaking supervisors.[25] Thus, the sample is small and results may not be completely representative of the e-rulemaking population. Nonetheless, the survey of federal rulemakers is the first of its kind on this topic and does provides some insights and early indications of the attitudes and perceptions of those on the "firing line" of this new technology.

After a little more than a month of collecting responses, I had amassed seventy-four responses from a wide variety of agencies. The breakdown was as follows:[26]

Agency	**No. of Responses**
Department of Transportation (DOT)	17
Department of Homeland Security (DHS)	12
Environmental Protection Agency (EPA)	8
Department of Labor (DOL)	8
Department of the Treasury (Treasury)	7
Department of Energy (DOE)	5
Department of Veterans Affairs (DVA)	4
Department of Commerce (DOC)	2
Department of the Interior (DOI)	2
Department of Health & Human Services (HHS)	1
Department of Housing & Urban Development (HUD)	1
Federal Communications Commission (FCC)	1
Federal Election Commission (FEC)	1
Unidentified	4

25. Neil Eisner periodically convenes a "brown bag lunch group" of his peers from other agencies. I circulated an e-mail to each of them with the survey link and asked for their help in circulating it to their staffs. I also used the Federal Yellow Book to look for other such supervisors and sent e-mails to those that I found. Finally, I attended a conference of agency rulemakers and solicited their cooperation.

26. Some respondents also mentioned their subagencies. For example, five of the Department of Transportation's respondents were from the Federal Aviation Administration, and five were from other different subagencies. Five of the Department of Homeland Security's respondents were from the Coast Guard, and three from the Transportation Safety Administration.

A large majority of the respondents spent most of their work time on rulemaking activities:

Percentage of Time	**No. of Respondents**
100%	20
90–99%	17
75–89%	14
50–74%	9
25–49%	7
10–24%	5
5%	1
No answer	1

About three-fifths of the respondents described themselves as "more of a line employee" (n=45) and two-fifths as more of a "supervisor" (n=29).

Most were attorneys:

Position	**No. of Respondents**
Attorney	47
Policy Expert in the Field	8
Technical Expert in the Field	5
Economist	2
Political Scientist	1
Other	11

The "other" category included four "regulations analysts," including one with a J.D. degree; two "writer-editors"; two "IT"; one "project manager"; and one with a "varied" background.

The age of the respondents skewed rather high:

Age Range	**No. of Respondents**
Below 30	7
30–39	14
40–49	21
50–59	22
Above 60	10

Rulemaking experience was also correspondingly high, but was well distributed:

Years of Experience	**No. of Respondents**
0–2	12
3–5	11
5–10	18
10–20	16
20+	17

Most of the respondents had worked with rulemaking both before and after the advent of e-rulemaking (forty-seven), although eleven had worked only with the new system. (Fifteen skipped this question and one had not worked at all with e-rulemaking.) Some of the questions discussed below were only asked of those that had worked before and after.

A. Positive Effects of e-Rulemaking (from the Rulemakers' Perspective)

I asked a series of sixteen questions attempting to see whether e-rulemaking has made it more or less easy to undertake some positive rulemaking activities: "When compared to the old system of paper comments, has the advent of e-rulemaking made it more difficult or easier for your agency to do the following."

I used a seven-point range: (1) Much more difficult under the new system; (2) More difficult under the new system; (3) A little more difficult under the new system; (4) The same as under the old system; (5) A little easier under the new system; (6) Easier under the new system; (7) Much easier under the new system. I also allowed an *N/A* answer ("Insufficient experience with this issue").

Sixty-four of the seventy-four respondents answered this long question, although some of those (including presumably those that had not worked with both systems) answered *N/A* for many of them. But all but two of the sixteen questions elicited at least thirty-six ranked answers.

Question 1: When compared to the old system of paper comments, has the advent of e-rulemaking made it more difficult or easier for your agency to do the following?

a. Conduct proactive notification and outreach to the public by maintaining target mailing lists (or listservs) of people who are interested in selected aspects of your rulemaking agendas?

Answer No.	**Corresponding Written Answer**	**No. of Respondents**	**Percentage**
1	Much more difficult under the new system	0	0%

2	More difficult under the new system	0	0%
3	A little more difficult under the new system	1	3%
4	The same as under the old system	11	30%
5	A little easier under the new system	3	8%
6	Easier under the new system	12	32%
7	Much easier under the new system	10	27%
	N/A	27	

Response Count: 64
Average Score: 5.51 (n=37)

Thus, only one respondent answered that it was harder to undertake targeted outreach under the e-rulemaking system and twenty-five said it was easier to some degree. The average score on this question was a high 5.51.

To save space, the full results for the remaining subparts of Question 1 are contained in the Appendix; here are the summary results:

Question	**Average Score**
b. Identify and find appropriate stakeholders?	4.81 (n=42)
c. Disseminate information relevant to the agency's proposed rulemaking (e.g., studies, economic analyses, legal analyses), so as to generate more informed commenters?	5.67 (n=46)
d. Present to the public competing or multiple alternatives to the proposed rules?	4.73 (n=37)
e. Stimulate public comments generally?	5.33 (n=51)
f. Sort and analyze public comments generally?	5.02 (n=51)
g. Obtain public comments specifically addressed to particular portions or segments of the proposed rule?	4.64 (n=47)
h. Sort and analyze public comments specifically addressed to particular portions or segments of the proposed rule?	4.70 (n=46)
i. Use the concept of "reply comments"?	5.44 (n=25)

j. Place summaries of ex parte communications in the record more quickly?	5.16 (n=38)
k. Coordinate the rulemaking internally by allowing many people to look at the same rulemaking docket without getting in each others' way?	5.70 (n=43)
l. Coordinate the rulemaking externally with O[ffice of] M[anagement and] B[udget] or other interested government entities?	5.23 (n=40)
m. Conduct interactive proceedings in rulemaking, such as "negotiated rulemaking"?	4.19 (n=16)
n. Craft a preamble to the final rule that responds to comments and includes all relevant studies and analyses?	5.05 (n=44)
o. Develop and implement appropriate archival practices relating to rulemakings (such as retiring records, etc.)?	5.25 (n=36)
p. Periodically evaluate and review the rule (and related rules), once promulgated?	5.19 (n=37)

Significantly, after tabulating an average of the ranked answers for each of the sixteen questions, all of them exceeded "4" ("same as under the old system") and twelve of them exceeded "5." This means that the advent of e-rulemaking has been "positive" for each activity. The activities with the highest average scores were "Coordinate the rulemaking internally by allowing many people to look at the same rulemaking docket without getting in each others' way" (5.70), and two activities dealing with "proactive notification and outreach" (5.51) and information dissemination (5.67). The four questions that led to only mildly positive responses were those relating to negotiated rulemaking (4.19), obtaining comments on segments of the rule (4.64), sorting such comments (4.70), and identifying and finding stakeholders (4.73).

B. Worrisome Effects of e-Rulemaking (from the Rulemakers' Perspective)

Using a similar seven-point scale, I then asked a series of ten questions attempting to see whether e-rulemaking has indeed increased the level of concern about some of the worries mentioned above. All but one of the ten questions elicited at least thirty-six ranked answers.

Question 2. When compared to the old system of paper comments, has the advent of e-rulemaking caused your agency to worry more or less about the following:

a. Outside intervention ("hacking") into your rulemaking proceedings?

Answer No.	Corresponding Written Answer	No. of Respondents	Percentage
1	Worry much more under the new system	3	8%
2	Worry more under the new system	9	24%
3	Worry a little more under the new system	12	32%
4	The same as under the old system	7	19%
5	Worry a little less under the new system	1	3%
6	Worry less under the new system	2	5%
7	Worry much less under the new system	3	8%
	N/A	25	

Response Count: 62
Average score: 3.32 (n=37)

Thus, only six respondents answered that they were less worried about hacking in the new system and twenty-four worried more to some degree. The average score on this question was a low 3.32.

Again, for brevity's sake, the full results for the remaining subparts of Question 2 are contained in the Appendix; here are the summary results:

Question	Average Score
b. Acquiring viruses via attachments submitted in comments?	3.31 (n=36)
c. Inappropriate exposure of materials in the rulemaking docket that might contain confidential business information?	3.11 (n=45)

d. Inappropriate exposure of materials in the rulemaking docket that might contain copyrighted materials?	3.20 (n=46)
e. Inappropriate exposure of materials in the rulemaking docket that might contain indecent or obscene language or materials?	3.30 (n=44)
f. Inappropriate exposure of information in the rulemaking docket that might lead to national security problems?	3.82 (n=28)
g. Risk of information destruction or other irretrievable loss of rulemaking information?	4.09 (n=43)
h. Integrating (scanned) paper comments with e-mailed or electronically submitted comments?	4.14 (n=49)
i. The authenticity of comments?	3.81 (n=47)
j. Ensuring the protection of the privacy of commenters?	3.13 (n=46)

After tabulating an average of the ranked answers for each of the ten questions, eight of them were below "4" ("same as under the old system") meaning that the advent of e-rulemaking has produced some heightened worries. The greatest worries (lowest average scores) concerned "Inappropriate exposure of materials in the rulemaking docket that might contain confidential business information" (3.11), "Ensuring the protection of the privacy of commenters" (3.13), "Inappropriate exposure of materials in the rulemaking docket that might contain copyrighted materials" (3.20), and "Inappropriate exposure of materials in the rulemaking docket that might contain indecent or obscene language or materials" (3.30). It should be noted that attorney respondents were even more worried about these last four categories (3.00, 2.90, 3.03, 3.25).

Only two of the hypothesized concerns were less worrisome under the e-rulemaking system: "Integrating (scanned) paper comments with e-mailed or electronically submitted comments?" (4.14) and "Risk of information-destruction or other irretrievable loss of rulemaking information?" (4.09).

C. Other Effects of e-Rulemaking (from the Rulemakers' Perspective)

The following questions were to be answered only by those forty-seven respondents who had indicated that they had worked with rulemaking both before and after the advent of e-rulemaking. (Those that had not were directed to skip these questions.) As the number of respondents for these answers varied only from forty-nine to fifty, it appears that this direction

was followed assiduously.[27] The survey advised respondents: "This and questions 3–11 may be difficult to answer with great certainty. Please provide your impressions as one who has been involved in rulemaking both before and after e-rulemaking."

An important issue is whether e-rulemaking has led to an increase in public comments. Only one respondent reported fewer comments, thirteen reported the same, while thirty-one reported some level of increase. The average "score" among those who provided a ranking was a high 5.36 out of 7.

2. Number of comments?

Response	**No. of Respondents**	**Percentage of Total**
Many Fewer	0	0%
Fewer	1	2%
Slightly Fewer	0	0%
The same	13	26%
Slightly More	9	18%
More	12	24%
Many More	10	20%
Don't Know	5	10%

Response Count: 50
Average Score: 5.36 (n=45)

What about the usefulness of the comments? The responses on whether the advent of e-rulemaking has led to more or fewer comments "that provide new useful information or arguments" led to a split decision. The average of the rankings here was 3.8 (or close to "the same"). Three-fifths of the respondents indicated no difference in this respect.

3. Comments with new useful information or arguments?

Response	**No. of Respondents**	**Percentage of Total**
Many Fewer	2	4%
Fewer	5	10%
Slightly Fewer	1	2%
The Same	30	60%

27. Note that it is possible that a few of those who had skipped the indicator question might have nonetheless answered some of the follow-up questions. On the other hand, a number answered each question "Don't know."

Slightly More	4	8%
More	2	4%
Many More	0	0%
Don't Know	6	12%

Response Count: 50
Average Score: 3.80 (n=44)

Not only was e-rulemaking perceived by many as not generating more useful comments, it was also strongly perceived to generate more comments "that provide only opinions without supporting facts or arguments." No one reported fewer such comments while twenty-five respondents reported an increase.

4. Comments that only provide opinions without supporting facts or arguments?

Response	**No. of Respondents**	**Percentage of Total**
Many More	10	20%
More	5	10%
Slightly More	10	20%
The Same	17	34%
Slightly Fewer	0	0%
Fewer	0	0%
Many Fewer	0	0%
Don't Know	8	16%

Response Count: 50
Average Score: 2.81 (n=42)

Even more telling is the high number of people who reported an increase in the number of comments that "are identical or nearly identical." Twenty reported an increase, with thirteen of these answering "many more." Only one respondent reported fewer such comments.

5. Comments that are identical or nearly identical?

Response	**No. of Respondents**	**Percentage of Total**
Many More	13	26.5%
More	7	14.3%
Slightly More	4	8.2%
The Same	14	28.6%

Slightly Fewer	0	0%
Fewer	1	2%
Many Fewer	0	0%
Don't Know	10	20.4%

Response Count: 49
Average Score: 2.59 (n=39)

Despite this tendency toward more opinionated and more similar comments, most rulemakers nonetheless reported that e-rulemaking has not caused them to place less "value on the comments by the average citizen." Three-fourths of the respondents answered "the same" for this question.

6. Value of the comments of average citizens?

Response	**No. of Respondents**	**Percentage of Total**
Much Less	1	2%
Less	3	6%
Slightly Less	3	6%
The Same	38	76%
Slightly Higher	0	0%
Higher	2	4%
Much Higher	3	6%

Response Count: 50
Average Score: 4.27 (n=50)

Does e-rulemaking perhaps lead to more commenters responding to others' comments or to economic analyses in the docket? One might hypothesize that this would be the case since such comments and analyses are easier to access online by potential commenters. There is at least some indication that this is occurring, especially with respect to reacting to others' comments.

7. In your experience, with the advent of e-rulemaking, have you seen more commenters responding to comments already in the docket?

Response	**No. of Respondents**
Yes	20
No	16
Don't Know	14

8. In your experience, with the advent of e-rulemaking, have commenters made more references to economic analyses and other supporting documents in the docket?

Response	**No. of Respondents**
Yes	7
No	22
Don't Know	20

E-rulemaking has also led to a slight increase in the number of questions agencies receive about ongoing rulemakings.

9. Number of questions to your office about ongoing rulemakings?

Response	**No. of Respondents**	**Percentage of Total**
Many Fewer	0	0%
Fewer	4	8%
Slightly Fewer	2	4%
The Same	22	44%
Slightly More	5	10%
More	5	10%
Many More	1	2%
Don't Know	11	22%

Response Count: 50
Average Score: 4.21 (n=39)

Most agency rulemakers reported some opportunity to consult with and learn from their counterparts in other agencies about e-rulemaking issues, but more than half reported that this opportunity was less than adequate.

10. As an agency rulemaker, how much opportunity have you had to consult with and learn from your counterparts in other agencies about e-rulemaking issues?

Response	**No. of Respondents**	**Percentage of Total**
None	8	13.5%
Minimal Opportunity	25	42.3%
Adequate Opportunity	18	30.5%
Great Opportunity	8	13.5%

A few narrative responses were received to this question. Several commenters lauded the efforts of some agencies (e.g., EPA, Treasury) to conduct workshops and share information, but others wished for more: "Simply not enough."; "I hope there will be many more opportunities in the future."; "I believe that in general we do a terrible job of facilitating the exchange of knowledge, best practices, and lessons learned."; and "It would be great to have a more advanced training on the use of e-rulemaking dealing less with the mechanics."

The survey also sought to elicit information about how agencies deal with the e-comments. One question simply asked if agency rulemakers tended to make hard (paper) copies of e-comments. Of the fifty-nine responses, only eight said "never." Most (twenty-four) said "occasionally," nine said "usually," and eighteen said "always."

This question stimulated a number of narrative responses. One (from the DOL) reported, "We still legally have to keep a paper copy of all comments, once a docket closes and we post electronic comments we print them all out." Another (no agency identified) explained, "When a rulemaking results in litigation, hard copies of the administrative record (including comments) need to be made for the parties and the court. Also, if a rule is complex, requests for hard copies from within the agency are inevitable." A third had another pragmatic explanation: "It is virtually impossible to review complex or lengthy electronic comments without making a hard copy." The same goes for sharing comments with colleagues for review and consideration, particularly if those colleagues do not have access to the e-comments. Several others said something to the effect of "I will make copies of significant comments that will be used to change analysis or be responded to in [the] preamble."

Another question asked whether the respondent's agency used "computer based 'sorting' technology to help categorize (or identify duplicate) e-comments." Of the fifty-nine respondents, twenty did not know. Of those who did know, seventeen said "yes"; twenty-two said "no." Of the twelve who expressed an opinion on this, three said this technology was "very helpful," six said "helpful," two said "a little helpful," and only one said "not at all." Two respondents (from DOL and DHS) reported that their agency had hired a contractor to do this.

As to the "bottom line" questions of whether e-rulemaking helped agencies promulgate rules more efficiently or promulgate higher quality rules, the responses were encouraging. Twenty-three of forty-four respondents reported an increase in efficiency as opposed to only eight who reported a decline.

11. In toto, and as a general matter, has the advent of e-rulemaking allowed your

agency to promulgate rules less or more efficiently?

Response	**No. of Respondents**	**Percentage of Total**
Much Less	1	2%
Less	6	12%
Slightly Less	1	2%
The Same	13	26%
Slightly More	7	14%
More	13	26%
Much More	3	6%
Don't Know/No Opinion	6	12%

Response Count: 50
Average Score: 4.61 (n=44)

The responses as to higher quality rules were also positive, though somewhat less so, with twelve of forty-four respondents reporting an increase in quality and five reporting a decrease. The main difference with the efficiency question is that twice as many respondents answered "the same" as to quality.

12. In toto, and as a general matter, in your opinion, has the advent of e-rulemaking made it less or more easy for your agency to promulgate higher quality rules?

Response	**No. of Respondents**	**Percentage of Total**
Much Less	0	0%
Less	2	4%
Slightly Less	3	6%
The Same	27	54%
Slightly More	3	6%
More	7	14%
Much More	2	4%
Don't Know/No Opinion	6	12%

Response Count: 50
Average Score: 4.36 (n=44)

One factor to keep in mind concerning agency staff attitudes toward e-

rulemaking is the increasing use of electronic dockets for other agency actions. Of the fifty-nine respondents, thirty-six reported that their agency uses e-dockets for actions other than rulemaking, and twenty-three said they did not. According to the narrative answers, agencies were using e-dockets for adjudication, guidances, notices, Paperwork Reduction Act notices, draft legislation, peer-reviewed matters, and certain correspondence.

One commenter waxed enthusiastic about e-dockets:

> E-dockets are fantastic. Currently the Federal Transit Administration (FTA) is using an e-docket to formulate an agency policy statement. Also, FTA uses e-dockets as forms of electronic filing systems . . . for various administrative adjudications, such as charter service adjudications. Complainants may file complaints electronically on an e-docket. Once a complainant files a complaint on an e-docket, the respondent may respond electronically via the e-docket. FTA posts its decisions on the e-docket. Ultimately, this process increases transparency in government, and we have not received as many FOIA requests for these documents because the documents are easily accessible.

The last question on the survey was open-ended and asked "for any other comments." Twenty-six respondents took the time to answer. The following are the most significant comments—and they tend to divide equally between favorable and unfavorable—though we should bear in mind that usually those with a grievance may be more likely to respond to such a question.

D. Rulemakers' General Comments on e-Rulemaking

1. Generally Positive Comments

(a) E-rulemaking is the obvious choice for encouraging public comment and allowing easy access to records from anywhere and without risking the loss of original hard copies. My only complaint is that the process is not completely electronic—we still generate many paper copies of each rule or proposal.

(b) Having an electronic docket has enabled me to manage comments to my rulemaking projects much more easily. I now can just tell people on my rulemaking project how to go to *regulations.gov* instead of having to make hard copies of the comments and distributing them to the team members on a regular basis.

(c) E-rulemaking has improved public access and internal efficiency, but we are not yet using all the potential tools that it makes available.

(d) It is a very powerful tool. We need to continue to inform the public on how best to use the tool. We need to continue to add the next phase to

the Federal Docket Management System (FDMS), namely more rulemaking development tools for the rule writer.

(e) With more people using the Internet, it seems the right way to conduct rulemaking and promises to reach more folks who don't read the *Federal Register*. In addition to reaching older members of society, making the process available online makes it more likely we will reach members of Generation X and the Millennium Generation. I was informed by an IT person in a [regulations] development workshop, however, that an online rulemaking docket did not constitute a blog because you have to open the NPRM [notice of proposed rulemaking] (or other documents published in the *Federal Register*) to get to the core subject. But you could set up a blog with a link to the docket, webcast live public meetings[,] and record them as podcast files for downloading from the docket.

(f) Interesting topic where many questions are yet unanswered. I think at this point the benefits to the agency are not fully evident since much time is spent on learning the new systems, but hopefully in the near future it will prove more efficient than the previous paper-based system.

(g) E-rulemaking is better at letting the public know what the agencies are doing than it is at providing thoughtful input into the decisions themselves.

(h) I support it . . . . In addition to making agency rulemaking more accessible to the public, it makes it easy for me to check DOL and other agency rulemakings and comments. It's a great research tool.

(i) Good start but they need to further refine the process for better functionality.

(j) Makes it much easier for the public to see the comments, less work for the agency to respond to requests for copies of comments. Less likelihood that important comments will go missing due to mistake or design.

(k) E-rulemaking hasn't changed the process of rulemaking. What it has done is provide easier access to already public documents easier. That is, interested parties can get documents at their desktop rather than having to go to a docket room.

2. Generally Negative Comments

(a) Many of the initial fears (e.g., authenticity of comments, transmissions of viruses, etc.) have not yet come to pass, but they are a constant concern.

(b) Because of intermittent FDMS and *regulations.gov* system outages, we continue to maintain an in-house paper-based parallel process for managing comments. Unless the reliability of e-rulemaking-related systems increase to the point where we are comfortable enough to move away from paper, we will not fully realize the potential efficiencies that can be gained

by moving to the electronic platform.

(c) We have been "live" with FDMS less than a year, and have had only 2 or 3 rules in the system. One rule had only one (supportive) comment, and another has had well over a thousand so far, but mostly an industry-generated paper letter-writing campaign from individuals whose names and addresses we must type and load into FDMS, a royal pain in the neck for our tiny staff.

(d) If you work at an Agency or Bureau that doesn't do many regulations, it's difficult to remember all the technology steps that are required to post a regulation. I find myself having to relearn the process each time. That is frustrating.

(e) It's difficult to isolate the effect of e-rulemaking on the rulewriting process because as more tools become available, the pressure grows to delay decisions and rulewriting until closer to the deadline.

(f) The system is very user friendly for public commenters and very user unfriendly for government regulators. Indeed, the system design seems to thwart at every stage the efficient assembly and review of public comments. It is difficult to access the comments, print them out, sort them by topic, match up attachments with cover documents, etc. Each comment has to be downloaded or printed separately before it can be skimmed for content. When there are thousands of comments, that takes an unreasonably long time. It was much faster to take a stack of hard copy comments and page through them to sort out the duplicates and hone in on the helpful, substantive letters. Plus, they could be easily sorted, flagged, and tabbed with notes and comments. In addition, it now takes much longer for comments to work their way from the technical folks that manage the e-rulemaking system to the regulatory folks that actually write the regulations (which could be many people on a complex regulation). I used to get the comments within a day or two of the close of the comment period; now it can take weeks.

(g) As my agency's FDMS Administrator, I have found FDMS/*regulations.gov* hard to use, confusing, and not intuitive at all. I also believe that what is now *regulations.gov* should be integrated into the *Federal Register* so that the *Federal Register*'s online version of a rulemaking document contains a hotlink directly to the *regulations.gov* docket and comment form for that rulemaking document.

(h) It was much easier under the former USDOT e-docket system than under the *regulations.gov* system. More features and ability to analyze comments better. We have had quite a few technical glitches that I guess, over time, will be ironed out. For example, I cannot directly upload documents to the docket in one of my rules.

(i) E-rulemaking, including drafting and review of rulemaking

documents[,] has resulted in reduction in the quality of the reviews and rise in inclination of reviewer to revise text to meet personal style. Overall, this affects the timing and quality of rules.

(j) I believe it is more costly to my agency because we have had to maintain two systems—our old electronic system and the FDMS.

(k) We view it as a benefit for the public, not necessarily as providing a great advantage for the agency.

CONCLUSION

It is fair to conclude, based on this relatively small sample, that agency rulemakers are generally receptive to e-rulemaking, although a common theme of their early evaluations was that the new system is a "boon for the public but a bane for the agency." Indeed, a large majority of respondents reported a general increase in rulemaking efficiency and a smaller majority reported a general increase in rulemaking quality. They said this even though they were also generally dubious about the usefulness of the resulting additional comments. In addition, a series of questions asked whether e-rulemaking has made it more or less easy to undertake some positive rulemaking activities, and in each case the answer was that it was easier.

On the other hand, another series of questions asked whether e-rulemaking has increased the level of concern about some of the worries hypothetically associated with e-rulemaking, and in the case of eight of them, the answer was that their worries had increased.

Thus, the early picture is still mixed—no one doubts that the new system is better at engendering more public participation, although most agency rulemakers did not report receiving a concomitant increase in useful information or arguments among the additional comments. Moreover, while rulemakers *are* quite impressed with the internal administrative and coordination benefits provided by the new technology, they also have heightened concerns about hacking and the potential problems of inappropriate worldwide exposure of certain information in their electronic dockets.

APPENDIX

Question 1: When compared to the old system of paper comments, has the advent of e-rulemaking made it more difficult or easier for your agency to do the following?

a. Conduct proactive notification and outreach to the public by maintaining target mailing lists (or listservs) of people who are interested in selected aspects of your rulemaking agendas?

Answer Number	Corresponding Written Answer	No. of Respondents	Percentage of Respondents
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	0	0%
3	A little more difficult under the new system	1	3%
4	The same under the new system	11	30%
5	A little easier under the new system	3	8%
6	Easier under the new system	12	32%
7	Much easier under the new system	10	27%
	N/A	27	

Response Count: 64
Average Score: 5.51 (n=37)

b. Identify and find appropriate stakeholders?

Answer Number	Corresponding Written Answer	No. of Respondents	Percentage of Respondents
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	1	2%
3	A little more difficult under the new system	3	7%
4	The same under the new system	16	38%
5	A little easier under the new system	9	21%
6	Easier under the new system	9	21%
7	Much easier under the new system	4	10%
	N/A	22	

Response Count: 64
Average Score: 4.81 (n=42)

c. Disseminate information relevant to the agency's proposed rulemaking (e.g., studies, economic analyses, legal analyses), so as to generate more informed commenters?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	0	0%
3	A little more difficult under the new system	2	4%
4	The same under the new system	10	22%
5	A little easier under the new system	5	11%
6	Easier under the new system	13	28%
7	Much easier under the new system	16	35%
	N/A	18	

Response Count: 64
Average Score: 5.67 (n=46)

d. Present to the public competing or multiple alternatives to the proposed rules?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	0	0%
3	A little more difficult under the new system	1	3%
4	The same under the new system	22	59%
5	A little easier under the new system	5	14%
6	Easier under the new system	3	8%
7	Much easier under the new system	6	16%
	N/A	26	

Response Count: 63
Average score: 4.73 (n=37)

e. Stimulate public comments generally?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	0	0%

2	More difficult under the new system	2	4%
3	A little more difficult under the new system	0	0%
4	The same under the new system	17	33%
5	A little easier under the new system	5	10%
6	Easier under the new system	14	27%
7	Much easier under the new system	13	25%
	N/A	13	

Response Count: 64
Average score: 5.33 (n=51)

f. Sort and analyze public comments generally?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	4	8%
2	More difficult under the new system	2	4%
3	A little more difficult under the new system	2	4%
4	The same under the new system	12	24%
5	A little easier under the new system	5	10%
6	Easier under the new system	13	25%
7	Much easier under the new system	13	25%
	N/A	13	

Response Count: 64
Average score: 5.02 (n=51)

g. Obtain public comments specifically addressed to particular portions or segments of the proposed rule?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	0	0%
3	A little more difficult under the new system	7	15%
4	The same under the new system	21	45%
5	A little easier under the new system	4	8%
6	Easier under the new system	12	26%

7	Much easier under the new system	3	6%
	N/A	17	

Response Count: 64
Average score: 4.64 (n=47)

h. Sort and analyze public comments specifically addressed to particular portions or segments of the proposed rule?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	4	9%
2	More difficult under the new system	1	2%
3	A little more difficult under the new system	4	9%
4	The same under the new system	15	33%
5	A little easier under the new system	4	9%
6	Easier under the new system	10	22%
7	Much easier under the new system	8	17%
	N/A	18	

Response Count: 64
Average score: 4.70 (n=46)

i. Use the concept of "reply comments"?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	1	4%
2	More difficult under the new system	1	4%
3	A little more difficult under the new system	1	4%
4	The same under the new system	3	12%
5	A little easier under the new system	3	12%
6	Easier under the new system	9	36%
7	Much easier under the new system	7	28%
	N/A	34	

Response Count:59
Average score: 5.44 (n=25)

j. Place summaries of ex parte communications in the record more quickly.

Answer Number	Corresponding Written Answer	No. of Respondents	Percentage of Respondents
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	1	2.6%
3	A little more difficult under the new system	1	2.6%
4	The same under the new system	16	42.1%
5	A little easier under the new system	5	13.1%
6	Easier under the new system	5	13.1%
7	Much easier under the new system	10	26.3%
	N/A	20	

Response Count: 58
Average score: 5.16 (n=38)

k. Coordinate the rulemaking internally by allowing many people to look at the same rulemaking docket without getting in each others' way?

Answer Number	Written Answer	No. of Respondents	Percentage of Respondents
1	Much more difficult under the new system	1	2%
2	More difficult under the new system	1	2%
3	A little more difficult under the new system	0	0%
4	The same under the new system	9	21%
5	A little easier under the new system	4	9%
6	Easier under the new system	10	23%
7	Much easier under the new system	18	42%
	N/A	16	

Response Count: 59
Average score: 5.70 (n=43)

l. Coordinate the rulemaking externally with OMB or other interested government entities?

Answer Number	Corresponding Written Answer	No. of Respondents	Percentage of Respondents
1	Much more difficult under the new	1	2.5%

	system		
2	More difficult under the new system	0	0%
3	A little more difficult under the new system	0	0%
4	The same under the new system	13	32.5%
5	A little easier under the new system	6	15%
6	Easier under the new system	14	35%
7	Much easier under the new system	6	15%
	N/A	22	

Response Count: 62
Average score: 5.23 (n=40)

m. Conduct interactive proceedings in rulemaking, such as "negotiated rulemaking"?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	2	12.5%
2	More difficult under the new system	0	0%
3	A little more difficult under the new system	1	6%
4	The same under the new system	7	44%
5	A little easier under the new system	2	12.5%
6	Easier under the new system	2	12.5%
7	Much easier under the new system	2	12.5%
	N/A	43	

Response Count: 59
Average score: 4.19 (n=16)

n. Craft a preamble to the final rule that responds to comments and includes all relevant studies and analyses?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	1	2.2%
3	A little more difficult under the new system	1	2.2%
4	The same under the new system	17	38.65%
5	A little easier under the new system	7	15.9%

6	Easier under the new system	12	27.3%
7	Much easier under the new system	6	13.6%
	N/A	16	

Response Count: 60
Average score: 5.05 (n=44)

o. Develop and implement appropriate archival practices relating to rulemakings (such as retiring records, etc.)?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	1	3%
2	More difficult under the new system	0	0%
3	A little more difficult under the new system	1	3%
4	The same under the new system	11	31%
5	A little easier under the new system	5	14%
6	Easier under the new system	10	28%
7	Much easier under the new system	8	22%
	N/A	22	

Response Count: 58
Average score: 5.25 (n=36)

p. Periodically evaluate and review the rule (and related rules), once promulgated?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Much more difficult under the new system	0	0%
2	More difficult under the new system	1	2.7%
3	A little more difficult under the new system	0	0%
4	The same under the new system	15	40.5%
5	A little easier under the new system	4	10.8%
6	Easier under the new system	9	24.3%
7	Much easier under the new system	8	21.6%
	N/A	25	

Response Count: 62
Average score: 5.19 (n=37)

Question 2. "When compared to the old system of paper comments, has the advent of e-rulemaking caused your agency to worry more or less about the following:"

a. Outside intervention ("hacking") into your rulemaking proceedings?

Answer Number	Corresponding Written Answer	No. of Respondents	Percentage of Respondents
1	Worry much more under the new system	3	8.1%
2	Worry more under the new system	9	24.3%
3	Worry a little more under the new system	12	32.4%
4	The same under the new system	7	18.9%
5	Worry a little less under the new system	1	2.7%
6	Worry less under the new system	2	5.4%
7	Worry much less under the new system	3	8.1%
	N/A	25	

Response Count: 62
Average score: 3.32 (n=37)

b. Acquiring viruses via attachments submitted in comments?

Answer Number	Corresponding Written Answer	No. of Respondents	Percentage of Respondents
1	Worry much more under the new system	5	14%
2	Worry more under the new system	8	22%
3	Worry a little more under the new system	11	31%
4	The same under the new system	7	19%
5	Worry a little less under the new system	1	3%
6	Worry less under the new system	4	11%
7	Worry much less under the new system	0	0%
	N/A	26	

Response Count: 62
Average score: 3.31 (n=36)

c. Inappropriate exposure of materials in the rulemaking docket that might contain confidential business information?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	5	11%
2	Worry more under the new system	11	24%
3	Worry a little more under the new system	14	31%
4	The same under the new system	10	22%
5	Worry a little less under the new system	1	2%
6	Worry less under the new system	2	4%
7	Worry much less under the new system	2	4%
	N/A	17	

Response Count: 62
Average score: 3.11 (n=45)

d. Inappropriate exposure of materials in the rulemaking docket that might contain copyrighted materials?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	4	9%
2	Worry more under the new system	9	20%
3	Worry a little more under the new system	19	41%
4	The same under the new system	9	20%
5	Worry a little less under the new system	1	2%
6	Worry less under the new system	1	2%
7	Worry much less under the new system	3	7%
	N/A	16	

Response Count: 62
Average score: 3.20 (n=46)

e. Inappropriate exposure of materials in the rulemaking docket that might contain indecent or obscene language or materials?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	3	7%
2	Worry more under the new system	8	18%
3	Worry a little more under the new system	16	36%
4	The same under the new system	13	30%
5	Worry a little less under the new system	0	0%
6	Worry less under the new system	2	5%
7	Worry much less under the new system	2	5%
	N/A	18	

Response Count: 62
Average score: 3.30 (n=44)

f. Inappropriate exposure of information in the rulemaking docket that might lead to national security problems?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	2	7%
2	Worry more under the new system	0	0%
3	Worry a little more under the new system	7	25%
4	The same under the new system	15	54%
5	Worry a little less under the new system	1	4%
6	Worry less under the new system	2	7%
7	Worry much less under the new system	1	4%
	N/A	34	

Response Count: 62
Average score: 3.82 (n=28)

g. Risk of information-destruction or other irretrievable loss of rulemaking information?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	2	5%
2	Worry more under the new system	4	9%
3	Worry a little more under the new system	8	19%
4	The same under the new system	17	40%
5	Worry a little less under the new system	5	12%
6	Worry less under the new system	0	0%
7	Worry much less under the new system	7	16%
	N/A	19	

Response Count: 62
Average Score: 4.09 (n=43)

h. Integrating (scanned) paper comments with e-mailed or electronically submitted comments?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	4	8.2%
2	Worry more under the new system	5	10.2%
3	Worry a little more under the new system	10	20.4%
4	The same under the new system	13	26.5%
5	Worry a little less under the new system	4	8.2%
6	Worry less under the new system	4	8.2%
7	Worry much less under the new system	9	18.4%
	N/A	13	

Response Count: 62
Average score: 4.14 (n=49)

i. The authenticity of comments?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	3	6%
2	Worry more under the new system	6	13%
3	Worry a little more under the new system	5	11%
4	The same under the new system	25	53%
5	Worry a little less under the new system	2	4%
6	Worry less under the new system	3	6%
7	Worry much less under the new system	3	6%
	N/A	15	

Response Count: 62
Average score: 3.81 (n=47)

j. Ensuring the protection of the privacy of commenters?

Answer Number	**Corresponding Written Answer**	**No. of Respondents**	**Percentage of Respondents**
1	Worry much more under the new system	7	15%
2	Worry more under the new system	6	13%
3	Worry a little more difficult under the new system	15	33%
4	The same under the new system	14	30%
5	Worry a little less under the new system	1	2%
6	Worry less under the new system	2	4%
7	Worry much less under the new system	1	2%
	N/A	16	

Response Count: 62
Average score: 3.13 (n=46)

COMMENTS

OVER THE COUNTER, UNDER THE RADAR: HOW THE ZICAM INCIDENT CAME ABOUT UNDER FDA'S HISTORIC HOMEOPATHIC EXCEPTION

AMY GAITHER*

TABLE OF CONTENTS

* J.D. Candidate, May 2011, American University Washington College of Law; B.A. English, Communications, 2007, Boston College. I would like to thank Professor Lewis Grossman, Professor Andrew Popper, Colleen O'Boyle, Anita Ghosh, and the editors and staff of the *Administrative Law Review* for their instrumental roles in the development of this Comment. I am also deeply indebted to Tim Speros and my family for their endless encouragement and support that made this Comment possible.

INTRODUCTION

On June 16, 2009, Matrixx Initiatives, Inc., maker of Zicam Cold Remedy products, received a devastating piece of correspondence. Arguably even more than the 300 lawsuits waged against Zicam products since 1999,[1] this letter had the potential to impact Matrixx's business like nothing else, short of a complete cure for the common cold. It was a warning letter from the Food and Drug Administration (FDA), alerting the manufacturer that due to over 130 reports of anosmia—loss of sense of smell, which in some cases can be long lasting or permanent—FDA concluded that Zicam intranasal products posed a serious risk to consumers.[2] The agency thus intended to regulate the intranasal products as "new drugs" under the applicable provision of the Federal Food, Drug, and Cosmetic Act (FDCA).[3] Marketed as homeopathic drugs, the products were never subject to FDA premarket approval, which requires prescription and over-the-counter (OTC) drugs not generally recognized as safe and effective to be thoroughly tested before entering the market.[4] Unlike Zicam

1. *See* Jennifer Corbett Dooren, *FDA Warns Against Use of Zicam*, WALL ST. J., June 17, 2009, at B1 (highlighting the safety issues that have plagued the company's intranasal products—Zicam Cold Remedy Nasal Gel, Zicam Cold Remedy Gel Swabs, and Zicam Cold Remedy Swabs Kids' Size—since their inception, and describing Matrixx's settlement of its numerous lawsuits in 2006); Valerie Jablow, *Lawsuits Sniff Out Zinc Hazard in Nasal Cold Remedy*, TRIAL, Feb. 2005, at 78 (listing some of the suits' claims which included fraud, negligence, strict products liability, breach of warranty, and breach of state consumer protection statutes).

2. *See* Letter from Deborah M. Autor, Dir., Office of Compliance, Ctr. for Drug Evaluation & Research, FDA, to William J. Hemelt, Acting President, CFO, and COO, Matrixx Initiatives, Inc. (Jun. 16, 2009), http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/ucm166909.htm [hereinafter Warning Letter] (noting that loss of sense of smell can have serious consequences, such as inability to detect the smell of a gas leak, smoke, or spoiled food; the letter also stated that some Zicam users also lost their sense of taste).

3. *Id.*; *see* Federal Food, Drug, and Cosmetic Act (FDCA) § 201(p), 21 U.S.C. § 321(p)(1) (2006) (defining *new drug* as any drug not generally recognized among qualified experts as safe and effective for its intended use).

4. *See* Warning Letter, *supra* note 2 (asserting that "[n]othing in the [FDCA] or the regulations issued under it exempts homeopathic drugs from new drug approval requirements," but recognizing that FDA has traditionally made the discretionary choice not to enforce the requirements with regard to homeopathic drugs); FDCA § 505(a), 21 U.S.C. § 355(a) (2006) (banning new drugs from introduction into interstate commerce without an

intranasal products, Zicam oral products have posed no safety threat, and thus remain on the market under FDA's historic homeopathic exception.[5] The warning letter prompted a voluntary recall projected to cost nearly $10 million, effectively eradicating the targeted products from the market unless and until Matrixx can prove them to be safe and effective for their intended uses under FDA's new drug application regime.[6] Not only did the market negatively respond to the warning letter—Matrixx stock plummeted 70% the day of the letter's release—but the media took issue with the situation as well, questioning the ability of a product to exist on drug store shelves with seemingly no FDA oversight.[7]

Consumers may be similarly troubled by the questions raised in the Zicam incident. Many consumer advocate websites attempt to warn the public that if an OTC product states "homeopathic" on the label, buyers may not be getting what they expect—a drug approved by FDA to be safe and effective for use as directed.[8] What may increase the severity of the Zicam situation is FDA's discovery of over 800 similar adverse event reports in Matrixx's possession that were never turned over to FDA.[9] For

approved new drug application); FDCA § 505(b)(1)(A), 21 U.S.C. § 355(b)(1)(A) (2006) (requiring all new drug applications to contain—and thereby conditioning their approval on—reports of investigations showing that the drug is safe and effective for its intended use).

5. *See generally* FDA, COMPLIANCE POLICY GUIDES § 400.400, CONDITIONS UNDER WHICH HOMEOPATHIC DRUGS MAY BE MARKETED (1988, revised 1995), *available at* http://www.fda.gov/ICECI/ComplianceManuals/CompliancePolicyGuidanceManual/ucm074360.htm [hereinafter COMPLIANCE POLICY GUIDE] (setting forth the only comprehensive set of regulatory guidelines for the marketing of homeopathic drugs, including conditions under which FDA will discretionarily allow marketing of homeopathic drugs without approved new drug applications).

6. *See* Jonathan D. Rockoff, *Matrixx Receives SEC Inquiry Following Warning About Zicam*, WALL ST. J., June 24, 2009, at B3 (detailing the reactive measures Zicam was forced to take in the wake of the warning letter's publication, which included a pledge to reimburse consumers for prior purchases of Zicam intranasal products).

7. *See id.* (correlating the stock price drop with the large percentage of Matrixx's business that was represented by its Zicam intranasal products); Transcript for FDA Media Briefing on FDA's Advice to Consumers Not to Use Certain Zicam Cold Remedies, June 16, 2009, http://www.fda.gov/downloads/NewsEvents/Newsroom/MediaTranscripts/UCM168484.pdf (featuring questions from reporters of major news outlets to FDA representatives, some questions particularly focusing on the confusing regulatory posture of homeopathic drugs like the Zicam products).

8. *See, e.g.*, ConsumerReportsHealth.org, Homeopathic Drugs: Look-Alike Medicines, http://consumerreports.org/health/natural-health/homeopathic-drugs/overview/homeopathic-drugs-ov.htm (last visited Apr. 11, 2010) (citing the experience of eleven "mystery shoppers" who visited fifty-two drug stores across the United States and found homeopathic and approved OTC products directly next to each other on store shelves, purportedly demonstrating that it was conceivable for consumers to unwittingly buy a homeopathic product without understanding the significant differences between it and the neighboring approved product).

9. *See* Warning Letter, *supra* note 2 (acknowledging the existence of the 800 reports

the average consumer, a heightened concern regarding FDA's methods is understandable; yet the Zicam incident appears to be the first of its kind. The warning letters issued to homeopathic marketers in the past involved regulatory infractions, not serious adverse events.[10] In fact, this is precisely the reason cited by FDA for its discretionary lack of oversight—with a view to the agency's limited resources, homeopathic products have simply never aroused significant cause for concern—that is, until now.[11]

This Comment will examine the foundations of the current homeopathic drug regulatory framework, evaluate the strengths and weaknesses of FDA's seemingly hands-off approach, and provide an analysis of how FDA can preserve the system's strengths while incorporating more oversight into its homeopathic product regime. Although this Comment provides background on both prescription and OTC homeopathic drugs, the regulatory analysis and recommendations pertain strictly to the OTC class. Part I provides a background of homeopathy and its historic treatment by both Congress and FDA. Part II examines FDA's current system of homeopathic product regulation and its application to the Zicam incident in order to extrapolate the powers invoked by FDA and the implications of those powers on the homeopathic drug industry. Finally, Part III provides recommendations for a future approach to homeopathic drug regulation with a focus on how FDA can incorporate aspects of analogous regimes into its current system to better effectuate its purpose of protecting the public health. The goal of these recommendations is to find a balance between two somewhat competing goals: judicious allocation of limited FDA resources and protection of public consumers in their reasonable expectations of product safety.

---

related to anosmia and requiring Matrixx to promptly submit the reports to FDA); *see also* Dooren, *supra* note 1 (explaining that although OTC product manufacturers were not required to report adverse events to FDA until recently, the legislation that implemented the requirement has been in effect since 2007). The FDA is still investigating these reports.

10. *See* Isadora Stehlin, *Homeopathy: Real Medicine or Empty Promises?*, FDA CONSUMER, Dec. 1996, at 15, 18 (finding that the most common infraction was the sale of prescription homeopathic drugs over the counter). Other warning letters cited products being "promoted as homeopathic that contain nonhomeopathic active ingredients"; "lack of tamper-resistant packaging"; "lack of proper labeling"; and "vague indications for use that could encompass serious disease conditions," which would require prescription dispensing and labeling. *Id.* For a more detailed discussion of the Compliance Policy Guide from which these violations stem, see *infra* Part II.

11. *See* Suzanne White Junod, *An Alternative Perspective: Homeopathic Drugs, Royal Copeland, and Federal Drug Regulation*, 55 FOOD & DRUG L.J. 161, 178–79 (2000) (recounting the reasoning behind the exclusion of homeopathic treatments from the OTC Drug Review, which included perceptions that such treatments were harmless and that homeopathy was a dying specialty).

I. BACKGROUND

The cloud of ambiguity surrounding homeopathic drugs is best explained through the FDA regulatory framework, whose treatment of homeopathic products stems from a controversial history dating back to the enactment of the FDCA in 1938. Concomitantly, the best approach for FDA to address situations like the Zicam incident in the future must be charted within this framework in consideration of the policy issues that have shaped the current state of homeopathic drug regulation. This section provides a brief history of homeopathy and the historic development of homeopathic drug regulation in the United States.

A. What Is Homeopathy?

The National Center for Complementary and Alternative Medicine (NCCAM) within the National Institutes of Health (NIH) designates homeopathy as a "whole medical system," or a complete system of theory and practice that evolved separately from "conventional medicine."[12] Homeopathy was developed by Samuel Hahnemann, a German physician practicing in the late 1700s, a time when bloodletting was the most common medical practice in Europe and the United States.[13] Hahnemann's aversion to the harsh and ineffective treatments of his day led him to seek out new forms of therapy, through which he developed the first, and main, tenet of homeopathy: like cures like, or the law of similars.[14] This premise holds that if a substance causes certain symptoms in a healthy person, the substance can treat those symptoms when exhibited by a person

12. NAT'L CTR. FOR COMPLEMENTARY & ALTERNATIVE MED., U.S. DEP'T OF HEALTH AND HUMAN SERVS., CAM BASICS 2 (2007), http://nccam.nih.gov/health/whatiscam/D347.pdf.

13. LYN W. FREEMAN & G. FRANK LAWLIS, MOSBY'S COMPLEMENTARY & ALTERNATIVE MEDICINE 347 (John Schrefer ed., 2001); Stehlin, *supra* note 10, at 16. Other popular treatments included blistering, which involved placing scalding substances on the skin to "draw out" infection, and administration of large doses of toxic substances, such as opiates, chloroform, and calomel (mercury chloride) to relieve pain and induce purging. *Id.*; NATALIE ROBINS, COPELAND'S CURE 6 (2005).

14. ROBINS, *supra* note 13, at 6. This founding homeopathic theory shaped the practice's name: "homeopathy" was derived from the Greek terms *homoios* (like) and *pathos* (suffering). Stehlin, *supra* note 10, at 16. Fittingly, Hahnemann referred to conventional medicine as "allopathy," from the Greek term *allos* (other). ROBINS, *supra* note 13, at 6. The term has stuck, and many sources still refer to conventional medicine as allopathy. The National Council Against Health Fraud insists that this term has been misapplied since the time of Hahnemann and asserts that modern medical writers who refer to conventional doctors as "allopaths" do so with an intended alternate meaning, one that refers to a practice utilizing only those remedies "proved of value." NAT'L COUNCIL AGAINST HEALTH FRAUD, NCAHF POSITION PAPER ON HOMEOPATHY (1994), http://ncahf.org/pp/homeop.html.

who is ill.[15] Hahnemann developed the theory when he experimentally administered himself a strong dose of quinine and found that it caused him to develop symptoms similar to those caused by malaria.[16] He tested his theory on himself and others in a practice he called "provings"—if a substance brought about certain symptoms, it would be used by homeopaths to treat those symptoms.[17] Hahnemann began to decrease the dosage of his test substances (which were debilitating in high quantities) and thereby developed the second main tenet of homeopathy: the minimum dose, or the law of infinitesimals.[18] This premise called for diluting homeopathic preparations to an extreme degree and subjecting them to forceful shakings between successive dilutions.[19] This practice has been unsparingly criticized. First, if any active ingredient remains in a preparation, it is, as the homeopathic principal describes, infinitesimal. For example, the amount of original substance in a 30X product has been diluted 1,000,000,000,000,000,000,000,000,000,000 times, which is roughly equivalent to one drop in a container more than fifty times the size

15. Stehlin, *supra* note 10, at 16. Some sources suggest that the idea of like cures like goes back to the writings of Hippocrates. *E.g.*, FREEMAN & LAWLIS, *supra* note 13, at 347; *see also* PHILLIP A. NICHOLLS, HOMEOPATHY AND THE MEDICAL PROFESSION 16–17 (1988) (placing the writings between 430 and 330 B.C.). However, Phillip A. Nicholls asserts that those writings were likely not authored entirely by Hippocrates, resulting in the use of both similars *and* opposites in the Hippocratic texts—use of medicines that were thought to bring about and medicines that were thought to suppress the symptoms expressed by the patient. *Id.*

16. Stehlin, *supra* note 10, at 16. Quinine had been used for centuries to treat malaria and fever but why it helped was not known. ROBINS, *supra* note 13, at 7. This invited Hahnemann to apply his theory to the substance's curative value, which he did for the smallpox vaccine as well. As an injection of cowpox, a form of the same illness the vaccine inoculated against, Hahnemann praised the vaccine as a prime example of the law of similars at work. ROBINS, *supra* note 13, at 6–7.

17. ROBINS, *supra* note 13, at 6. Homeopathy purported to treat symptoms, whereas allopathy purported to alleviate symptoms by treating the disease. *See* NICHOLLS, *supra* note 15, at 33 (analyzing the dual therapeutic methods from a socioeconomic standpoint, revealing that the orthodox approach stems from, among other things, an effort at streamlined disease-based diagnosing to treat more patients). Hahnemann believed that most recurring symptoms stemmed from a common disease, referred to as the "itch" or "psora." ROBINS, *supra* note 13, at 10.

18. Stehlin, *supra* note 10, at 16.

19. ROBINS, *supra* note 13, at 8–9. In this process, which Hahnemann called "potentization," one drop of substance is placed into a 1:10, 1:100, or 1:1000 ratio of water or alcohol, designated with Roman numerals as 1X, 1C, and 1M respectively. After the shaking or forceful hitting of the substance's container, called "succussion," one drop of the first dilution is then placed into a new 1:10, 1:100, or 1:1000 water or alcohol ratio, followed again by succussion. The process can be done once or repeated many times; the number of a substance's successive dilutions is indicated by the number in front of the Roman numeral X, C, or M. *See generally* FREEMAN & LAWLIS, *supra* note 13, at 350 (detailing the homeopathic dilution process).

of earth.[20] Second, at such high dilutions, there may not be any active ingredient in some preparations at all. Critics often cite Avogadro's number—which theorizes a point in the process of dilution where a molecule of any given substance can no longer exist—as evidence of the ineffectiveness of homeopathic products.[21]

Hahnemann developed the final tenet of homeopathy, the doctrine of individualized therapy, as a means of employing the first two: in practicing his new kind of medicine, he insisted that homeopaths conduct lengthy patient evaluations, often up to one or two hours, in order to ascertain all emotional and physical symptoms for precise treatment.[22] Through this practice and the absence of side effects resulting from diluted medications, homeopathy quickly gained publicity for employing a gentler approach than traditional medicine.[23] Conventional doctors continually outnumbered homeopathic practitioners; however, homeopathy remained popular with the public, resulting in over one hundred homeopathic medical schools in major cities across the country by the 1880s.[24] Around this time, the homeopathic community experienced a general, though not complete, shift away from strict adherence to certain classical homeopathic

---

20. *See* Dan McGraw, *Flu Symptoms? Try Duck*, U.S. NEWS & WORLD REP., Feb. 17, 1997, at 51 (examining the popular homeopathic product oscillococcinum 200C, which uses the heart and liver of a single duck to create enough product to generate sales of over $20 million).

21. ROBINS, *supra* note 13, at 10. Several other unorthodox medical systems began to emerge in the United States around this time which must be distinguished from homeopathy. These include osteopathy, which holds that illness results from the failure of the body to have proper bone and muscle alignment; Christian Science, which believes that God alone promotes healing; chiropractic, which bases its healing on manipulation of the spine; and naturopathy, which purports to treat disease with natural elements, such as hot and cold air baths, massage, and diet. *Id.* at 24, 50.

22. W. STEVEN PRAY, A HISTORY OF NONPRESCRIPTION PRODUCT REGULATION 191 (2003); *see also id.* at 192 (pointing out that OTC homeopathic products, which obviously do not require individualized evaluation before purchase, violate this doctrine and thus "should not be considered homeopathic at all").

23. *See* ROBINS, *supra* note 13, at 5–6 (recounting that homeopathy was rumored to have greatly aided in the cholera epidemics of 1832 and 1849, which it likely did by replacing the harmful conventional treatment options of bloodletting and purging).

24. *See* Martin Kaufman, *Homeopathy in America: The Rise and Fall and Persistence of a Medical Heresy*, *in* OTHER HEALERS: UNORTHODOX MEDICINE IN AMERICA 99, 105 (Norman Gevitz ed., 1988) (noting that homeopathic and allopathic medical educations were very similar and that for the most part homoeopathic principals were taught in addition to, not in the place of, traditional medical training). Homeopathy's popularity led the American Medical Association to believe that people were being duped by homeopathic practitioner "gimmickry." ROBINS, *supra* note 13, at 19. This sentiment continues among critics today. *See, e.g.*, Leon Jaroff, *The Man Who Loves to Bust Quacks*, TIME, Apr. 30, 2001, at 61 (profiling Stephen Barrett, a former psychiatrist and well-known health fraud monitor who has dedicated thirty years to educating consumers on how not to be duped by "quacks" and their sales tactics, mainly on his widely read website entitled "Quackwatch").

laws, such as the practice of prescribing a single preparation for all of a patient's symptoms and the oxymoronic principle of the greater the dilution the more potent the preparation.[25]

One of the early death knells of homeopathic prominence in the United States was the widely publicized Flexner Report which surveyed the quality of medical education in the United States.[26] Its depictions of most homeopathic medical schools as subpar training facilities that were unscientific, even unsanitary, resulted in the majority closing or converting to allopathic medicine by the 1920s.[27] Another blow to homeopathy was dealt by the scientific advancements that abounded during the early- to mid-1900s, including the development of antibiotics as well as clinical studies utilizing placebos as controls to prove the effectiveness of medicines.[28] This progress greatly enhanced public desire for scientific medicine, leading to a significant decline in the practice and teaching of homeopathy in the late 1930s and early 1940s.[29]

Yet homeopathy has not gone away. The paradigm-shattering consciousness of the 1960s and 1970s saw a large increase in demand for all things unconventional, including medicine.[30] Today, homeopathy and many other forms of treatment are grouped together in the general category of complementary and alternative medicine (CAM), use of which continues to expand.[31] A recent survey by NIH and the Centers for

25. *See* Kaufman, *supra* note 24, at 106–07 (noting that homeopathic practitioners who streamlined their practices to combine homeopathy and allopathy were able to see more patients and thus increase their income, something the remaining "pure" Hahnemannian homeopaths reviled).

26. *Id.* at 111.

27. *Id.* at 112; ROBINS, *supra* note 13, at 111, 117.

28. ROBINS, *supra* note 13, at 123, 225.

29. *Id.* at 226. Dr. Morris Fishbein, editor of the *Journal of the American Medical Association* who would later become the Association's president, proclaimed "The Death of Homeopathy" as early as 1932, faulting Hahnemann's "unprovable theory." MORRIS FISHBEIN, FADS AND QUACKERY IN HEALING 27–29 (1932). Surmising this triumph, he wrote, "Thus passed the homeopathic system. Thus, in fact, pass all systems in the practice of medicine. Scientific medicine absorbs from them that which is good, if there is any good, and then they die." *Id.* at 28–29.

30. *See generally* Anne Taylor Kirschmann, *Making Friends for "Pure" Homeopathy: Hahnemannians and the Twentieth-Century Preservation and Transformation of Homeopathy*, *in* THE POLITICS OF HEALING: HISTORIES OF ALTERNATIVE MEDICINE IN TWENTIETH-CENTURY NORTH AMERICA 29 (Robert D. Johnston ed., 2004) (elaborating on the philosophical underpinnings of homeopathy which spurred the practice's reemergence when the 1960s counterculture identified with "pure" homeopathy's focus on individualized care and rejection of the mainstream medical establishment). *But see* ROBINS, *supra* note 13, at 241 (describing the time as one that brought homeopathy back into relevance, yet acknowledging that continual scientific advances would never allow homeopathy to escape the shadow of dominant conventional medicine).

31. *See* NAT'L CTR. FOR COMPLEMENTARY & ALTERNATIVE MED., *supra* note 12, at 1

Disease Control and Prevention found that Americans spend $34 billion a year on alternative therapies, a growth of more than 25% in the past decade.[32] The survey shows that $2.9 billion alone goes toward homeopathic products,[33] a number that gives new significance to the FDA warning letter to Matrixx by highlighting the obvious implications such enforcement actions pose to the homeopathic drug market.

B. Historic Congressional Treatment of Homeopathy

The Federal Food, Drug, and Cosmetic Act of 1938 was the landmark legislation that established FDA's power of premarket review for all new drugs. It also represents the point at which homeopathy first made an appearance in the *United States Code*. One provision of the FDCA definition of the term *drug* includes "articles recognized in the official United States Pharmacopeia, official Homeopathic Pharmacopeia of the United States, or official National Formulary, or any supplement to any of them."[34] FDA has generally not been able to regulate products as drugs based solely on their inclusion in one of the listed compendia, as the text appears to allow.[35]

(defining *complementary medicine* as medicine used together with conventional medicine, and *alternative medicine* as medicine used in place of conventional medicine). NCCAM lists the following examples as falling under the CAM umbrella: acupuncture, aromatherapy, ayurveda, chiropractic, dietary supplements, electromagnetic fields, homeopathic medicine, massage, naturopathic medicine, osteopathic medicine, qi gong, reiki, therapeutic touch, and traditional Chinese medicine. *Id.* at 3–4.

32. Liz Szabo, *More Trying Alternative Therapies*, USA TODAY, July 31, 2009, at 3A. A 1998 study from the Stanford Center for Research in Disease Prevention found that "the majority of alternative medicine users appear to be doing so not so much as a result of being dissatisfied with conventional medicine but largely because they find these health care alternatives to be more congruent with their own values, beliefs, and philosophical orientations toward health and life." John A. Astin, *Why Patients Use Alternative Medicine: Results of a National Study*, 279 JAMA 1548, 1548 (1998); *see also Alternative Medicines: Hearing Before the Subcomm. on Labor, Health and Human Servs., and Educ., and Related Agencies of the S. Comm. on Appropriations*, 106th Cong. 1, 2 (2000) (opening statement of Sen. Arlen Specter, Chairman, Subcomm. on Labor, Health and Human Servs., and Educ., and Related Agencies of the S. Comm. on Appropriations) (citing statistics that showed 42% of United States health care consumers utilized CAM treatments and discussing the founding of NCCAM within NIH to promote studies on CAM treatments).

33. Szabo, *supra* note 32.

34. FDCA § 201(g)(1)(A), 21 U.S.C. § 321(g)(1)(A) (2006).

35. *See* Nat'l Nutritional Foods Ass'n v. FDA, 504 F.2d 761, 788–89 (2d Cir. 1974) (dismissing the argument that recognition in an official compendia is sufficient to establish that a product as falls under the "drug" definition because such a premise would lead to the conclusion that all vitamins and minerals are drugs because they are listed in the official compendia, which runs counter to FDA's own regulations); *see also* Nat'l Nutritional Foods Ass'n v. Mathews, 557 F.2d 325, 337–38 (2d Cir. 1977) (rejecting a similar FDA argument as arbitrary, finding that FDA's position would result in conflicting treatment of only certain vitamins as drugs despite other vitamin listings in the official compendia).

However, this provision has been the source of much confusion as to why the 1938 Act bothered to recognize the Homeopathic Pharmacopeia of the United States (HPUS) in the first place.[36] The common answer is that the senator who sponsored the FDCA, Royal Copeland, was a homeopathic physician, and it is undeniable that Copeland favored inclusion of the HPUS in the FDCA because of his affiliation with the practice.[37] However, FDA historian Susan White Junod asserts that food and drug officials likely welcomed this proposal, not as a concession to Copeland but as part of a strategy to utilize the FDCA to prosecute fraudulent drug salesmen that peddled bogus homeopathic products.[38]

Another provision of the *United States Code* that references the HPUS is the FDCA section that defines *official compendium*.[39] The import of this provision lays in those sections of the FDCA that state the conditions under which a drug is rendered adulterated or misbranded—the FDCA utilizes the official compendia for public standards of strength, quality, and

36. *See* Junod, *supra* note 11, at 161 (pointing out the incongruity of the FDCA representing the modernization of drug regulation yet simultaneously recognizing the seemingly unscientific and waning practice of homeopathy). According to the Homeopathic Pharmacopeia Convention of the United States (HPCUS), the entity now responsible for publication of the HPUS, the HPUS was first published in 1897 and continues to be a source of homeopathic drug information, including drug monographs, general labeling, and manufacturing information. HPCUS, What is the HPUS?, http://www.hpus.com/whatishpus.php (last visited Apr. 11, 2010); HPCUS, Overview, http://www.hpus.com/overview.php (follow "The HPUS Revision Service Contents" hyperlink) (last visited Apr. 11, 2010). The official version of the HPUS is now referred to as the HPUS Revision Service, which is available only in web format through an online subscription via www.hpus.com.

37. *See* ROBINS, *supra* note 13, at 207 (describing the inclusion of the HPUS in the FDCA as a part of Copeland's lifelong effort to enhance homeopathy's legitimacy in society). Senator Copeland, who had served as dean of the New York Homeopathic Medical College and as the New York Health Commissioner, rejected much of the mysticism that accompanied early homeopathic doctrine and was at the fore of the homeopathic modernization movement. *See id.* at 100, 148–50, 166–67 (detailing the influential positions held by Copeland including his fortuitous election as New York State Senator in 1922, which was largely a product of his prominence in the health field); Junod, *supra* note 11, at 167 (describing Copeland's explanation of homeopathy as a "complementary medical discipline" to be used in conjunction with advancing science, and his downplaying of the law of infinitesimals).

38. *See* Junod, *supra* note 11, at 173–74 (contending that this strategy is one the legislators would have purposely withheld from the legislative record, and pointing out that, indeed, the record gives no reasoning behind inclusion of the HPUS in the 1938 Act where the 1906 Pure Food and Drugs Act already recognized the United States Pharmacopoeia and the National Formulary as official compendia).

39. FDCA § 201(j), 21 U.S.C. § 321(j) (2006). The other official compendia, the United States Pharmacopeia (USP) and the National Formulary (NF), are now published together in a single volume referred to as the USP–NF.

purity.[40] The HPUS was likely included in the "official compendium" definition as a result of its inclusion in the "drug" definition; however, the HPUS has historically been relied on to a very limited extent in the context of both FDCA provisions: FDA officials have always had many other tools under the Act by which to halt sales of bogus products claiming to be based on homeopathic theory, and legitimate homeopathic drugs historically have not posed safety threats, obviating use of the HPUS language in the Act.[41] The *United States Code*'s formal recognition of homeopathy has therefore had little impact—likely the reason the HPUS remains in the text today.[42]

C. Historic FDA Treatment of Homeopathic Drugs

FDA's interpretation and application of legislative amendments to the FDCA—as opposed to the statutory language itself—has been the main source of regulation historically governing homoeopathic products. The first wave of regulation that significantly affected the homeopathic community was the 1951 Durham–Humphrey Amendment to the FDCA.[43] The amendment states that if use of a drug is unsafe unless

40. *See* PETER BARTON HUTT, RICHARD A. MERRILL & LEWIS A. GROSSMAN, FOOD AND DRUG LAW 534 (3d ed. 2007) (explaining that if a drug fails to comply with the official compendium standard, the manufacturer may still use the common drug name listed in the compendium and state how the drug differs from the standard in strength, quality, or purity). If a drug is included in both the USP–NF and the HPUS, the USP–NF standards apply unless the product is clearly labeled and sold as a homeopathic product. FDCA §§ 501(b), 502(e), 502(g), 21 U.S.C. §§ 351(b), 352(e), 352(g). Although still an official compendium whose standards homeopathic drugs must generally conform to, the HPUS has played a more limited role in FDA adulteration and misbranding regulation than the USP–NF because of the limited nature of homeopathic drug regulation generally. *Cf.* Edward M. Cohen, *The Influence of the USP on the Drug Approval Process*, *in* THE PHARMACEUTICAL REGULATORY PROCESS 335, 335–39 (Ira R. Berry & Robert P. Martin eds., 2d ed. 2008) (detailing the development of the USP and its extensive historical involvement in FDA statutory enforcement schemes, which expanded with its acquisition of the NF in 1974).

41. *See* Junod, *supra* note 11, at 175–76 (citing the FDA's new authority under the 1938 Act to conduct factory inspections as a more straightforward approach to apprehending fraudulent product manufacturers); ROBINS, *supra* note 13, at 7–8 (revealing that many substances used in homeopathic products were used by ancient civilizations, and, in any event, homeopathy advocates believed that even potentially toxic substances were seldom dangerous at such high dilutions).

42. *See* Junod, *supra* note 11, at 179 (recounting that the proposed 1979 Drug Reform Act would have eliminated the HPUS provisions; however because the bill was defeated and no House hearings were held, the reasoning behind the proposal remains unclear and the HPUS remains in the statute).

43. *Id.* at 176; *see also* HUTT, MERRILL & GROSSMAN, *supra* note 40, at 488–90 (discussing the text of the amendment that codified FDA regulations distinguishing between prescription and nonprescription drugs). FDA's prescription requirement was based on the premise that adequate directions for use could not be formulated for certain drugs, which

supervised by a licensed practitioner because of its toxicity, other potentiality for harm, its method of use, or the collateral measures necessary to its use, the drug must be dispensed by prescription only.[44] After the law's passage, the leading homeopathic medical association, the American Institute for Homeopathy (AIH), advocated for most homeopathic drugs to be dispensed by prescription to keep in line with the traditional homeopathic practice of individualized treatment.[45] FDA approved, reservedly, stating that although the amendment did not appear to bring homeopathic drugs under its regime, FDA had no objection to their distribution with the prescription legend.[46] FDA was careful to note, however, that it would not bring enforcement actions against homeopathic

thus required dispensing and supervision by a medical practitioner. *See* Peter Temin, *The Origin of Compulsory Drug Prescriptions*, 22 J.L. & ECON. 91, 99 (1979) (noting that FDA never stated its reasoning behind this presumption, which inevitably had a great impact on the drug market).

44. FDCA § 503(b)(1), 21 U.S.C. § 353(b)(1) (2006). The original 1951 provision included "habit-forming" in the definition, which was deleted in 1997. HUTT, MERRILL & GROSSMAN, *supra* note 40, at 489. What constitutes a "licensed practitioner" is determined by each state as an inherent police power. MICHAEL H. COHEN, COMPLEMENTARY & ALTERNATIVE MEDICINE: LEGAL BOUNDARIES AND REGULATORY PERSPECTIVES 24 (1998). This policy was determined by the Supreme Court in *Dent v. West Virginia*, 129 U.S. 114 (1889), which held that the power of the states to provide for the general welfare authorizes state regulation of the medical profession to protect citizens from "ignorance and incapacity . . . deception and fraud." *Id.* at 122. *See generally* Michael H. Cohen, *A Fixed Star in Health Care Reform: The Emerging Paradigm of Holistic Healing*, 27 ARIZ. ST. L.J. 79 (1995) [hereinafter Cohen, *A Fixed Star*] (discussing the advent of medical licensing statutes in the United States and their function in entrenching traditional medical practices while excluding complementary and alternative medicine). Three states—Connecticut, Arizona, and Nevada—have established licensing boards for homeopathic practitioners; some states that do not have separate licensing boards nevertheless include homeopathy within the definition of complementary and alternative medicine (CAM) providers, who are subject to a licensing process, and other states limit the practice of homeopathy to licensed chiropractors only. *See* Patrick L. Sheldon, *The Truth About Homeopathy: A Discussion of the Practice and the Dangers That Inhere*, 8 QUINNIPIAC HEALTH L.J. 289, 295–97 (2005) (describing the range of restrictiveness in regulatory licensing regimes employed by the states, into which the professional practice of homeopathy falls differently from state to state).

45. *See* Junod, *supra* note 11, at 176–77 (describing the American Institute for Homeopathy (AIH) lobbying of FDA officials that resulted in most homeopathic drugs becoming prescription drugs in the 1950s). This was perhaps an effort on the part of AIH to establish integrity and legitimacy for a practice that, at the time, was a faint voice in the medical community.

46. *See id.* (asserting that FDA accepted the AIH's argument that adequate directions for use of homeopathic drugs could not be devised, in accordance with the theory behind the FDCA's prescription drug provision and the concomitant FDCA requirement that all OTC products contain adequate directions for use); *cf. id.* at 168 (recalling that prior to passage of the FDCA and its "adequate directions for use" provisions, a distinct benefit of homeopathic physicians was that they usually dispensed their own medicines, a practice that appealed to consumers who claimed pharmacists failed to provide satisfactory label directions).

products that were sold for minor conditions without a prescription, i.e., OTC homeopathic products, and it appears the homeopathic drug prescription labeling requirement was historically not strictly enforced.[47]

A far more serious alteration of the FDCA occurred with the Kefauver–Harris Amendments of 1962, which threatened to wipe homeopathic drugs from the market with a new requirement for proof of drug efficacy for intended uses.[48] Under the 1938 Act, drug makers were required to submit new drug applications (NDAs) to FDA demonstrating a drug's safety.[49] An NDA went into effect within approximately two weeks unless FDA took affirmative action against it, allowing the NDA drug as well as subsequent chemically similar products—or "me-too" drugs—to be marketed.[50] Under the 1962 Amendments, however, not only was affirmative approval required before a new drug could be marketed, but the Secretary was also required to reject an NDA or rescind a previously approved NDA if there was lack of substantial evidence of the drug's effectiveness.[51] Under these Amendments, FDA commenced the Drug Efficacy Study Implementation (DESI) by which it undertook to ascertain the efficacy of all drugs that were covered by former safety NDAs.[52] More importantly for the homeopathic community, FDA also undertook the large task of the Over-the-Counter Drug Review. Unlike the DESI Review that examined each individual 1938–1962 NDA-covered drug, and did so solely to determine efficacy for intended uses, the OTC Drug Review was conducted on a therapeutic category basis by which FDA advisory panels examined information on active ingredients, not individual OTC products, and did so to determine

47. *See id.* at 177 (revealing that this lack of enforcement was reportedly due to some FDA officials believing that the prescription legend lent "undeserved credibility" to homeopathic drugs).

48. *See generally* PRAY, *supra* note 22, at 147–70 (detailing the legislative history of the 1962 Act and how the thalidomide tragedy, where a sleeping pill approved in Europe caused birth defects in thousands of children, contributed to the passage of the premarket approval regime and its retroactive efficacy requirements).

49. 21 U.S.C. § 355(a)–(b) (1940).

50. *Id.* at § 355(c); *see also* HUTT, MERRILL & GROSSMAN, *supra* note 40, at 579–80 (explaining that the original NDA drug was referred to as the "pioneer" and that all subsequent copy drugs had to correspond with an approved pioneer NDA, which was then said to "cover" all the "me-too" drugs as well).

51. *See* Drug Amendments of 1962, Pub. L. No. 87-781, § 102, 76 Stat. 780, 781 (1962) (inserting "and effectiveness" in various provisions within FDCA § 505 and setting forth the § 505(d) criteria by which an NDA is to be approved and the § 505(e) criteria by which an NDA is to be withdrawn—both provisions contain the language "lack of substantial evidence that the drug will have the effect it purports or is represented to have").

52. *See* HUTT, MERRILL & GROSSMAN, *supra* note 40, at 580 (highlighting that industry members protested the idea of FDA examining 1938–1962 NDAs for efficacy and demanded an outside authority conduct examinations, resulting in FDA's contract with the National Academy of Sciences to conduct testing for the DESI Review).

both safety and efficacy for intended uses.[53] Drugs that did not meet FDA safety and efficacy standards were deemed "unapproved new drugs" and subjected to the NDA provisions of the FDCA; drugs that met the standard could remain on the market on the condition that they complied with published regulatory monographs specific to each active ingredient on issues such as dosing, labeling, warnings, and other important issues.[54]

By the time the OTC Drug Review commenced in 1972, approximately one-third of all homeopathic products sold were OTC.[55] However, with a view to limited FDA resources and the mounting work to be done on nonhomeopathic OTC products, FDA decided not to subject homeopathic drugs to review for safety and efficacy.[56] There appear to have been several unstated reasons for FDA's decision. First, homeopathic products generally did not pose serious safety threats.[57] Second, FDA was focusing its regulatory efforts elsewhere—not only on the OTC Drug Review, but on a large-scale attempt to regulate vitamins and minerals—and may have considered it impracticable to review homeopathic drugs at the time, or at least of little benefit in relation to the effort that would have been required.[58] And finally, FDA may have realized that homeopathic drug efficacy testing would be problematic from the outset. Commentators have

---

53. *See generally* Over-the-Counter Drugs: Proposal Establishing Rule Making Procedures for Classification, 37 Fed. Reg. 85, 86–89 (Jan. 5, 1972) (describing FDA's reasoning for the OTC Drug Review); Procedures for Classification of Over-the-Counter Drugs, 37 Fed. Reg. 9464, 9464–73 (May 11, 1972) (the final rule, as amended, is now codified at 21 C.F.R. § 330.10) (addressing comments regarding the unique process by which OTC drugs would be reviewed).

54. *See* Over-the-Counter Drugs: Proposal Establishing Rule Making Procedures for Classification, 37 Fed. Reg. at 85 (introducing the monograph enforcement approach).

55. Junod *supra* note 11, at 178. AIH's attempt to relegate homeopathic drugs to the prescription domain had apparently failed.

56. *See* Over-the-Counter Drugs: Proposal Establishing Rule Making Procedures for Classification, 37 Fed. Reg. at 85–86 (noting the indomitable size of the OTC Drug Review, which encompassed between 100,000 and 500,000 OTC products containing an estimated 200 active ingredients). In the lengthy Procedures for Classification of Over-the-Counter Drugs, FDA dedicated a short and decisive paragraph to the issue:

> The American Institute of Homeopathy requested that homeopathic medicines be excluded from the OTC review. Because of the uniqueness of homeopathic medicine, the Commissioner has decided to exclude homeopathic drugs from this OTC drug review and to review them as a separate category at a later time after the present OTC drug review is complete.

37 Fed. Reg. at 9466. In the thirty-seven intervening years since this regulation, the plan to review homeopathic drugs has never materialized.

57. *See* Junod, *supra* note 11, at 177–79 (representing that FDA made a judgment call that, at the time, homeopathic products were not used nearly to the extent that other OTC products were used, and thus they were less of a concern for FDA in light of a history of relative safety).

58. *See id.* at 178–79 (asserting that FDA's preoccupation with the vitamin regulation, which eventually failed, made it easier for FDA to ignore homeopathic drugs).

suggested that due to the need to rely on the HPUS as an official compendium of homeopathic drug standards, any form of a homeopathic drug review may require drug testing by homeopathic experts using provings as efficacy tests.[59] FDA might have considered it imprudent to expend the resources necessary for another review that would have so vastly differed in approach from the scientific foundations of the OTC Drug Review.[60] In 1972, these very compelling reasons resulted in FDA leaving homeopathy well enough alone; that is, until 1988 when it issued the Compliance Policy Guide that governs sales of homeopathic drugs today.

II. CURRENT HOMEOPATHIC DRUG REGULATION AND THE ZICAM INCIDENT

The driving force behind the warning letter to Matrixx is FDA's power under the FDCA to regulate a product as a drug if it is "intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease" or, for any product that is not food, if it is "intended to affect the structure or any function of the body."[61] The intended use of Zicam intranasal products, as

59. *See, e.g.*, *id.* at 177 (reasoning that scientific experts would apply standards incompatible with the official homeopathic compendium, the HPUS, which would be contrary to the language of the FDCA); Rebecca Gelfond, Regulating Homeopathic Drugs: Pragmatic Solutions for the Food and Drug Administration 44 (Feb. 8, 1999) (unpublished manuscript, on file with author), *cited in* HUTT, MERRILL & GROSSMAN, *supra* note 40, at 619 (proposing that FDA require homeopathic drugs to be effective where effectiveness is viewed from the standpoint of homeopathy).

60. *See* 21 C.F.R. § 330.10(a)(4)(ii) (2009) (setting forth the Review's efficacy standard as a reasonable expectation that the pharmacological effect of the drug will provide clinically significant relief in a significant proportion of the target population). HPUS monographs were developed by homeopaths through traditional homeopathic provings, which themselves followed traditional homeopathic practices of dilution and succussion. These processes, and the homeopathic theories on which they rely, do not lend themselves to scientific evaluation, which is exactly what the AIH maintained in its discussions with FDA. *See* Kaufman, *supra* note 24, at 118 (depicting the homeopathic position as hostile to the OTC Drug Review process on the basis that homeopathic drugs could not be evaluated by "allopathic review"). Further, even in the unlikely event that FDA was comfortable relying on homeopathic provings as an acceptable form of efficacy testing, it is doubtful such tests would satisfy the Review's evidentiary requirement: the premise of like cures like may not scientifically comport with the intended meaning behind "clinically significant relief," 21 C.F.R. § 330.10(a)(4) (2006), and this ambiguous standard means it would be difficult for FDA to logically affirm that there was a "reasonable expectation" of achieving it. *Id.*

61. FDCA § 201(g)(1)(B)–(C), 21 U.S.C. § 321(g)(1)(B)–(C) (2006). The FDA has exercised this power to regulate products that were not generally considered drugs but that fell under the FDCA definition when manufacturers made drug-like claims demonstrating an intent that the products be used in the way the Act contemplates. *See, e.g.*, Letter from W. Charles Becoat, Dir., FDA Minneapolis District, to Ken Powell, Chairman and CEO, General Mills, Inc. (May 5, 2009), http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/ucm162943.htm (alerting the prominent food manufacturer that FDA would regulate Cheerios as a new drug

stated on the product labeling, was reduction of the duration of the common cold and the severity of cold symptoms, including sore throat, stuffy nose, sneezing, coughing, and congestion.[62] This means that the products were intended for use in the mitigation and treatment of disease, as well as intended to affect the function of the body, and thus were drugs under the FDCA.[63] Further, the warning letter asserts that the products are not generally recognized as safe and effective for their intended uses, which makes them not only drugs but also new drugs, for which approved NDAs are required.[64] Zicam intranasal products, like all homeopathic OTC products, always fell into this definition as a formal matter because they were never subjected to definitive safety and efficacy testing. Yet homeopathic products have been allowed to be marketed despite this fact due to, first, their exemption from the OTC Drug Review and, second, FDA's issuance in 1988 of a Compliance Policy Guide (CPG) that specifically allows such marketing. This CPG serves three main functions: (1) it establishes the conditions under which homeopathic drugs may be marketed, which do not include premarket NDA approval or compliance with an OTC drug monograph; (2) it warns the industry of how and when sale of a purportedly homeopathic drug will constitute health fraud; and (3) it sets forth specific regulations applicable to homeopathic products, including labeling requirements within the CPG as well as other requirements mandated by the FDCA and the *Code of Federal Regulations* (CFR).[65] This section will examine the CPG and analyze how these governing regulations applied to, and were invoked against, Zicam intranasal cold remedy products.

A. Conditions Under Which Homeopathic OTC Drugs May Be Marketed

Under the CPG, to qualify as a homeopathic drug a product must (1) be labeled as homeopathic; (2) contain an active ingredient listed in the HPUS; (3) be in a potency specified in terms of dilution, e.g. 1X, 2C, etc.; (4) contain diluents commonly used in homeopathic pharmaceuticals; and

based on labeling on the box and website indicating an intended use for the prevention, mitigation, and treatment of hypercholesterolemia—excessive cholesterol—and the correlative coronary heart disease).

62. Warning Letter, *supra* note 2.

63. FDCA § 201(g)(1)(B)–(C), 21 U.S.C. § 321(g)(1)(B)–(C) (2006). Zicam intranasal products are not drugs based solely on their active ingredient's inclusion in the HPUS. *See supra* note 35 (describing why the official compendia provision of the drug definition cannot be taken literally).

64. Warning Letter, *supra* note 2; FDCA § 201(p)(1), 21 U.S.C. § 321(p)(1) (2006).

65. COMPLIANCE POLICY GUIDE, *supra* note 5, at 105–06.

(5) not be combined with any nonhomeopathic active ingredients.[66] Matrixx's Zicam intranasal products strictly complied with these CPG requirements: the packaging stated "Homeopathic";[67] the drug's only active ingredient, zinc gluconate, is included in the HPUS;[68] the ingredient was displayed on the labeling as "zincum gluconium 2X";[69] and the substance was diluted with water and other inactive ingredients providing the gel-like texture.[70] However, as the warning letter makes clear, compliance with the CPG does not preclude FDA action.

The CPG further mandates that products offered for the treatment of serious disease conditions be dispensed under the care of a licensed practitioner.[71] The key to the intended meaning of *serious disease condition* within the CPG is what the CPG allows to be sold *without* a prescription—products offered for use in "self-limiting conditions" that are recognizable by consumers and amendable to self-diagnosis.[72] Therefore, any drug that

66. *Id.* Diluents are generally water, alcohol, or both combined. HOMEOPATHIC PHARMACOPEIA CONVENTION OF THE U.S., THE HOMEOPATHIC PHARMACOPOEIA OF THE UNITED STATES REVISION SERVICE (online subscription, Dec. 2004), http://www.hpus.com [hereinafter HPUS REVISION SERVICE]. The HPUS lists several types of alcohols that may be used, including ethyl alcohol (found in alcoholic beverages), sucrose, lactose, and glycerin (which are all sugar alcohols). *Id.* Orally ingested homeopathic drugs are exempted from regulations mandating that OTC drugs intended for oral ingestion not contain alcohol as an inactive ingredient above certain prescribed concentrations. 21 C.F.R. § 328.10(a), (g)(3) (2009).

67. *See* MedShopExpress, Zicam No-Drip Liquid Nasal Gel, Cold Remedy Swabs, http://www.medshopexpress.com/602163.html (last visited Apr. 11, 2010) (providing a product description which includes detailed labeling text and a picture of the product packaging).

68. *Id.*; *see* HPUS REVISION SERVICE, *supra* note 66 (featuring the zinc gluconate monograph).

69. MedShopExpress, *supra* note 67. The designation 2X indicates that the product was diluted 1:10 twice, leaving 1 part per hundred (1:100 or 1 percent) of zinc gluconate in the final product. For an explanation of the homeopathic dilution process and the roman numeral designations, see *supra* note 19.

70. MedShopExpress, *supra* note 67.

71. COMPLIANCE POLICY GUIDE, *supra* note 5, at 105. For the meaning of *licensed practitioner*, see *supra* note 44.

72. COMPLIANCE POLICY GUIDE, *supra* note 5, at 105. The term *self-limited* is generally defined outside of the CPG as "limited by . . . its own nature" and "running a definite and limited course." WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 2060 (1986). This definition squarely applies to cold symptoms, minor aches and pains, sleeplessness, and other similar conditions for which OTC products are generally sold. Conversely, this definition would not apply to infectious diseases, such as strep throat or chlamydia, or life-threatening conditions such as cancer or AIDS. The CPG states that the prescription drug provision of FDCA is also applicable; therefore, any homeopathic drug that is unsafe if used without medical supervision due to its toxicity or other potential danger, including lack of adequate directions for use, is forbidden from OTC sale. Another provision of the CPG mandates that if the HPUS specifies certain distinctions between nonprescription and prescription status based on strength of the product, then the stricter criteria—either the prescription

is intended to treat a non-self-limiting condition must be dispensed by prescription. The Zicam products, again, strictly complied with these requirements, claiming only to reduce the duration and symptoms associated with the common cold—a classic example of a self-limiting condition.[73]

Notably, the CPG does not generally require NDA approval or compliance with an OTC drug monograph. The CPG is explicit in stating that all products falling within the drug definition are drugs under the FDCA regardless of whether they are homeopathic,[74] yet there is no mention of a homeopathic drug ever constituting a new drug under the FDCA, and there is no requirement for proof of safety or efficacy.[75] In the warning letter to Matrixx, however, FDA cites to a clause in the CPG that preserves its power of enforcement of the FDCA new drug provisions.[76] The clause asserts that the CPG delineates "those conditions under which homeopathic drugs may *ordinarily* be marketed in the [United States]."[77] Thus, as FDA interprets this clause in the warning letter, a homeopathic drug "is not subject to the enforcement discretion set forth in the CPG when there is evidence of a safety risk associated with the product."[78]

In order to effectively invoke the new drug provisions of the FDCA in the warning letter, FDA needed to show that the active ingredient in Zicam intranasal products was not generally recognized as safe and effective for its intended use.[79] The active ingredient in question—zinc gluconate—is a common ingredient in vitamins and dietary supplements.[80] To effectively

---

drug provision of the FDCA or the listing in the HPUS—will apply to determine prescription status. COMPLIANCE POLICY GUIDE, *supra* note 5, at 108.

73. *Cf.* Anne Gadomski, *A Cure for the Common Cold?: Zinc Again*, 279 JAMA 1999, 2000 (1998) (advising medical practitioners to give patients balanced information on the effectiveness of zinc in treating the common cold because no alternative therapies have the weight of evidence behind them and the benign nature of the cold, which inevitably ceases on its own, does not necessitate drug treatment).

74. This position was confirmed in *United States v. Writers & Research, Inc.*, in which the court held that a product labeled as homeopathic that was being promoted as a treatment for cancer, AIDS, and other chronic and degenerative diseases was subject to the requirements of the FDCA regardless of its homeopathic status. 113 F.3d 8, 11 (2d Cir. 1997).

75. COMPLIANCE POLICY GUIDE, *supra* note 5, at 107.

76. Warning Letter, *supra* note 2.

77. COMPLIANCE POLICY GUIDE, *supra* note 5, at 106 (emphasis added).

78. Warning Letter, *supra* note 2. The warning letter also deemed Zicam misbranded because its labeling did not bear adequate warnings of the risk of anosmia. *Id.*

79. *See* FDCA § 201(p), 21 U.S.C. § 321(p) (2006) (defining *new drug* as a drug not generally recognized as safe and effective for its intended use); FDCA § 505(a), 21 U.S.C. § 355(a) (prohibiting the introduction of new drugs into interstate commerce without an approved NDA).

80. *See* Gadomski, *supra* note 73, at 1999 (comparing the use of zinc in attempts to

eliminate Matrixx's ability to claim that zinc gluconate is generally recognized as safe and effective based on such common oral ingestion, the warning letter points to the existence of published literature that salts of zinc can damage olfactory function.[81] Throughout the FDCA, the terms *safety* and *efficacy* are couched with the phrase *for intended use*—this makes safety for oral ingestion irrelevant if the same ingredient is unsafe for a different use, in this case intranasal application.[82]

B. When Sale of a Homeopathic Drug May Constitute Health Fraud

The CPG only briefly mentions health fraud, yet it is an important inclusion in the face of the many complaints waged against the homeopathic regulatory regime.[83] The CPG defines *health fraud* as the "deceptive promotion, advertisement, distribution or sale" of drugs that are

---

"cure the common cold" with vitamins C and A, noting that all have "immune-enhancing properties"). Pursuant to the Dietary Supplement Health and Education Act of 1994, dietary ingredients in dietary supplements are regulated as a special subset of foods under the FDCA which are not subject to premarket approval. *See* HUTT, MERRILL & GROSSMAN, *supra* note 40, at 261–62 (describing the wide latitude dietary supplements achieved in the legislation that was enacted over FDA's strong objections); FDCA § 201(ff)(1)–(2), 21 U.S.C. § 321(ff)(1)–(2) (2006) (defining *dietary supplement* to include vitamins, minerals, herbs or other botanicals, amino acids, and any other dietary substance used to supplement the diet that is not represented as a conventional food). Zicam zinc-based products for oral ingestion have remained on the market as homeopathic products; however, their counterparts in the market—other orally ingested OTC cold drugs containing zinc—mostly fall under the dietary supplement regime. *See, e.g.*, Airborne, What's in Airborne, http://airbornehealth.com/about/whats-in-airborne (last visited Apr. 11, 2010) (listing the ingredients, including zinc, of the popular dietary supplement marketed to boost immunity to prevent or treat the common cold). The Zicam intranasal products could never qualify as dietary supplements due to the statutory definition of *dietary supplement* that specifies "ingestion" as the method of administration. FDCA § 201(ff)(2)(A)(i), 21 U.S.C. § 321(ff)(2)(A)(i) (2006); *see* United States v. Ten Cartons Ener-B Nasal Gel, 888 F. Supp. 381, 395 (E.D.N.Y. 1995) (holding that a Vitamin B-12 gel designed to be applied to the inside of the nose did not qualify for regulation as a dietary supplement because it was not "intended for ingestion," which implies administration through swallowing and entrance into the gastrointestinal tract).

81. Warning Letter, *supra* note 2.

82. The OTC monograph listing active ingredients found to be safe and effective for the uses indicated in Zicam product labeling—"Cold, Cough, Allergy, Bronchodilator, and Antiasthmatic Drug Products for Over-the-Counter Human Use"—does not include zinc. 21 C.F.R. § 341 (2009).

83. *See, e.g.*, ConsumerReportsHealth.org, *supra* note 8 (suggesting that consumer confusion between homeopathic products and more stringently regulated OTC products is a result of misleading marketing that is characteristic of health fraud). The Federal Trade Commission, not FDA, is responsible for regulating false advertising of OTC products, as distinguished from the fraudulent products themselves. HUTT, MERRILL & GROSSMAN, *supra* note 40, at 809–10. But the agencies often work in concert on such issues and have memoranda of understanding promoting collaboration. *Id.* at 810.

"represented as being effective" but which have not been scientifically proven as such.[84] Critics would be quick to point out the irony of such a provision appearing in a homeopathic drug guidance when homeopathic products have never been, and may never be, proven effective by FDA standards.[85]

The health fraud provision explicitly deems products that follow the "customary practice of homeopathy" as not constituting health fraud.[86] This could mean one of two things. On one hand, FDA may be indicating a willingness to consider homeopathic provings to be scientific evidence of safety and efficacy. On the other hand, the provision may be a simple recognition of the cognitive dissonance FDA must employ to uphold the current state of its regulation, by which homeopathic drugs are freely sold without any form of review or approval. This latter interpretation is far more probable—the CPG's inclusion of the health fraud provision was most likely intended to reinforce the CPG's policy of voluntary discretion not to enforce the FDCA, which FDA is free to disregard in cases of patently fraudulent behavior by drug manufacturers.

C. Labeling Requirements and Other Regulations Applicable to Homeopathic OTC Drugs

The CPG's general labeling requirements for homeopathic drugs reference provisions governing drugs in general in the FDCA and FDA-promulgated regulations. For example, the requirement for a "Name and Place of Business" on homeopathic product labeling must be carried out "in conformance with" the specific requirements for all other "name and place of business" drug labeling found in the FDCA and the CFR.[87] These provisions of the CPG are the reasons for the similarity between

84. COMPLIANCE POLICY GUIDE, *supra* note 5, at 106.

85. *See infra* part III.B (discussing the fact that homeopathic drugs would be eliminated from the market if a homeopathic drug efficacy review were instated). The CPG contains definitions of *homeopathy* and *homeopathic drugs*, and it cites two books that are offered as guides to the use of homeopathic drugs, including "potencies, dosing, and other parameters." COMPLIANCE POLICY GUIDE, *supra* note 5, at 106, 107. These resources are intended to give FDA officials a general understanding of how to differentiate between a genuine and a bogus homeopathic drug, though many would argue they are one and the same.

86. COMPLIANCE POLICY GUIDE, *supra* note 5, at 106.

87. *See id.* at 107 ("Each product must bear the name and place of business of the manufacturer, packer, or distributor in conformance with Section 502(b) of the [FDCA] and 21 CFR 201.1."). The CPG has similar labeling requirements for "Directions for Use," "Statement of Ingredients," "Established Name," and "Container Size" that apply to all homeopathic drugs, prescription and OTC. OTC homeopathic products in particular have other referential requirements for "Principal Display Panel," "Statement of Identity," "Indications for Use," and specific warning requirements for different indications that conform to OTC drug regulations. *Id.* at 107–08.

homeopathic and conventional OTC drug packaging; they are presumably a reflection of what FDA has determined to be necessary OTC labeling. The CPG also lists additional requirements specific to homeopathic products that are not found in the FDCA or the CFR: (1) the quantity and amount of ingredients in the product must be expressed in homeopathic terms, e.g., 1X, 2C, etc.; (2) if products or ingredients are not recognized by the HPUS, documentation must be provided to FDA to support that the products or ingredients are "generally recognized as homeopathic"; and (3) all labeling must be in English but may include the Latin name of the active ingredient in addition to the English name.[88]

These requirements have been the topic of much debate. The difference between conventional and homeopathic OTC drug labeling is essentially the word *homeopathic* and the listing of the active ingredient(s) in terms of dilution, as opposed to traditional units of measurement. On one hand, the CPG labeling requirements are beneficial in ensuring that the labels of homeopathic products include basic information that FDA considers necessary on all OTC drugs. On the other hand, such labeling may serve to confuse consumers as to what kind of product they are purchasing. If consumers do not notice the word *homeopathic* on the product label, or if they are unaware of what it means, they are likely to assume that the OTC drug has been tested for safety and efficacy by FDA (and that *efficacy* has its

88. *Id.* at 107. The requirement for supporting documentation for non-HPUS products presumably addresses the issue of manufacturers attempting to evade regulation by labeling a nonhomeopathic product as homeopathic. This has been the subject of warning letters to homeopathic drug manufacturers where FDA was quick to go after obvious attempts to game the system. *See, e.g.*, Letter from Steven A. Masiello, Dir., Office of Compliance & Biologics Quality, FDA Center for Biologics Evaluation and Research, to Bill Gray, M.D., Bill Gray Med. Corp. (Apr. 2, 2003), http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2003/ucm147405.htm (warning the manufacturer that the product, Dr. Gray's Smallpox Shield, would be regulated as a biologic for containing variolinum which is not listed in the HPUS, as well as for stating that the product required a prescription via a fee-based prescribing service on its website). With regard to the English-labeling requirement, the CPG notes that many homeopathic products bear Latin names that correspond with listings in the HPUS, which mandates Latin names as the primary titles on its monographs. COMPLIANCE POLICY GUIDE, *supra* note 5, at 107; HPUS REVISION SERVICE, *supra* note 66. The CPG goes on to list certain general requirements that homeopathic products are subject to, which are helpful in ensuring that FDA has adequate information and resources to go after a homeopathic drug manufacturer when necessary. These provisions state that all firms that manufacture or otherwise process homeopathic drugs must register as drug establishments and all homeopathic drug products must be listed, consistent with such requirements for drug manufacturer registration and listing in the FDCA and CFR. Similar requirements apply for packaging regulations and current good manufacturing practice regulations. Conversely, the CPG specifically exempts homeopathic products from expiration date regulations. COMPLIANCE POLICY GUIDE, *supra* note 5, at 109.

traditional meaning, not a meaning defined by the law of similars).[89]

Further, consumers who are unfamiliar with the meanings behind homeopathic terms of dilution may not know how much (or how little) of an active ingredient is in a product.[90] Moreover, even many consumers familiar with homeopathy are not aware that many homeopathic OTC drugs do not dilute substances to the extent that traditional homeopathic theory suggests.[91] Thus if labeling fails to clearly reveal the amount of an active ingredient, even those consumers who know that homeopathic drugs are supposed to be highly diluted may be deceived. Zicam is a prime example of this phenomenon: consumers who were familiar with homeopathic theory may not have realized that they were purchasing a product that contained a full 1% of active ingredient—an amount that is surely not infinitesimal.[92]

III. FUTURE HOMEOPATHIC DRUG REGULATION

Consumer-protection advocates have continually demanded that homeopathic drugs be required to demonstrate safety and efficacy as all other OTC drugs must.[93] Many advocate for this proposal primarily because homeopathic products would likely be eradicated from the market for failure to meet scientific standards.[94] The Zicam incident thus appears

89. *See* ConsumerReportsHealth.org, *supra* note 8 (explaining how consumers might unwittingly buy a homeopathic product without understanding the significant differences between it and an FDA-approved product).

90. *See* NAT'L COUNCIL AGAINST HEALTH FRAUD, *supra* note 14 (complaining that homeopathic labeling only informs consumers about the number of serial dilutions the product has undergone, which does not comport with standard drug labeling that informs consumers of the quantity of active ingredients per dose in known forms of volume measurement, such as milligrams). For the FDA regulations governing nonhomeopathic OTC drug content labeling, see 21 C.F.R. § 201.62 (2009).

91. *See* NAT'L COUNCIL AGAINST HEALTH FRAUD, *supra* note 14 (noting that many homeopathic dosages, although dilute, may contain enough of a substance to affect the body, and asserting that this fact is troubling from a consumer standpoint in light of the lack of proof of safety or effectiveness of homeopathic drugs).

92. *Compare* MedShopExpress, Afrin No Drip Nasal Decongestant Mist, 12 Hour, Original, http://www.medshopexpress.com/586545.html (last visited Apr. 11, 2010) (showing that the active ingredient oxymetazoline hydrochloride in a nonhomeopathic OTC, and thus CFR-compliant, drug is present in a concentration of 0.05% per dose), *with* MedShopExpress, *supra* note 67 (showing that the active ingredient in Zicam nasal swabs was present at a dilution level of 2X, indicating a concentration of 1% per dose).

93. *See, e.g.*, PRAY, *supra* note 22, at 203 (describing a 1994 petition to FDA filed by forty-two health professionals "asking the agency to develop rulemaking procedures to require that all homeopathic drugs be proven safe and effective").

94. *See, e.g.*, NAT'L COUNCIL AGAINST HEALTH FRAUD, *supra* note 14 (recommending that FDA require homeopathic products to meet the efficacy standards of all other drugs); Stephen Barrett, *Homeopathy: The Ultimate Fake*, QUACKWATCH, http://www.quackwatch.org/01QuackeryRelatedTopics/homeo.html (referencing the

to center on FDA's choice of either (1) initiating some form of review to proactively screen homeopathic drugs or (2) continuing with the current regulatory regime, which runs the risk of appearing to disregard FDA's traditional mandate of protecting the public health.[95] Of course protection of the public health is the primary goal of FDA; however, there may be an intermediate path by which to accomplish it in the context of homeopathic drugs.

A. Concerns

First, FDA is an entity of limited resources. Appropriations to the agency have gradually increased over the years,[96] yet many would argue that they have not done so proportionally to the increase in FDA responsibilities. These observations accumulate weight in light of the recent developments at the agency that have accompanied the new Administration. As compared to the noticeable decrease in FDA enforcement actions during the Bush Administration,[97] the Obama Administration FDA has increased its enforcement initiatives.[98] With the recent legislation establishing FDA authority over tobacco products, the agency has also acquired a whole new industry to regulate.[99] These

---

author's Freedom of Information Act request that revealed FDA has not found any homeopathic product to be safe and effective, and asserting that if such requirements were finally employed, homeopathic drugs "would face extinction in the United States"—an outcome for which the author strongly advocates).

95. *See* FDCA § 1003(b), 21 U.S.C. § 393(b) (2006 & Supp. 2009) (redesignated from FDCA § 903(b) by Pub. L. No. 111-31, § 101(b)(1), (2), 123 Stat. 1784 (2009)) (setting forth the mission of FDA to protect and promote the public health by effectively monitoring the products it regulates); *cf.* Kevin Gauntt Barker, Comment, *Thank You for Regulating: Why Philip Morris's Embrace of FDA Regulation Helps the Company but Harms the Agency*, 61 ADMIN. L. REV. 197, 221–22 (2009) (maintaining that FDA regulation of tobacco would force the agency to violate its own founding ideology by overseeing a marketed product that is harmful to health and has no offsetting benefit).

96. *See* FDA, Appropriations History Tables, http://www.fda.gov/AboutFDA/ReportsManualsForms/Reports/BudgetReports/2005FDABudgetSummary/ucm112957.htm (last visited Apr. 11, 2010) (charting appropriations for the years 1995–2005, which show a consistent increase each year).

97. *See* SPECIAL INVESTIGATIONS DIV., MINORITY STAFF OF H. COMM. ON GOV'T REFORM, 109TH CONG., PRESCRIPTION FOR HARM: THE DECLINE IN FDA ENFORCEMENT ACTIVITY 1 (2006) (summarizing the report that found a "precipitous drop" in FDA enforcement activities from 2000 to 2005, including instances where FDA headquarters rejected enforcement recommendations of FDA field offices despite serious violations and safety threats found by its officers).

98. *See* Jennifer Corbett Dooren, *Cheerios' Health Claims Break Rules, FDA Says*, WALL ST. J., May 13, 2009, at B1 (noting that the FDA is showing signs of taking a more aggressive stance toward regulated entities under its new leadership, such as the warning letter to General Mills that drew widespread attention).

99. *See generally* Family Smoking Prevention and Tobacco Control Act, Pub. L. No.

developments have resulted in increased appropriations at the request of the President.[100] Yet it remains to be seen how FDA will begin to handle its expanded responsibilities, and any proposal for future homeopathic drug regulation must be viewed within this strained context.

Second, the issue of consumer choice could create significant obstacles to any attempt to increase regulation of homeopathic drugs. FDA battled with the public and inevitably with Congress in its attempts to regulate vitamins and minerals, and subsequently dietary supplements.[101] These products, and all of CAM medicine for that matter, have been shown to be very important to American consumers; the public demands choice.[102] Many argue that the continual consumption of homeopathic drugs by the public is a result of the placebo effect, which could be considered both positive and negative.[103] Yet a significant portion of the population feels that homeopathic drugs truly work, or at least provide refreshing and

111–31, 123 Stat. 1776 (2009) (granting FDA authority to regulate tobacco products "to protect the public health").

100. *See* Press Release, Senate Comm. on Appropriations, Senate Approves FY 2010 Agriculture, Rural Development and FDA Appropriations (Aug. 4, 2009), http://appropriations.senate.gov/news.cfm?method=news.view&id=fc50e8f8-5a35-4349-876a-ab83515a7de9 (announcing that FDA will receive $299 million above the amount allocated to the agency for fiscal year 2009).

101. *See generally* PRAY, *supra* note 22, at 205–17 (providing a history of dietary supplement regulation, including the extensive back-and-forth struggle between FDA and Congress resulting in three separate statutory amendments, in which the entities had starkly opposing views as to how restrictive an approach the agency should take with regard to dietary supplements).

102. *See, e.g.*, Draft Guidance for Industry on Complementary and Alternative Medicine Products and Their Regulation by the Food and Drug Administration, 72 Fed. Reg. 29,337, 29,338 (May 25, 2007) (stating that FDA had received such a large volume of comments that it was unable to identify and respond to extension requests, and clarifying that the outcry was a result of "misinterpretation"; to calm it, FDA made clear that it did not propose any new regulatory requirements for CAM products, licensing of CAM practitioners, or consumer ability to purchase CAM products or be treated by a CAM practitioner)*; see also* Kathleen M. Boozang, *National Policy on CAM: The White House Commission Report*, 31 J.L. MED. & ETHICS 251, 251 (2003) (examining the White House Commission on Complementary and Alternative Medicine policy, which was established by President Clinton in 2000 to maximize the benefits of CAM in America as a response to consumers "voting with their health care dollars").

103. For an explanation of the placebo effect—the psychological phenomenon of patients purportedly recovering from various conditions when they think they are taking a new drug, but are actually receiving a placebo—see Tamar Nordenberg, *The Healing Power of Placebos*, FDA CONSUMER, Jan.–Feb. 2000, at 14, 14–17. For a critical view that analyzes the beneficial claims made by placebo proponents, see Harriet Hall, *The Placebo Effect*, 15 SKEPTIC 56 (2009), and Stephen Barrett, *Spontaneous Remission and the Placebo Effect*, QUACKWATCH, http://www.quackwatch.org/04ConsumerEducation/placebo.html. For information on the 1955 groundbreaking medical study on the placebo effect, see Henry K. Beecher, *The Powerful Placebo*, 159 JAMA 1602 (1955).

hopeful alternatives to conventional medicines.[104] Congress has done what it can to enable consumer choice in this forum,[105] thus any attempt by FDA to limit it could be seen as a contravention of congressional intent. This may or may not provoke congressional action rendering stricter FDA regulation moot.

Third, there are the issues of safety and consumer protection. Critics will be quick to point out that consumers cannot be exercising a meaningful choice when they do not know what they are choosing.[106] If consumers are unaware of the premises behind homeopathy, or if they are unaware of the differences in regulatory oversight between homeopathic and traditional OTC drugs, they are not only being misled, but they could be unknowingly subjecting themselves to harm. This harm could be the result of substituting homeopathic drugs for products that have been proven effective or of taking an uninformed risk on a product that has not been tested for safety. Ultimately, consumers cannot protect themselves from a product that not even FDA knows is unsafe when it enters the market—case in point: Zicam.

---

104. *See* Lawrence J. Schneiderman, *The (Alternative) Medicalization of Life*, 31 J.L. MED. & ETHICS 191, 195 (2003) (recounting a survey conducted by the author and one of his medical students which consisted of interviewing 100 people in San Francisco who were consulting homeopathic practitioners—he states that these "were not unsophisticated people who were unaware of modern medicine" but rather highly educated people with chronic and painful conditions, such as chronic asthma and chronic arthritis, who were seeking treatment outside scientific medicine which "did not give them the cure they were hoping for"); *The Role of Early Detection and Complementary and Alternative Medicine in Women's Cancers: Hearing Before the H. Comm. on Government Reform*, 106th Cong. 8 (1999) (opening statement of Rep. Dan Burton, Chairman, H. Comm. on Government Reform) (orating the committee's goal to "break through barriers of institutional bias" against complementary and alternative therapies for cancer in an effort to improve the availability of information and treatment options to citizens suffering from the disease).

105. *See, e.g.*, *Alternative Medicines: Hearing Before the Subcomm. on Labor, Health, and Human Servs., and Education, and Related Agencies of the S. Comm. on Appropriations*, *supra* note 32, at 2–3 (opening statement of Sen. Tom Harkin, Member, Subcomm. on Labor, Health, and Human Servs., and Education, and Related Agencies of the S. Comm. on Appropriations) (touting the benefits of CAM and describing efforts, including the establishment of NCCAM within NIH and the grant of funding to the White House Commission on Complementary and Alternative Medicine, to enable the public to access such treatments). *But see* Bridget M. Kuehn, *Despite Health Claims by Manufacturers, Little Oversight for Homeopathic Products*, 302 JAMA 1631, 1631 (2009) (reporting that clinical trials on CAM products have been hampered in the homeopathic context due to many trials on such products containing methodological flaws, such as not having an appropriate placebo control, resulting in proposals to NCCAM being necessarily rejected for funding).

106. *See* NAT'L COUNCIL AGAINST HEALTH FRAUD, *supra* note 14 (accusing drug manufacturers of labeling their products as homeopathic in order to evade regulation, resulting in an "explosion" of such products in recent years that consumers do not fully understand).

B. A Homeopathic OTC Drug Review?

If FDA finally followed through with its 1972 promise to subject homeopathic OTC products to a separate official drug review, it would face many of the same problems that it did in 1972.[107] In addition to limited resources and complicated logistics, the largest problem facing a full review of OTC homeopathic drugs is the efficacy issue. The HPUS states, "Because homeopathic drug provings are pharmacological studies on healthy volunteers, they are quite similar to Phase I clinical trials."[108] Examination of this deceptively vague statement reveals the incompatibility of the HPUS with the standards articulated by FDA, and thus the futility of a homeopathic drug efficacy review.

First, although FDA-mandated Phase I studies on conventional drugs are performed on healthy volunteers, they mainly focus on discerning the initial safety picture of the drug and the exact pharmacological effect of the drug on the human body.[109] They are a very preliminary step in the long path to FDA drug approval, as a drug's NDA approval will generally be conditioned on documentation of several subsequent controlled clinical trials definitively proving a positive risk–benefit ratio between the safety and effectiveness of the drug.[110] Conversely, a single homeopathic proving satisfies the evidentiary standards of the HPUS.[111] Provings are

107. *See supra* notes 59–60 and accompanying text.

108. HPUS REVISION SERVICE, *supra* note 66.

109. *See* HUTT, MERRILL & GROSSMAN, *supra* note 40, at 630 (excerpting from a FDA Center for Drug Evaluation and Research handbook that describes the purposes of Phase 1 trials on investigational new drugs, which include, as a purely secondary measure, the gathering of early evidence of effectiveness "if possible").

110. *See generally* 21 C.F.R. § 314.50 (2009) (setting forth the required content of an NDA, including information on chemistry and pharmacology of the drug, animal studies, multiple human studies, and patent information); Gary L. Yingling & Ann M. Begley, *Clinical Research Requirements for New Drug Applications*, *in* THE PHARMACEUTICAL REGULATORY PROCESS, *supra* note 40, at 199–212 (describing the three phases of clinical studies—phase one: toxicology; phase two: dose range; and phase three: efficacy—and detailing the many compliance considerations that must be taken into account during this testing, including sponsor submission requirements, clinical investigator oversight, institutional review board approval, and proper record keeping and reporting to FDA). Prescription-to-OTC switching and approval of generic medications do not require the extensive testing mandatory for NDA submissions; however, these processes occur subsequent to NDA approval of a drug which has thus already satisfied safety and efficacy requirements. *See generally* PRAY, *supra* note 22, at 180–81 (discussing methods by which approved prescription drugs can be switched to OTC and showing that "new" OTC drugs generally arrive on the OTC market by first satisfying NDA requirements to be sold as prescription drugs and later switching to OTC); Marc S. Gross et al., *Generic Drug Approval Process: Hatch-Waxman Update*, *in* THE PHARMACEUTICAL REGULATORY PROCESS, *supra* note 40, at 61 (detailing the Drug Price Competition and Patent Term Restoration Act, or Hatch–Waxman Act, of 1984, which established the current generic drug regime).

111. *See* HPUS REVISION SERVICE, *supra* note 66 (listing the documents required for

administrations of substances to healthy persons to observe what symptoms are produced, and thus what symptoms the substance can potentially treat. Notably, this process fails to test the actual product on people experiencing those symptoms. This fact flies in the face of FDA's traditional clinical trial requirements.[112]

Second, the HPUS does not list the symptoms that a substance produced in a proving. HPUS monographs list only descriptive elements of active ingredients, details on drug preparations, and the lowest potency at which a product may be sold for OTC use; they therefore do not clearly correlate drugs with their effect on the body, which would be necessary for any traditional FDA efficacy determination.[113] In comparison, a United States Pharmacopeia (USP) monograph for an active drug ingredient features a detailed description of a chemical assay that produces the pharmacological effect of the drug in the human body.[114] Additionally, each OTC drug monograph in the CFR states the specific conditions to be treated by each active ingredient, which are thus permitted on the labeling of products containing approved active ingredients because the ingredients have been clinically proven safe and effective for use with such conditions.[115] FDA's

---

acceptance of a homeopathic remedy into the HPUS, which include (1) documents showing the qualification of the principal investigator, which must show that he or she has been a homeopathic drug prescriber for five years and have at least two years' experience in drug provings; (2) the proving protocol, which "should be carried out according to Hahnemann's classical directions" as stated in his ORGANON OF HOMEOPATHIC MEDICINE (North American Academy of the Homeopathic Healing Art eds., 1836); and (3) the final report of the homeopathic drug proving, which must include, among other things, a compilation and classification of the symptoms observed and evaluated).

112. *See* 21 C.F.R. § 314.126(b)(2) (2009) (listing requisite elements of an "adequate and well-controlled" study, one element of which is a study design "that permits a valid comparison with a control to provide a quantitative assessment of drug effect," i.e., to evaluate if the drug has the effect it is supposed to have). The HPUS states that other research methods may be considered for clinical verification and potential inclusion in the HPUS, in an acknowledgement "that a range of methods are currently being used in homeopathic drug provings." HPUS REVISION SERVICE, *supra* note 66. Such methods include "case series, outcome studies, prospective observational studies, longitudinal data collection networks, or randomized controlled trials." *Id.* Although such alternative testing may provide more insight into traditional effectiveness information on homeopathic drugs, it does not shift the premise of homeopathic drugs, and thus the efficacy analysis, away from the law of similars.

113. HPUS REVISION SERVICE, *supra* note 66; *supra* note 110 and accompanying text.

114. *See, e.g.*, COMM. OF REVISION, U.S. PHARMACOPEIAL CONVENTION, INC., USP XXII, at 12–13 (1990) (featuring the assay for acetaminophen). Some HPUS monographs feature an assay for accurate preparation of the monograph ingredient; however, the proven pharmacological effect associated with USP assays is absent in HPUS assays due to the reliance of the HPUS on the law of similars.

115. *See, e.g.*, 21 C.F.R. § 341.20 (2008) (listing nasal decongestant active ingredients approved for OTC marketing, such as pseudoephedrine hydrochloride); *id.* § 341.80 (setting forth the precise mandatory labeling for nasal decongestant OTC drugs, including a

governing CPG does refer industry personnel to two century-old books that provide information on the relationship of homeopathic drugs to symptoms and indications for use.[116] The texts, however, are far from comparable to modern scientific proof of efficacy. It is thus unclear how FDA would designate a definitive endpoint for homeopathic drug efficacy testing—it could surely look to homeopathic publications and individual product claims to determine what conditions a drug is supposed to treat, and thus what indications to test for; however, the clinical uncertainty accompanying these sources, evidenced by their reliance on the law of similars, would make any study highly questionable and likely futile.

In considering the option of a homeopathic OTC drug review, one cannot escape the efficacy issue, which has been around since homeopathy's inception. If FDA decided to confront it, the agency would be endorsing the end of homeopathic OTC drugs—few if any could satisfy current efficacy standards, and for this reason even fewer manufacturers would be willing to expend the time and money necessary to prove otherwise. This fact advises against a full homeopathic OTC drug review if homeopathic drugs are to remain on the market as an option for the consuming public. Unlike the efficacy issue, however, the occurrence of serious adverse events associated with homeopathic drugs renders the safety issue a new and very important concern that FDA would be remiss to ignore.[117] A limited homeopathic OTC drug review is warranted to ensure the safety of such products in light of new safety information exposed by the Zicam incident.[118] The modern drug market, which features Internet

---

requirement for a stated indication of temporary relief from nasal decongestion).

116. *See* COMPLIANCE POLICY GUIDE, *supra* note 5, at 107 (stating that John Henry Clarke, M.D.'s volumes on homeopathic drugs, *A Dictionary of Practical Materia Medica* (1902) and the concomitant *A Clinical Repertory to the Dictionary of Materia Medica* (1904), should be reviewed along with other available information by agency personnel in order better understand homeopathic drugs).

117. One potentially positive aspect of the HPUS is that drug provings can be inherent safety tests—by administering substances in larger doses than will be administered in the final diluted products, they can ensure a certain margin of safety. However the HPUS itself states that provings require an alteration of the definition of *adverse event* in the homeopathic context to reflect the fact that all side effects are recorded for purposes of determining how the drug will be used, not what the drug labeling should warn against. HPUS REVISION SERVICE, *supra* note 66. Accordingly, the safety issue remains uncertain.

118. An argument against this approach may point out the existence of other drug safety measures at FDA's disposal. FDA currently houses an adverse event reporting program called MedWatch on the FDA website. FDA, MedWatch: The FDA Safety Information and Adverse Event Reporting Program, http://www.fda.gov/Safety/MedWatch/default.htm (last visited Apr. 11, 2010). The agency is also in the process of implementing a new comprehensive safety data network, titled the Sentinel Initiative, which will pull data from several sources to supply a one-stop-shop for comprehensive adverse event and drug safety information. OFFICE OF CRITICAL

pharmacies, pervasive advertising, and growing expenditures on health care, highlights the potential regulatory and public health implications associated with homeopathic OTC drugs. With a view to FDA priorities, which have traditionally followed a risk-based enforcement approach,[119] the safety issue associated with these products outweighs the efficacy issue, leading to the need to address only homeopathic OTC drug safety in the context of limited FDA resources.[120]

C. The Safety Issue: A Limited Homeopathic OTC Drug Safety Review

When FDA initiated the gargantuan undertaking of the OTC Drug Review, it came up with the novel approach of establishing advisory review panels comprised of experts specially qualified to evaluate distinct

---

PATH PROGRAMS, FDA, THE SENTINEL INITIATIVE: NATIONAL STRATEGY FOR MONITORING MEDICAL PRODUCT SAFETY 13 (2008) http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/UCM124701.pdf. FDA's power over postmarket drug safety was also greatly increased by the Food and Drug Administration Amendments Act of 2007, under which FDA may now require postmarket studies and clinical trials upon discovery of new safety information (i.e., upon new adverse experiences with a drug). FDCA § 505(o)(3)(A)–(B), 21 U.S.C. § 355(o)(3)(A)–(B) (Supp. I 2007). These safety measures, however, were established to monitor drugs that have already undergone traditional safety and efficacy testing. They are additional reactive measures—undertaken after the proactive FDA review and approval process—which are intended to address adverse events that can only emerge when a drug is released for widespread use by the public. Homeopathic drugs, in contrast, are currently subject only to the reactive approach exemplified by the Matrixx warning letter, such drugs having never been subject to any form of premarket safety review that would indicate potential side effects.

119. *See generally* FDA, COMPLIANCE POLICY GUIDES § 440.100, MARKETED NEW DRUGS WITHOUT APPROVED NDAS AND ANDAS (2006), *available at* http://www.fda.gov/ICECI/ComplianceManuals/CompliancePolicyGuidanceManual/ucm074382.htm (outlining how FDA will exercise its enforcement discretion with regard to drugs marketed illegally for failure to obtain required FDA premarket approval). The enforcement priorities of this CPG primarily focus on unapproved marketed drugs that pose safety threats, consistent with FDA's foremost mission of protecting the public health. Leaving the unique situation of homeopathic drugs aside, the policies of this CPG are indicative of FDA's general enforcement approach.

120. *But see generally id.* (discussing the priority of enforcement against both ineffective drugs and health fraud drugs which are likely to pose "indirect health hazards" if the consumer is likely to delay or discontinue appropriate medical treatment in reliance on the drug). Although the health fraud focus of this CPG brings the efficacy issue to the fore, it does not follow that homeopathic drug efficacy should be the primary focus of FDA's attention in regulating homeopathic drugs. The homeopathic CPG addresses health fraud, and in doing so, it indicates that ineffective homeopathic drugs deemed to pose serious indirect health risks will not be overlooked under the dual coverage of the homeopathic CPG and FDA's health fraud regime. *See, e.g.*, United States v. Writers & Research, Inc., 113 F.3d 8, 10–11 (2d Cir. 1997) (upholding FDA's seizure of homeopathic drugs claiming to cure life-threatening diseases). The homeopathic drugs at issue in this case can be characterized as classic serious health fraud drugs for claiming a false curative value which, if relied on, posed the threat of imminent fatality.

therapeutic categories of OTC drugs.[121] The panels evaluated data that came from OTC drug makers, other interested parties, and the available scientific literature to determine safety and effectiveness of an active ingredient for its intended use.[122] In light of the preceding efficacy discussion, the OTC Drug Review model would require alteration to fit the homeopathic context. However, it is a valuable tool with which to approach the task.

The HPUS specifies the minimum dilution—i.e., the highest concentration—at which each listed ingredient can be sold OTC. Although seemingly an adequate safety barrier, the last update of these standards was in 1998,[123] and the Zicam incident has exposed areas that create a cause for concern. Specifically, the HPUS specifies the minimum dilution at which each drug can be sold for "external use," which is defined as application to the eyes, ears, nose, or other bodily surface other than the mouth or other bodily orifice.[124] Minimum dilutions for external use are generally lower than minimum dilutions for general OTC use—i.e., drugs made for topical application can be sold with higher concentrations of active ingredient than drugs intended for ingestion. For example, the HPUS sets the minimum OTC dilution for "zincum gluconium" at 1X, for "zincum bromatum" at 3X, and for "zincum muriaticum" at 6X; however, the minimum external use dilution for all three of these ingredients is listed as "N/A."[125] The HPUS explains that these standards were developed

121. *See* PRAY, *supra* note 22, at 174–75 (discussing the panels and noting that although prior experience with the DESI Review of prescription drugs greatly aided FDA, the task was still "one of the most ambitious and comprehensive programs ever undertaken by the agency"); *see also* 21 C.F.R. § 330.10(a)(1) (2009) (explaining the advisory panel review system).

122. *See* PRAY, *supra* note 22, at 176 (setting out the criteria the panels used to analyze all the relevant data, including the existence of any clinical trials, the development of scientific opinion on the ingredient, and the marketing experience of the OTC drugs, including sales volume and the amount of complaints received); *see also* 21 C.F.R. § 330.10(a)(2) (2009) (setting forth the process by which FDA would request data via publication in the *Federal Register*; all "interested persons" were asked to submit pertinent materials); *id.* § 330.10(a)(4) (stating the safety and efficacy showing requirements of the review).

123. HPUS REVISION SERVICE, *supra* note 66.

124. *Id.* The HPUS also states lower dilutions at which a drug can be sold by prescription. Although lower dilutions inherently implicate greater safety concerns, homeopathic prescriptions should be an ancillary concern for FDA because they are monitored by a licensed practitioner. As previously mentioned, this Comment's discussion pertains strictly to homeopathic OTC products in recognition of the many additional considerations accompanying the issue of homeopathic practitioners, such as licensing, standard of care, reimbursement, and medical integration. For a discussion on these issues, see generally Cohen, *A Fixed Star*, *supra* note 44.

125. HPUS REVISION SERVICE, *supra* note 66.

using "acute toxicity data from the literature."[126] Thus, despite a stated literature examination by the Homeopathic Pharmacopoeia Convention of the United States (HPCUS), the HPUS featured no limit as to the level of zinc concentration that Matrixx could use in its Zicam intranasal zinc products. This is alarming considering that risk indicators linking intranasal zinc application to anosmia have been around for decades in the literature and in practice.[127] The HPUS further states that the minimum dilution levels were determined within a "100-fold margin of safety" based on accidental ingestion by a 10-kilogram (approximately 22-pound) child of an average amount of ingredient in a full drug container, 30 milliliters or 16.2 grams.[128] This methodology seems to ignore external use entirely, despite the knowledge that the term *external use* includes sensitive bodily areas.

In light of the seriousness of the Zicam incident and the shortcomings that it has revealed, FDA action is warranted. The most efficient proposal is for a very limited safety review to evaluate the scientific literature on homeopathic drug ingredients and ensure that the HPUS minimum dilutions do in fact render homeopathic OTC drugs safe under FDA standards. To accomplish this, FDA should follow the OTC Drug Review model to establish an advisory review panel comprised of homeopathic, herbal, and chemical experts that are qualified to evaluate relevant information on each ingredient featured in the HPUS. As with the OTC Drug Review, FDA can issue a call for data submissions, which will enable manufacturers to show safe public experience with a particular ingredient if possible. There are 1,286 monographs in the HPUS, and in comparison to the OTC Drug Review that examined approximately 200 active ingredients, this number may appear daunting. However, a limited homeopathic safety review would only require evaluation for safety, not efficacy for intended uses; unlike the panels of the OTC Drug Review, the homeopathic drug safety panel would not be responsible for drafting

126. *Id.*

127. *See* Warning Letter, *supra* note 2 (acknowledging existing evidence "in the published scientific literature" linking zinc to olfactory damage); Kuehn, *supra* note 105, at 1632 (citing the existence of case studies in the medical literature of anosmina in patients that used zinc intranasally, as well as historical information on how the intranasal use of zinc in an attempt to prevent polio in the 1930s was linked to anosmia); Jablow, *supra* note 1, at 78–80 (describing the failed polio vaccine that resulted in anosmia as well as a more recent University of Colorado Study that concluded there was a direct link between nasal exposure to zinc and olfactory receptor cell damage); *see also supra* Part II.A (discussing Matrixx's inability to claim safety of its intranasal zinc products despite the general safety of oral zinc products in light of published literature showing zinc to cause olfactory harm).

128. HPUS REVISION SERVICE, *supra* note 66.

monographs with detailed labeling and warning requirements.[129] Further, the nature of homeopathic ingredients will render them much easier to review—a long history of safe experience will cover many herb and mineral ingredients, and some ingredients are used in vitamins and dietary supplements, for which abundant safety data should be available.[130] The panel should be responsible only for determining the potencies at which homeopathic ingredients can be safely marketed without going through more thorough testing for full NDA approval. With the aid of the HPCUS and the cooperation of the industry, this should not be a burdensome review. FDA can also considerably preserve resources by addressing the new safety policy through a new CPG instead of promulgating regulations—compliance with FDA minimum dilution standards can be another condition under which homeopathic products may be marketed.[131]

Incorporating a limited safety review into FDA's homeopathic drug regime would likely assuage public concern about homeopathic drug oversight and safety by ensuring more congruence with the existing OTC drug regime that was precipitated by the OTC Drug Review advisory panel approach. Additionally, limiting the review to a discrete safety purpose would be congruent with FDA's risk-based priorities. Ultimately, the benefit of such an approach would be to recognize the potential for

129. FDA's overarching standard in approving a new drug has historically been a determination that the benefits of the drug outweigh its risks. *See* HUTT, MERRILL & GROSSMAN, *supra* note 40, at 694–95 (discussing FDA's choice to employ an integrative approach to the FDCA requirements for safety and efficacy, which textually can be read as independently satisfied standards). The concept of efficacy is inherent in a consideration of benefits; therefore, FDA's analysis is essentially a weighing of efficacy versus safety, with the many detailed considerations such an analysis entails. However, because the current CPG provisionally exempts homeopathic drugs from the FDCA safety and efficacy requirements, homeopathic drugs are not subject to the benefit–risk standard, leaving FDA free to shape a different standard. It appears from the Zicam incident that FDA has already formulated a threshold at which it considers a homeopathic drug too dangerous to be marketed OTC. Thus, FDA can likely utilize existing criteria to set forth a concrete standard of review.

130. *See, e.g.*, HPUS REVISION SERVICE, *supra* note 66 (featuring the monograph for caffeine).

131. The main drawback to this approach is the fact that compliance policy guides are just that—policy guides, not rules that carry the force of law. *See generally* Stephen M. Johnson, *Good Guidance, Good Grief!*, 72 MO. L. REV. 695 (2007) (addressing the use of interpretive policy documents by agencies to create binding rules outside of the notice-and-comment process prescribed in the Administrative Procedure Act—one complaint against this method is the uncertainty surrounding the vitality of such policy documents created by conflicting opinions on the degree of judicial deference owed to them). However, considering the facts that the current homeopathic drug regime is anchored in a CPG, and that FDA has historically been able to utilize its enforcement discretion to obtain compliance with the policies enumerated therein, continuing with the CPG approach for any new safety criteria coming out of the homeopathic drug advisory panel review is likely the most efficient approach.

harm and prevent it—surely such a review would have prevented hundreds of unknowing consumers from using Zicam intranasal products and temporarily or permanently losing their sense of smell.

D. The Consumer Protection Issue: New Homeopathic OTC Drug Labeling Requirements

In the Dietary Supplement Health and Education Act of 1994 (DSHEA), Congress allowed dietary supplement manufacturers to make labeling claims indicating a supplement's effect on the structure or function of the body, under the condition that the claims be substantiated.[132] The claims must also be accompanied by a disclaimer that states, "This statement has not been evaluated by the Food and Drug Administration. This product is not intended to diagnose, treat, cure, or prevent any disease."[133] The regulatory regimes governing dietary supplements and homeopathic drugs are divergent, and should be, in recognition of the differing conceptions behind the products—unlike dietary supplements, homeopathic drugs are very much intended to diagnose, treat, cure, or prevent disease. Nevertheless, although the dietary supplement regime may not be applicable in the homeopathic drug context, the labeling approach that Congress took in DSHEA is valuable in considering revised labeling requirements for homeopathic products.

The policy behind DSHEA was to facilitate consumer access to safe dietary supplements.[134] The legislation encouraged dissemination of truthful information regarding such products in the form of structure and function effectiveness claims, which FDA had previously disallowed.[135] Although most homeopathic drug claims are different from the claims

---

132. *See* FDCA § 403(r)(6), 21 U.S.C. § 343(r)(6) (2006) (delineating the types of structure and function claims manufacturers may make regarding dietary supplements and the conditions under which such claims will not render the products misbranded); *see also* FDCA § 201(g)(1)(D), 21 U.S.C. § 321(g)(1)(D) (2006) (stating that a dietary supplement does not become a drug if it makes a claim in accordance with the FDCA provisions regulating such claims).

133. FDCA § 403(r)(6)(C), 21 U.S.C. § 343(r)(6)(C) (2006).

134. *See* Dietary Supplement Health and Education Act of 1994, Pub. L. No. 103-417, § 2(13), 108 Stat. 4325, 4325–26 (1994) (stating congressional intent not to impose unreasonable regulatory barriers on the flow of safe dietary supplements and accurate information regarding their benefits to consumers).

135. *See id.* § 2(5) (touting the societal benefits to be gained from promotion of education regarding good nutrition and safe use of nutritional supplements); PRAY, *supra* note 22, at 212–16 (discussing the two prior pieces of legislation in 1990 and 1992 in which Congress attempted to loosen FDA's restrictions on dietary supplement claims, and the 1994 legislation, DSHEA, which finally succeeded). Senator Orrin Hatch, the driving force behind DSHEA, stated that FDA's "single-minded" approach forced Congress to intervene. *Id.* at 216.

allowed by DSHEA, the congressional method of ensuring that the public did not take away the wrong message from the claims—the required disclaimer—can be utilized for the same purpose with homeopathic OTC drugs. To alert consumers to the truthful scientific and regulatory posture of homeopathic OTC drugs, FDA should require a labeling statement indicating that FDA has not evaluated the effectiveness of the product for its intended use. In addition, to ensure that consumers know the precise contents of a drug, FDA should require that homeopathic OTC drug labeling reflect the amount of active ingredient in traditional units of measurement and percentage next to the homeopathic dilution level. This latter requirement will enable consumers who are not familiar with homeopathic theory to understand exactly what they are purchasing—if a product features such a small percentage of an active ingredient as to render it essentially nonexistent, this labeling requirement will let consumers know. Concomitantly, it will apprise consumers who may be expecting a very high dilution from a homeopathic product if a drug is actually much less diluted (i.e., more potent) than traditional homeopathic theory would suggest. Akin to the policy behind DSHEA of enabling access while promoting informed purchasing, such an approach would sufficiently apprise consumers of necessary purchasing information that is currently lacking in homeopathic OTC drugs.

This approach has the advantage of not only preserving consumer choice but also promoting a more meaningful choice by providing consumers with the details necessary to make an informed decision. To implement the new labeling requirements, FDA can again utilize the CPG approach.[136] Although the homeopathic regulatory framework would continue to be a more attractive option than the traditional drug regime for manufacturers who deliberately seek to avoid efficacy requirements, informative labeling will at least decrease instances of mistaken purchases of such products. Moreover, by utilizing labeling as a vehicle for preserving consumer choice, FDA would have the benefit of following the lead of Congress who implemented the method with regard to dietary supplements. This could create a potential safe harbor for FDA's new homeopathic OTC drug regime.

CONCLUSION

The governing CPG listing conditions under which homeopathic drugs may be marketed was released over twenty years ago at the brink of the Internet age and the early stage of growth in CAM popularity among the

136. FDA already prescribes its homeopathic drug labeling requirements via the governing CPG. *See supra* Part II.C; COMPLIANCE POLICY GUIDE, *supra* note 5, at 107–09.

American consuming public. Not only have these phenomena proven to be much more than passing trends, but traditional concepts of information dissemination and product availability have greatly expanded along with them. The Zicam incident brings to the fore FDA's insufficient oversight of homeopathic drugs and raises concern over the potential effects of such a lax approach in modern society. FDA can make the choice to continue with the reactive enforcement scheme of its current CPG despite changing conditions, which may be a valid resource-conserving approach. However, in light of the heightened public scrutiny raised by the Zicam incident and a new Administration that is clearly encouraging FDA to meaningfully fulfill its mission, now may be the best time to reexamine the homeopathic drug regulatory regime and shape it to reflect modern expectations of safety and information in medical choices.

STATUTORY STRUGGLES OF ADMINISTRATIVE AGENCIES: THE DIRECTOR OF NATIONAL INTELLIGENCE AND THE CIA IN A POST-9/11 WORLD

LAUREN C. CLARK*

TABLE OF CONTENTS

* J.D. Candidate, May 2011, American University Washington College of Law; M.S. Global Leadership, University of San Diego School of Business Administration, 2007; B.S. Political Science, United States Naval Academy, 2004. I would like to thank Professor James Zirkle for his guidance in developing this topic as well as his constant feedback throughout the writing process. I would also like to thank Professor Adeen Postar, Colleen B. O'Boyle, and the rest of the *Administrative Law Review* staff for their generosity in assisting with research, editing, and supporting my work. This Comment is dedicated to my parents for their never-ending support and encouragement, and to Chad, whose eternal optimism, love, strength, and patience continue to inspire me daily.

INTRODUCTION

Following the terrorist attacks of September 11, 2001, Congress passed the Intelligence Reform and Terrorism Prevention Act of 2004 (IRTPA)[1] in hopes of changing numerous perceived failures of the United States intelligence community.[2] The IRTPA established, among other things, the position of the Director of National Intelligence (DNI) to oversee, coordinate, and improve the performance of the various United States intelligence entities.[3] The DNI is a cabinet-level official who serves as the principle advisor to the President and National Security Council on intelligence-related matters.[4] With centralized access and enhanced oversight into various intelligence activities, the DNI would presumably improve the United States intelligence community and prevent another 9/11-style attack on American soil.

Prior to the passage of IRTPA in 2004, the United States intelligence community was a compartmentalized and competition-based system of civilian and military intelligence assets,[5] held loosely together by a Director

1. Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004 , Pub. L. No. 108-458, § 1011, 118 Stat. 3638, 3643–44 (2004) (codified as amended in scattered sections of the National Security Act of 1947, 50 U.S.C. §§ 401 to 403-6 (2006)).

2. RICHARD A. BEST, JR., CONG. RESEARCH SERV., INTELLIGENCE ISSUES FOR CONGRESS 2 (2009), http://www.fas.org/sgp/crs/intel/RL33539.pdf.

3. IRTPA § 1011.

4. Office of the Director of National Intelligence, About the ODNI, http://www.dni.gov/who.htm (last visited Jan. 14, 2010).

5. National Security Act of 1947, 50 U.S.C. § 401a(4) (2006) (defining the members of the United States intelligence community to include the following: Central Intelligence Agency (CIA), Department of State Bureau of Intelligence and Research (INR), Defense Intelligence Agency (DIA), National Security Agency (NSA), National Reconnaissance Office (NRO), National Geospatial-Intelligence Agency (NGA), Federal Bureau of Investigation (FBI), Army Intelligence, Navy Intelligence, Air Force Intelligence, Marine Corps Intelligence, Department of Homeland Security (DHS), Coast Guard (CG), Treasury Department, and Energy Department).

of Central Intelligence (DCI).[6] The DCI had three main duties: to direct the Central Intelligence Agency (CIA), to be the intelligence advisor to the President, and to be the central coordinator of the various intelligence agencies and departments.[7] To accommodate these tasks, through the years the President would issue orders in an attempt to expand the DCI's power and centralize the DCI's role within the intelligence community.[8] However, more power often resulted in more responsibility, leaving the DCI with too many tasks and not enough resources to complete them.[9] The IRTPA, acknowledging these previous struggles, sought to separate the DCI's three tasks, giving the newly created DNI responsibility for overseeing the United States intelligence community and acting as an advisor to the President[10] while leaving the task of running the day-to-day operations of the CIA to the Director of the Central Intelligence Agency (DCIA).[11] The new DNI authority, outlined in the IRTPA and codified as amended in the National Security Act of 1947, includes authority to specify the intelligence budget, transfer funds and personnel across the intelligence community, and develop priorities for intelligence collection and analysis.[12] But even with stronger statutory powers and a more centralized structure, the DNI has received his share of criticism in the last five years.

One of the chief complaints against the DNI was that even with enhanced authority, Congress still had not bestowed the DNI with enough

---

6. *See generally* Exec. Order No. 12,333, 3 C.F.R. 200, 202–03 (1982) (a competition-based, decentralized system).

7. *See id.* at 202–04 (describing the range of the DCI's duties as coordinator of U.S. intelligence activities); *see also* GEORGE J. TENET, DIRECTOR OF CENTRAL INTELLIGENCE DIRECTIVE 1/1: THE AUTHORITIES AND RESPONSIBILITIES OF THE DIRECTOR OF CENTRAL INTELLIGENCE AS HEAD OF THE U.S. INTELLIGENCE COMMUNITY (1998), http://www.fas.org/irp/offdocs/dcid1-1.htm.

8. *See* CIA CTR. FOR THE STUDY OF INTELLIGENCE, CENTRAL INTELLIGENCE: ORIGIN AND EVOLUTION 6–11 (Michael Warner ed., 2001), https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Origin_and_Evolution.pdf [hereinafter CIA ORIGIN AND EVOLUTION REPORT] (noting the various attempts to expand the DCI powers through executive orders in an effort to reach the centralization envisioned by President Harry S. Truman when he signed the National Security Act of 1947 into law, while accommodating the DCI's duty to run effective operational intelligence activities at the CIA).

9. *Id.*

10. OFFICE OF THE DIR. OF NAT'L INTELLIGENCE, VISION 2015: A GLOBALLY NETWORKED AND INTEGRATED INTELLIGENCE ENTERPRISE 21 (2008), http://www.dni.gov/Vision_2015.pdf.

11. Under IRPTA the Director of Central Intelligence (DCI) became the Director of the Central Intelligence Agency (DCIA). IRTPA, Pub. L. No. 108-458, § 1071, 118 Stat. 3638, 3689 (2004) (replacing "Director of Central Intelligence" with "Director of National Intelligence" or "Director of the Central Intelligence Agency" where applicable).

12. *Id.* § 1011(a) (adding to the National Security Act of 1947, 50 U.S.C. §§ 403(b)(1), (b)(2), 403-1(c)(1)(B), (f)(1)(A), (3)(A), (5), (g)(1) (2006)).

authority to control and unify the historically autonomous intelligence departments and agencies.[13] In 2008, President George W. Bush responded to this complaint with Executive Order 13,470, further delineating the specific DNI responsibilities under the IRTPA.[14] Despite this executive order, questions of authority continue to arise when the DNI promulgates changes within the intelligence community.[15]

There are several recent examples of statutory ambiguity and overlap of legal authority between the DNI and the CIA as a result of IRTPA's implementation. First is the over-publicized turf battle between the CIA and the DNI concerning appointment of overseas station chiefs.[16] Traditionally, the CIA has been in charge of appointing these positions, but the new DNI statutory authority suggests the DNI may also have some control.[17] The issue of who has the power to appoint these positions fueled national news headlines for months before the White House resolved the issue.[18]

Another example of statutory ambiguity is the DNI's Intelligence Community Directive (ICD) establishing the National Intelligence Civilian Compensation Program (NICCP).[19] NICCP is a DNI initiative to replace individualized pay systems currently used by each of the intelligence entities with a uniform, community-wide, compensation-based pay system.[20] Although the individual agencies and departments appear to have adopted this directive voluntarily, IRTPA does not give the DNI explicit authority to make these entities comply.[21]

A third example of statutory ambiguity can be seen in the DNI's administrative authority to address Freedom of Information Act (FOIA) classification and declassification issues as they relate to the CIA's FOIA

---

13. *See* Pam Benson, *In Today's Intelligence Hierarchy, Who* Really *Runs the Show?*, CNN.COM, Feb. 12, 2009, http://www.cnn.com/2009/POLITICS/02/12/cia.dni/index.html; *see also* Fred Kaplan, *You Call That a Reform Bill?*, SLATE, Dec. 7, 2004, http://www.slate.com/id/2110767.

14. Exec. Order No. 13,470, 3 C.F.R. 218 (2009).

15. *See infra* notes 50–51.

16. *See generally* David Ignatius, *Duel of the Spy Chiefs*, REAL CLEAR POLITICS, June 11, 2009, http://www.realclearpolitics.com/articles/2009/06/11/duel_of_the_spy_chiefs_96947.html (detailing heated exchanges between the DNI and DCIA on who should appoint the overseas station chiefs, with insiders calling the DCIA's response "an act of insubordination" and President Obama being "peeved" with the entire ordeal).

17. *See infra* notes 71–74 and accompanying text.

18. *See infra* notes 75–76 and accompanying text.

19. INTELLIGENCE COMMUNITY DIRECTIVE NO. 650: NATIONAL INTELLIGENCE CIVILIAN COMPENSATION PROGRAM; GUIDING PRINCIPLES AND FRAMEWORK (2008), http://www.fas.org/irp/dni/icd/icd-650.pdf.

20. *Id.* at 1–3.

21. *See infra* notes 80, 82 & 88 and accompanying text.

authority. The intelligence reorganization under the IRTPA granted the DNI exclusive authority to "protect intelligence sources and methods from unauthorized disclosure" and to prepare intelligence products for dissemination.[22] However, as was the practice before the IRTPA, the CIA continues to respond individually to FOIA requests and process declassification requests.[23]

Lastly, not only do the statutory ambiguities create uncertainty in agency administration of FOIA and pay systems, but they also have operational implications. The DNI's National Counterterrorism Center (NCTC), created under the IRTPA, was an effort to centralize various counterterrorism efforts throughout the intelligence and homeland security communities.[24] However, the NCTC still competes with the long-established CIA Counterterrorism Center (CTC) and demonstrates the IRTPA's failure to resolve operational conflict and redundancy.[25]

While the spirit and intent of the IRTPA suggest intelligence agencies such as the CIA will work in concert with the DNI when implementing these directives and initiatives, the DNI has acknowledged there are legal inconsistencies as to how this will take place.[26] Unless future amendments through Congress or through executive orders fix these ambiguities and overlaps,[27] potential conflicts over future intelligence directives will continue to threaten the success of a centralized intelligence community, detracting from its vital mission of securing the nation.

This Comment addresses whether the DNI, under the IRTPA, has the proper authority to effectively integrate and unify the United States intelligence community by evaluating the current statutory guidelines and clashes of authority between the DNI and the CIA. Part I of this Comment examines the development of the DNI's statutory authority under the National Security Act of 1947, as amended by the IRTPA in 2004 and

---

22. 50 U.S.C. § 403-1(i)(1), (2)(c) (2006).

23. Exec. Order No. 12,958, 3 C.F.R. 333, 346–47 (1996), *as amended in* 70 Fed. Reg. 21,609 (Apr. 26, 2005) (allowing agencies that receive a FOIA request to respond with declassification of the information or state a valid exemption).

24. *See infra* notes 102–104 and accompanying text.

25. *See infra* note 97 and accompanying text.

26. INTELLIGENCE COMMUNITY DIRECTIVE NO. 650, *supra* note 19, at 1 n.1 ("A legal determination as to whether the language in this [Implementation and Administration] paragraph is necessary under the IRTPA, in order for the DNI to execute this ICD, has not been made.").

27. *See* IRTPA, Pub. L. No. 108-458, § 1018, 118 Stat. 3638, 3670 (2004) (insisting that the DNI's authority "respects and does not abrogate the statutory responsibilities of the heads of the departments of the United States Government" including the CIA); Exec. Order No. 13,470 § 1.3(c), 3 C.F.R. 218, 224 (2009) (restating the language of the IRTPA, that statutory authorities of intelligence agencies like the CIA will not be abrogated by decisions of the DNI).

Executive Order 13,470 in 2008. Part II analyzes the areas of statutory ambiguity of the DNI's current power and the apparent overlap between the DNI and CIA administrative authority and the effect it has had, and will continue to have, on the relationship between the DNI and CIA. Examples discussed include appointment of overseas station chiefs, efforts to streamline the intelligence community's employee pay system, overlap of classification and declassification procedures as they relate to the intelligence community's FOIA request process, and the operational redundancy of counterterrorism centers. Finally, Part III evaluates various proposed solutions to these statutory problems and suggests ways to improve the relationship of the DNI over the intelligence community by setting forth what authority should stay with intelligence entities like the CIA and what power should be designated to the DNI.

I. BACKGROUND OF INTELLIGENCE COMMUNITY STATUTORY POWERS

Over the last fifty years, volumes of amendments, National Security Council intelligence directives, and executive orders detail an ongoing struggle to find the most effective organization of the intelligence community following Congress's original plan under the National Security Act of 1947.[28] The 2004 IRTPA was not the first attempt to reorganize the intelligence structure established in 1947 but a concerted effort to again effectuate change after a long line of marginally successful attempts to address decades of perceived shortcomings of a less-than-cohesive intelligence community.[29]

A. Pearl Harbor: A Catalyst for Change

Congress developed the National Security Act in 1947 in response to United States intelligence failures that contributed to the successful Japanese attack on Pearl Harbor in 1941 and led to United States involvement in World War II.[30] The *Joint Committee Report on the Investigation*

28. *See* CIA ORIGIN AND EVOLUTION REPORT, *supra* note 8, at 1–2 (detailing frustrations of the DCI's inability to truly run and coordinate national intelligence collection, which manifested into years of attempts to reform the position by the National Security Council, presidents, and Congress, each time being tempered with fears of excessive concentration of power in such a covert arena of government).

29. *See id.* (noting the numerous NSC intelligence directives and executive orders aimed at reforming the intelligence community).

30. JOINT COMM. ON THE INVESTIGATION OF THE PEARL HARBOR ATTACK, 79TH CONG., INVESTIGATION OF THE PEARL HARBOR ATTACK 252–54 (Comm. Print 1946) (reporting the intelligence deficiencies discovered through the Committee's investigation and outlining recommendations to ensure unity in the United States intelligence system).

of the Pearl Harbor Attacks demanded a "centralization of authority and clear-cut allocation of responsibility" within the intelligence community to prevent another attack.[31] Congress responded with the National Security Act of 1947,[32] which established the CIA as an independent agency responsible for "overseeing strategic analysis and coordinating clandestine activities abroad."[33] At the same time, its director, the DCI, would advise the National Security Council of all intelligence matters and would also produce "national intelligence" by coordinating with the various intelligence departments and agencies.[34] Through the past several decades, amendments, intelligence directives, and executive orders have attempted to provide the DCI more power to effectively centralize intelligence-gathering tasks.[35] However, these efforts seemed to not be working.[36] By 1992, members of Congress began to introduce new bills to reorganize and develop a more coherent and unified intelligence community under a "Director of National Intelligence."[37] Proponents of the reorganization argued that the DCI was overtasked and lacked the power necessary to exercise proper authority over the intelligence community.[38] It was not

31. *Id.* at 254.

32. *See* HISTORY STAFF, CTR. FOR THE STUDY OF INTELLIGENCE, CIA, CIA COLD WAR RECORDS: THE CIA UNDER HARRY TRUMAN 131–35 (Michael Warner ed., 1994) (providing a reproduction of the original intelligence section of the National Security Act of 1947); *see also* Loch K. Johnson, *A Centralized Intelligence System: Truman's Dream Deferred*, 23 AM. INTELLIGENCE J. 6, 6–8 (2005) (suggesting that President Truman's desire to commission a single, cohesive intelligence report became an executive order creating the CIA and the DCI).

33. CIA.gov, *A Look Back . . . The National Security Act of 1947*, https://www.cia.gov/news-information/featured-story-archive/2008-featured-story-archive/national-security-act-of-1947.html.

34. *See generally* NATIONAL SECURITY COUNCIL INTELLIGENCE DIRECTIVE NO. 1: DUTIES AND RESPONSIBILITIES (1950), http://fas.org/irp/offdocs/nscid01.htm.

35. CIA ORIGIN AND EVOLUTION REPORT, *supra* note 8, at 7–12; *see also* Exec. Order No. 12,333, 3 C.F.R. 200 (1982) (further distinguishing the role and responsibilities of the DCI from what they were in the National Security Act of 1947).

36. *See* CIA ORIGIN AND EVOLUTION REPORT, *supra* note 8, at 7–12 (explaining that although Cold War administrations added to DCI's responsibilities, these changes were limited in scope).

37. *See, e.g.*, *S. 2198 and S. 421 to Reorganize the United States Intelligence Community: Joint Hearing Before the S. Select Comm. on Intelligence and the H.R. Permanent Select Comm. on Intelligence*, 102d Cong. 2 (1992). Recommendations from this proposed legislation and the companion bill offered in the House of Representatives, H.R. 4165, were partially incorporated into the Intelligence Organization Act of 1993, which strengthened the powers of the DCI by codifying increased budgetary powers and provided the DCI with expanded authority to shift certain foreign intelligence program funds.

38. *See* NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., THE 9/11 COMMISSION REPORT 402–03 (2004) (explaining that eventually DCI George Tenet and his chief aides were coordinating interagency meetings almost every day and that as he became more of a "lead coordinator" of the intelligence community, it became more difficult for him

until after the terrorist attacks of 9/11 and the perceived failures of the intelligence community that contributed to them that Congress finally took action.[39]

B. 9/11: A Second Catalyst for Change Spurs the Creation of the DNI

In December 2004, Congress passed the IRTPA, beginning the most comprehensive reform of the intelligence community since its creation over fifty years ago.[40] This legislation was the result of numerous perceived intelligence failures outlined in the 9/11 Commission Report.[41] The report details an intelligence system geared to "wage the Cold War," and by the late 1990s, the entire system was the product of "the dispersal of effort on too many priorities, the declining attention to the craft of strategic analysis, and security rules that prevented adequate sharing of information."[42] The goal of Congress in enacting the IRTPA was to ensure the new DNI had more authority, and thus more ability to affect change, than the DCI of the original National Security Act of 1947.

Under the IRTPA, the DNI's responsibilities are to serve as the head of the intelligence community and advise the President and National Security Council on intelligence matters.[43] Other new and enhanced authorities include authorizing the DNI to transfer or reprogram funds after "consulting" with the DCIA or other intelligence community department heads.[44] The DNI is also authorized to transfer personnel within the intelligence community for up to two years[45] and exercise authority over the appointment of intelligence community leadership.[46] Lastly, the

"to play all the position's other roles, including that of analyst in chief").

39. *See id.* at 86–91 (analyzing the various intelligence failures preceding 9/11 and the recommendations made to unify the intelligence effort in response to those failures).

40. RICHARD A. BEST, JR. & ALFRED CUMMING, CONG. RESEARCH SERV., DIRECTOR OF NATIONAL INTELLIGENCE STATUTORY AUTHORITIES: STATUS AND PROPOSALS 2 (2008), http://www.fas.org/sgp/crs/intel/RL34231.pdf.

41. *See* NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 86–91 (referencing the structure and organization of the intelligence community and outlining how the various changes in technological capabilities, legislative priorities, and decentralized control contributed to a structure that proved to be ineffective in detecting and responding to the growing threat of terrorism).

42. *Id.* at 91.

43. IRTPA, Pub. L. No. 108-458, § 1011, 118 Stat. 3638, 3644 (2004) (codified at National Security Act of 1947 § 102 (b), 50 U.S.C. § 403(b) (2006)).

44. *Id.* § 1011, 118 Stat. at 3646 (codified at National Security Act of 1947 § 102A(d)(1)(A), (3), 50 U.S.C. § 403-1(d)(1)(A), (B)(3) (2006)); BEST & CUMMING, *supra* note 40, at 1.

45. IRTPA § 1011, 118 Stat. at 3647–48 (codified at National Security Act of 1947 § 102A(e)(2)(A), 50 U.S.C. § 403-1(e)(2)(A)).

46. *Id.* § 1014, 118 Stat. at 3663–64 (codified at National Security Act of 1947 § 106, 50 U.S.C. § 403-6 (2006)) (providing the DNI with the ability to recommend to the

IRTPA gave the DNI greater budgetary authority than that of the DCI.[47]

However, as the DNI started to carry out his new tasks, issues with the IRTPA's ambiguous statutory authority became apparent. By 2007, reports surfaced that the DNI, Michael McConnell, was requesting stronger and clearer delineations on his authority to run the intelligence community.[48] President Bush quickly responded with Executive Order 13,470, which augments the IRPTA by delineating twenty-four specific responsibilities of the DNI.[49] While Executive Order clearly explained the DNI's authorities, it is questionable whether the order actually expanded them.[50] The only new authority Executive Order 13,470 may have added to the authority of the DNI under the IRTPA was the ability to recommend removal of various intelligence community officials.[51]

C. What Is Left for the CIA

While the DNI remained busy determining his new role, the various intelligence agencies and departments were adjusting as well. The agency with the largest adjustment was the CIA. Of the sixteen departments and agencies that comprise the intelligence community,[52] all but the CIA fall under the control of a cabinet-level official.[53] The CIA is the only

---

President individuals to fill the vacancies of the head of the individual intelligence collection agencies and departments).

47. BEST & CUMMING, *supra* note 40, at 8. *Compare* IRTPA § 1011, 118 Stat. at 3644–45 (codified at National Security Act of 1947 § 102A(c)(1)(B), 50 U.S.C. § 403-1(c)(1)(B)) (authorizing the DNI to "develop and determine" the National Intelligence Program (NIP) budget), *with* 50 U.S.C. § 403-3(c)(1)(A) (2000) (authorizing the DCI to "facilitate the development" of the NIP budget).

48. Shaun Waterman, *State of Security: DNI: Lacking Power–1*, UPI.COM, Apr. 10, 2007, http://www.upi.com/Security_Industry/2007/04/10/State-of-Security-DNI-Lacking-power-1/UPI-43201176209633/.

49. Exec. Order No. 13,470 § 1.3(b)(1)–(24), 3 C.F.R. 218, 220–24 (2009) (amending Exec. Order 12,333, 3 C.F.R. 200 (1982)).

50. *See* Joseph Anzalone et al., *National Security*, 43 INT'L LAW. 929, 937–38 (2009) (explaining that Executive Order 13,470 merely reiterates most of the authorities granted to the DNI by the original text of IRTPA and clarifies the IRTPA authority by enumerating responsibilities, but it fails to bestow any new, substantial authority to the DNI beyond the original IRTPA legislation).

51. *See id.* at 938 (noting that Executive Order 13,470 also highlighted the importance of DNI consultation with the heads of the various intelligence community agencies and departments, which could be construed as enhancing that power relative to the other members of the intelligence community).

52. *See supra* note 5 (listing the sixteen agencies that make up the intelligence community).

53. *See* Intelligence.gov, Members of the Intelligence Community, http://www.intelligence.gov/1-members.shtml (last visited Jan. 22, 2010) (stating that all of the intelligence offices or agencies fall under the control of a cabinet-level position with the exception of the CIA).

intelligence unit exposed to the direct authority of the DNI, subjecting it to closer scrutiny and less protection than its counterparts with nonintelligence cabinet-level leadership.[54]

The IRTPA effectively stripped the DCI of two of his three primary responsibilities—he no longer serves as the President's advisor on national-intelligence issues, and he no longer has the authority to set collection and analysis priorities as the head of the intelligence community.[55] Pursuant to the IRTPA, the DCI's new responsibilities include "collect[ing] intelligence through human sources and by other appropriate means"; correlating, evaluating, and disseminating intelligence related to national security; "providing overall direction for and coordination of the collection of national intelligence outside the United States through human sources"; and performing other functions, under DNI direction, such as coordinating relationships between the intelligence services of other countries, or other tasks from the DNI.[56] Additionally, the DCI's title was changed from Director of Central Intelligence to Director of the Central Intelligence Agency.[57] Although IRTPA altered some of the CIA's authority, the basis of its statutory authority is still the Central Intelligence Agency Act of 1949.[58]

The IRTPA of 2004 focused instead on shifting powers to the new players, like the DNI, to unify intelligence efforts. However, evaluating the effectiveness of this restructure is just beginning.

---

54. The significance of cabinet-level protection from the DNI is that Department of Defense intelligence agencies like the National Security Agency and the Defense Intelligence Agency are afforded certain budgetary protections by the Secretary of Defense, which limits perceived control over them. *See* IRTPA § 1011, 108 Pub. L. No. 458, § 102A(c)(3)(A), 118 Stat. 3638, 3645 (2004) (authorizing the DNI to "participate in the development by the Secretary of Defense of the annual budgets for the Joint Military Intelligence Program and for Tactical Intelligence and Related Activities Program"); *see also* Benson, *supra* note 13 (pointing out that while the National Security Agency and the National Reconnaissance Office report directly to the Defense Secretary and not the DNI, the CIA acknowledges that its only "boss" is the DNI, which highlights the disproportionately larger amount of control the DNI has over the CIA compared with the Department of Defense intelligence agencies).

55. *See* 50 U.S.C. § 403(b)(1)–(2) (2006) (reassigning two of the roles previously held by the DCI, serving as the head of the intelligence community and serving as the President's intelligence advisor, to the DNI); BEST & CUMMING, *supra* note 40, at 1–2 (discussing IRTPA's reassignment of roles).

56. § 403-4a(d)(1)–(4).

57. IRTPA § 1071, 118 Stat. at 3689–92 (replacing "Director of Central Intelligence" with "Director of National Intelligence" or "Director of the Central Intelligence Agency" where applicable); BEST & CUMMING, *supra* note 40, at 2.

58. Central Intelligence Agency Act of 1949, Pub. L. No. 81-110, 63 Stat. 208 (codified as amended at 50 U.S.C. §§ 403a–403s (2006)).

II. PROBLEMS WITH THE IRTPA AND DNI STATUTORY AUTHORITY

During the confirmation hearing of DNI nominee Mike McConnell in early 2007, Senator John D. Rockefeller stated,

> [B]eyond the act of separating the two jobs, it is less clear whether the structure of the DNI office is ideal to accomplish its mission . . . . We did not pull the technological collection agencies out of the Defense Department and we did not give the DNI direct authority over the main collection or analytical components of the community. We gave the DNI the authority to build the national intelligence budget, but we left the execution of the budget with the agencies. We gave the DNI tremendous responsibilities. The question is, did we give the position enough authority for him to exercise those responsibilities?[59]

In many ways, it seems Senator Rockefeller is correct. While the spirit of unity and cooperation is apparent from the text of the IRTPA, its real-world impact will meet numerous roadblocks and require modification.

A. The Loopholes

The first issue hindering the progress of the Office of the DNI is § 1018 of the IRTPA, Presidential Guidelines on Implementation and Preservation of Authorities.[60] This section states that the President will provide the DNI with guidelines to implement and execute his mission as long as it is done "in a manner that respects and does not abrogate the statutory responsibilities of the heads of the departments of the United States Government."[61] This phrase has drawn its fair share of criticism from the legal community at large.[62] Statutory authority of intelligence community members had been established in an atmosphere of relative autonomy prior to 9/11, leaving control over intelligence operations, personnel, and budgets in the hands of the respective agencies and departments rather

59. *Nomination of Vice Admiral Michael McConnell to Be Director of National Intelligence: Hearing Before the S. Select Comm. on Intelligence*, 110th Cong. 1 (2007) (opening statement of Sen. John D. Rockefeller, Chairman of the U.S. Senate Select Committee on Intelligence).

60. IRTPA § 1018, 118 Stat. at 3670 (referenced in the codification of the IRTPA at 50 U.S.C. § 403(b)(3) (2006)).

61. IRTPA § 1018, 118 Stat. at 3670–71 (stating that the applicable department heads list presented is "not limited to" the ones listed, thus allowing the CIA, as an independent government agency, to qualify).

62. *See* BEST, *supra* note 2, at 8 (stating that the concession of the DNI to not abrogate the statutory responsibilities of the individual intelligence units was a hotly debated issue in the drafting of the IRTPA); *see also* Kaplan, *supra* note 13 (noting that the clause in IRTPA § 1018 is a huge loophole hindering the ability of the DNI to enforce any changes and enhancements within the intelligence community, specifically within the Department of Defense, which controls about 80% of the U.S. intelligence community's budget).

than under the coordinated control of a DNI.[63]

As recently as February 2008, DNI Mike McConnell suggested that an executive order was necessary to strengthen the statutory authority the DNI needed to allow him to perform the task of integrating the intelligence community.[64] However, the much-anticipated Executive Order 13,470 merely reiterates the IRTPA § 1018 loophole. It states that the DNI's authority should "not abrogate the statutory or other responsibilities of the heads of departments of the United States Government or the Director of the Central Intelligence Agency."[65] It also states that if any members of the intelligence community believe that the DNI issued a directive or abrogated their individual statutory authority, they can appeal the issue to the National Security Council.[66] This limitation of authority over the intelligence community seems to be the origin for several instances of overlap and friction between the DNI and individual intelligence entities like the CIA.

B. Examples of Statutory Ambiguity and Overlap

In dealing with current issues regarding conflicting authorities between the DNI and the CIA, Congress has been slow to reevaluate the perceived conflicts. Rather, it has opted to deal with each issue as it arises.[67] The problem with this approach is that it prompts Congress to react to each individual problem rather than fix the statute once and save itself future time and effort.[68] Without a clear delineation of authority, whether or not something becomes an issue rests within the discretion of individual

63. *See* Kaplan, *supra* note 13 (referring to the pre-9/11 intelligence community as a "vast, disparate, and sometimes quarrelsome array of federal departments, agencies, and sub-agencies").

64. *See DNI Authorities and Personnel Issues: Hearing of the S. Select Comm. on Intelligence*, 110th Cong. 23 (2008), http://www.dni.gov/testimonies/20080214_transcript.pdf (statement of J. Michael McConnell, Director of National Intelligence) [hereinafter *McConnell Hearing*] (discussing the statutory shortcomings of IRTPA and his anticipation of an executive order which would expand DCI statutory authority over the various intelligence community assets).

65. Exec. Order No. 13,470, 3 C.F.R. 218, 224 (2009) (amending Exec. Order No. 12,333, 3 C.F.R. 200 (1982) and clearing up past ambiguity as to whether or not the CIA, as an agency rather than a department, qualifies for protection).

66. *Id.*; *see also* Anzalone et al., *supra* note 50, at 937–38 (arguing that Executive Order 13,470 did not address or ameliorate the failures of IRTPA but only reinforced the troublesome loopholes and appeal process).

67. *See* INTELLIGENCE AUTHORIZATION ACT FOR FISCAL YEAR 2010, S. REP. NO. 111-55, at 50 (1st Sess. 2009) (providing congressional interpretation of the conflicting authorities of the IRTPA relative to the DNI and DCIA's respective roles in appointing overseas station chief positions, without directly addressing any possible changes to the law to clarify the current issue or prevent future ones).

68. *See id.* (providing an example of Congress's piecemeal response to issues).

agencies to challenge the DNI's directive. The following examples demonstrate statutory ambiguity causing administrative overlap of the DNI and DCIA's powers.

1. Overseas Station Chiefs: Statutory Ambiguity

On May 19, 2009, DNI Dennis Blair issued Intelligence Community Directive 402—a classified directive proclaiming that the DCI would now be able to appoint the top spy in each country, known as an overseas station chief, a job that was traditionally held by the CIA.[69] News outlets, however, proclaimed that the DCIA refused to concede the CIA's traditional duty to appoint station chiefs, igniting controversy as to which position, the DNI or the DCIA, retained the right to appoint the overseas station chief position.[70]

Executive Order 13,470, the Bush Administration's attempt to further clarify and define the authority of the DNI, states that the DNI has authority to enter into agreements with foreign governments and international organizations, as well as the authority to "formulate policies concerning" and "align and synchronize" intelligence relationships with foreign governments and international organizations.[71] This wording likely provides the DNI with the expectation that he would be responsible for the appointment of U.S. station chiefs at overseas intelligence posts.[72] At the same time, however, the CIA's authority states that the DCIA "shall coordinate the relationships between elements of the intelligence community and the intelligence or security services of foreign governments

69. Mark Mazzetti, *Turf Battles on Intelligence Pose Test for Spy Chiefs*, N.Y. TIMES, June 9, 2009, http://query.nytimes.com/gst/fullpage.html?res=9400E0DA1331F93AA35755C0A96F9C8B63; *see also* Darrell Issa, *CIA's Panetta, DNI Blair Must End Turf War and Switch Jobs*, USNEWS.COM, June 18, 2009, http://www.usnews.com/articles/opinion/2009/06/18/cias-panetta-dni-blair-must-end-turf-war-and-switch-jobs.html (stating that the distinction in authority of the DNI and the CIA appointing station chiefs "couldn't be more apparent" and that the DNI's authority is administrative oversight, leaving DCIA Leon Panetta in charge of the "active 'command and control' of the CIA's foreign intelligence officers").

70. *See* Ignatius, *supra* note 16 (arguing that "[t]he right division of labor is to let the CIA run operations, which begins with picking the people who will be America's point of contact with foreign intelligence services" and that IRTPA added "unnecessary new layers of bureaucracy . . . partly duplicating jobs that used to be done by the CIA.").

71. Exec. Order No. 13,470 § 1.3(b)(4)(A)–(C), 3 C.F.R. 218, 220-21 (2009); IRTPA, Pub. L. No. 108-458 § 1011(a), 118 Stat. 3638, 3651–52 (2004) (codified at 50 U.S.C. § 403-1 (2006)).

72. *See* Issa, *supra* note 69 (explaining the difficulty facing both the DNI and the DCIA in confining themselves to the boundaries created by IRTPA, specifically the DNI "resist[ing] the urge to assert command and control" and the DCIA "working within a legal framework that potentially buffers his direct access to the President").

or international organizations on all matters involving intelligence related to national security or involving intelligence acquired through clandestine means."[73]

From a plain-text reading of both of these current authorities, the DNI is tasked with "overseeing" the coordination of intelligence community relationships with foreign governments, while the CIA is tasked with the actual "coordination" of those relationships.[74] Given the current wording of the law coupled with the long-standing tradition of being the sole entity to appoint overseas station chiefs, this similarity in statutory mission explains why the DCIA might feel that the DNI overstepped his statutory authority. Congress threw in its support for the DNI in July 2009.[75] After months of attempting to resolve the issue, the White House finally issued its decision, siding with the DCIA on the issue but also reinforcing the DNI's authority over the intelligence community as a whole.[76]

Even with the dispute currently resolved, this station chief debacle remains an example of how shifting authorities between the DNI and CIA, if not clearly defined in the IRTPA and ensuing legislation, creates problems for unification and cooperation within the intelligence community. With a lack of clear-cut statutory authority, the DNI's powers are only effective when the individual intelligence community entities agree to cooperate.[77]

---

73. IRTPA § 1011(a), 118 Stat. 3638, 3660–61 (codified at 50 U.S.C. § 403-4a(f) (2006)).

74. *Compare* IRTPA § 1011(a) (codified at 50 U.S.C. § 403-1(k) (2006)) ("oversee the coordination"), *with* IRTPA § 1011(a) (codified at 50 U.S.C. § 403-4a(f) (2006)) ("shall coordinate").

75. *See* INTELLIGENCE AUTHORIZATION ACT FOR FISCAL YEAR 2010, S. REP. NO. 111-55, at 50 (1st Sess. 2009) (stating that Intelligence Community Directive 402 "recognizes the value of turning to the CIA Chief of Station to be the DNI's representative in foreign countries" and that in exercising his authority, the DNI "has made the decision that the directive is the right choice for the Intelligence Community. The Committee supports the DNI in that choice and looks forward to the CIA's prompt adherence to his decision.").

76. *See* Posting of Jake Tapper to Political Punch, http://blogs.abcnews.com/politicalpunch/2009/11/white-house-backs-cia-over-dni-in-turf-battles.html (Nov. 12, 2009, 23:05 EST) (reporting that after months of back-and-forth between National Security Advisor Jim Jones and Vice President Joe Biden attempting to resolve the issue, the White House eventually made a decision that the CIA-appointed overseas station chiefs will remain the representatives abroad for the United States intelligence community).

77. For the IRTPA to successfully transform the intelligence community from its once-individualized and autonomous system into a unified and cooperative body, the DNI and DCIA must address statutory conflicts and ambiguity privately rather than detailing rifts and competition. *But see* Issa, *supra* note 69 (calling the issue an outright feud between the CIA and DNI); Ignatius, *supra* note 16 (detailing a duel and a battle over "turf"); Benson, *supra* note 13 (characterizing the issue as a "clash of the titans" with a visible "trench line").

2. National Intelligence Civilian Compensation Program: Statutory Ambiguity

For another area of ambiguity, consider employee compensation. In early 2008 the DNI promulgated Intelligence Community Directive 650 (ICD 650), instructing the various intelligence community entities to abandon their individualized pay systems and adopt a uniform pay-for-performance system.[78] The purpose of the National Intelligence Civilian Compensation Program (NICCP) is to enable the intelligence community to "recruit, motivate, and retain highly qualified individuals . . . and facilitate the rotation of [intelligence community] employees between [intelligence community] components."[79] The IRTPA and subsequent amendments provide the DNI with the power to "*encourage* and *facilitate* the recruitment and retention . . . of highly qualified individuals,"[80] but do not delineate how the DNI should do so. As a result, the lack of explicit DNI authority to control the payment of CIA personnel coupled with the DCIA's customary role of paying CIA employees, the codified loophole in the IRTPA that prohibits the DNI from "abrogating" the CIA's statutory authority,[81] and the wording included within ICD 650[82] could technically allow the CIA to challenge the NICCP.

For instance, the CIA could argue that it retains the authority to pay its employees through historical and codified law.[83] Additionally, the CIA has the statutory authority to control personnel expenses related to travel and transportation costs for employees and their families stationed overseas[84] and to pay for certain medical and physical exams of officers and

78. INTELLIGENCE COMMUNITY DIRECTIVE NO. 650, *supra* note 19, at 2.

79. *Id.* at 1.

80. 50 U.S.C. § 403-1(f)(3)(A)(iii) (2006) (emphasis added).

81. IRTPA, Pub. L. No. 108-458, § 1018, 118 Stat. 3638, 3670 (2004).

82. *See* INTELLIGENCE COMMUNITY DIRECTIVE NO. 650, *supra* note 19, at 6 ("Where applicable, the heads of executive departments and independent agencies with [intelligence community] employees may use their respective authorities to deviate from this ICD when necessary to carry out their independent missions and functions.").

83. *See* 50 U.S.C. § 403j(a)(1) (2006) ("Notwithstanding any other provisions of law, sums made available to the Agency by appropriation or otherwise may be expended for purposes necessary to carry out its functions, including—(1) personal services . . . without regard to limitations on types of persons to be employed . . . ."); *see also* BUS. EXECUTIVES FOR NAT'L SEC., PAY FOR PERFORMANCE AT THE CIA: RESTORING EQUITY, TRANSPARENCY AND ACCOUNTABILITY; THE ASSESSMENT OF THE INDEPENDENT PANEL ON THE CENTRAL INTELLIGENCE AGENCY'S COMPENSATION REFORM PROPOSALS 8 (2004), http://www.bens.org/mis_support/cia-reform-report.pdf ("The Central Intelligence Agency is exempt from certain provisions of Title 5 of the US Code (the federal law governing employment in the civil service) in particular, those provisions concerning compensation and federal employment regulations. . . . [E]ach Agency Senior Manager heads a separate career service and has authority to unilaterally determine salary levels for positions within their span of control with little centralized oversight.").

84. 50 U.S.C. § 403e(a)(1)(A)–(F) (2006).

employees[85] or other allowances and benefits related to "travel, personnel and physical security activities, operational activities, and cover-related activities."[86] Because this authority extends beyond a base salary, it could thus undermine the DNI's intent to normalize salaries across the overseas intelligence community. When coupled with the various loopholes[87] in the IRTPA, the CIA could likely challenge any DNI request to adhere to the NICCP. Absent clear changes to the IRTPA clarifying who retains control over administrative personnel functions, future attempts by the DNI to unify other administrative processes—like agency hiring standards, retirement programs, or employee health care benefits—might also be plagued by the same hypothetical arguments made above. At the same time it is important to note that the DNI implemented ICD 650 without any challenges from members of the intelligence community. The DNI's ability to get these entities to the negotiating table, keeping its authority while still agreeing to abide by the NICCP goals, is a success acknowledging the DNI's power as a centralizing force within the intelligence community, even without the requisite statutory authority on this issue.[88]

3. Freedom of Information Act Requests: Statutory Overlap

Another issue that arose during the IRTPA reorganization of the intelligence community was the ambiguity over who would be in charge of protecting the sources and methods of intelligence concerning FOIA requests. FOIA, enacted in 1966, provides public access to U.S. government records and outlines the responsibilities of agencies carrying out the procedures.[89] FOIA requests make these government records

85. *Id.* § 403e(a)(5)(A)–(D).

86. *Id.* § 403e(b)(2).

87. *See* IRTPA, Pub. L. No. 108-458, § 1018, 118 Stat. 3638, 3670 (2004) (codified at 50 U.S.C. § 403(b)(3)) (directing the DNI to not abrogate the statutory authority of any individual intelligence agency or department); *see also* INTELLIGENCE COMMUNITY DIRECTIVE NO. 650, *supra* note 19, at 6. (clarifying that despite the spirit and intent of collaboration under this directive, the heads of the independent intelligence agencies, under their respective authority may "deviate from this ICD when necessary to carry out their independent missions and functions").

88. *See* Mike McConnell, Dir. of Nat'l Intelligence, Media Briefing on National Intelligence Civilian Compensation Program (NICCP) at the Office of the Director of National Intelligence Headquarters in Washington, D.C. 5 (May 15, 2008), http://www.dni.gov/interviews/20080515_interview.pdf [hereinafter NICCP Briefing] (acknowledging that the intelligence community initially questioned the DNI's authority but recounting that the senior leadership preferred to focus on coming together and establishing a set of policies to "move forward as a community" rather than worry about legal authority).

89. Pub. L. No. 89-487, 80 Stat. 250 (1966) (codified as amended at 5 U.S.C. § 552 (2006)).

available to "any person," unless the agency can show that the requested record contains information outlined in one of the nine statutory exemptions.[90] FOIA legislation outlines the responsibilities of agencies carrying out the procedures.[91] The intelligence community is afforded certain FOIA exemptions for areas of national defense and foreign security.[92]

Under the IRTPA, the DNI was given the authority to "protect intelligence sources and methods from unauthorized disclosure," including "access to and dissemination of intelligence" and "preparation of intelligence products . . . for dissemination."[93] In some ways, the IRTPA language conflicts with the administrative authority granted to the individual intelligence agencies under Executive Order 12,958 to classify and declassify their own information for instances such as FOIA requests.[94] For example, consider when citizens file FOIA requests for information and records from an intelligence agency like the CIA.[95] Although the FOIA request would traditionally go directly to the CIA, IRTPA could be read to give the DNI a role in the CIA's declassification process, adding a layer of review to the request which elongates the time to file a response and generally decentralizing an important administrative process. However, DNI has presently delegated authority back to the agencies through a classified memorandum.[96] This private resolution may indicate the DNI acknowledges that his role is to guide the intelligence community rather than to control all administrative functions—even if the current law does not reflect that understanding.

---

90. *See* 5 U.S.C. § 552(b)(1)–(9) (2006) (exempting, inter alia, documents properly classified as secret for national defense or foreign policy reasons and documents related solely to internal agency personnel rules and practices).

91. *See id.* § 551(1)(A)–(H) (providing the definition of *agency* and allowing the CIA to qualify).

92. *See id.* § 552(b)(1)(A)–(B) (stating that this section does not apply to matters "specifically authorized under criteria established by an Executive order to be kept secret in the interest of national defense or foreign policy" as well as those matters "properly classified pursuant to such Executive order").

93. 50 U.S.C. § 403-1(i) (2006).

94. Exec. Order No. 12,958, 3 C.F.R. 333–34 (1996), *as amended in* 70 Fed. Reg. 21,609 (Apr. 21, 2005) (allowing agencies, upon receiving a FOIA request, to respond with declassification of the information or state a valid exemption).

95. *See* 50 U.S.C. § 431 (2006) (stating that the exemption of operational CIA files is at the discretion of the DCIA and must be done with "the coordination of" the DNI but failing to identify the DNI's role relative to "coordination").

96. The memorandum shifting authority back to the CIA to declassify information in response to FOIA requests is currently classified.

4. Counterterrorism Centers: Statutory Redundancy

One danger of adding another layer of authority to the intelligence community is the possibility of duplicating existing efforts. The 9/11 Commission surmised that the counterterrorism efforts before 9/11 were scattered and resources were spread thin.[97] The CIA had both a Terrorist Threat Integration Center (TTIC) and a Counterterrorism Center (CTC), while the FBI had the Counterterrorist Screening Center.[98] The Commission noted that a "'smart' government would *integrate* all sources of information to see the enemy as a whole."[99] In response to this recommendation, the IRTPA established The National Counterterrorism Center (NCTC) and placed it under the control of the DNI.[100]

Counterterrorism efforts today are still somewhat duplicative. The overlap between the CTC and the NCTC illustrates this point. The CIA's CTC presently coordinates both operational and analytical intelligence efforts, working closely with various United States government agencies and foreign liaisons to disrupt terrorist activities.[101] Meanwhile the DNI's NCTC, by law, is the "primary organization in the United States Government for analyzing and integrating all intelligence passed or acquired by the United States Government pertaining to terrorism and counterterrorism."[102]

It is unclear how much of the NCTC's operations duplicate the CIA's CTC efforts and how much NCTC merely synthesizes the intelligence provided by CIA and other entities.[103] It is also unclear how much collaboration takes place between the entities. On the surface, however,

97. NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 400–01; *see also* RICHARD L. RUSSELL, SHARPENING STRATEGIC INTELLIGENCE: WHY THE CIA GETS IT WRONG, AND WHAT NEEDS TO BE DONE TO GET IT RIGHT 153 (2007) (explaining that the IRTPA creation of the NCTC within the DNI to consolidate counterterrorism assets does not solve the problem of duplication of effort because the CIA's Counterterrorism Center (CTC) still exists).

98. NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 401.

99. *Id.*

100. *See* Exec. Order No. 13,354, 69 Fed. Reg. 53,589 (Sept. 1, 2004) (vesting the DCI with authority over the NCTC), *rescinded by* Exec. Order No. 13,470 (codified at 3 C.F.R. 218, 220–25 (2009)) (including the NCTC in the missions of the DNI).

101. Central Intelligence Agency, Centers in the CIA, https://www.cia.gov/library/publications/additional-publications/the-work-of-a-nation/cia-director-and-principles/centers-in-the-cia.html (last visited May 4, 2010).

102. 50 U.S.C. § 404o(d)(1) (2009) (outlining the primary missions of the National Counterterrorism Center).

103. *See* National Counterterrorism Center, About the National Counterterrorism Center, http://www.nctc.gov/about_us/about_nctc.html (last visited Apr. 27, 2010) (designating the NCTC with a mission of "integrating and analyzing *all* intelligence pertaining to counterterrorism").

the NCTC as established by the IRTPA demonstrates another possible area where congressional intent to integrate interagency counterterrorism efforts has not been achieved in actual practice.[104]

C. Implications

The station chief issue, NICCP ambiguity, FOIA overlap, and CTC–NCTC conflict are just a few examples of how a vague or redundant statutory authority may become troublesome. While some issues, such as the NICCP implementation, have progressed without objection from the CIA, they still demonstrate an inherent problem with the current distribution of administrative authority within the United States intelligence community. These IRTPA statutory loopholes, ambiguities, and administrative overlaps point out flaws in the enabling legislation, providing us with an opportunity to reevaluate the respective roles of the DNI and the DCIA.

The specific instances of agency overlap and statutory ambiguity discussed above provide glimpses into larger organizational problems. Ambiguity over the appointment of overseas station chiefs affects the immediate task of filling overseas intelligence posts, but perhaps more importantly, the level of publicity around this issue demonstrates that the DNI has added to the bureaucratic tensions that the IRTPA was supposed to break down.[105] Likewise, the NICCP directive and the FOIA conflict demonstrate that the statutory overlap becomes an issue beyond bureaucratic tensions and affects the daily administrative tasks of the various intelligence community agencies.[106] Implementing a common pay system throughout the community has already taken considerable time and resources that might have been used elsewhere.[107] Looking beyond purely administrative burdens, IRTPA provisions that do not take into account preexisting infrastructure jeopardize the operational effectiveness of national intelligence efforts. The operation of the NCTC—an organization

104. The 9/11 Commission Report recommendations, which are reflected in the statutory language of the IRTPA, call for the NCTC to "absorb a significant portion of the analytical talent now residing in the CIA's Counterterrorist Center." NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 404.

105. Ignatius, *supra* note 16.

106. *See* discussion *supra* Parts II.B.2, II.B.3.

107. Although there is little public information on the amount of resources used to evaluate the NICCP, DNI Mike McConnell said of the evaluation process,

> We looked at how it would fit across all of the community, and we worked through all of that in a coordinated way, probably taking a little more time than we should have, we would like to, but we're at a point now where we signed off on this and we're going to put this in action.

NICCP Briefing, *supra* note 88, at 2.

that parallels the counterterrorism mission of the well-established CIA CTC—has spread thin already-limited intelligence-community resources rather than integrating and unifying them.[108]

The resolution of each of these conflicts of authority should result in solutions that mend any tension between the DNI and the rest of the intelligence community. However, as evidenced by the current struggles between the DNI and the CIA to determine proper authority for station chief appointment, personnel, FOIA, and counterterrorism center issues, many potential conflicts remain unresolved. Ultimately, legislation should provide the DNI with a succinct scope of authority and clear power to implement that authority without loopholes, ambiguity, or overlap.

III. PROPOSED AMENDMENTS AND RECOMMENDATIONS

When Congress established the DNI position, many critics felt that it would be no more successful at centralization than the now-defunct DCI, merely adding another layer of bureaucracy over an arguably already-stove-piped system.[109] Given the amount of effort and development in building the DNI—the employee count is now well over one thousand[110]—it would be ineffective to argue that the right solution would be merely to undo its creation. Before advancing solutions, it is important to look back to the intent of Congress in framing the DNI's authority under the IRTPA.

A. Remembering Congressional Intent: Administrative Versus Operational

One of Congress's chief goals in creating the DNI was to alleviate some of the pressure on the DCI, who up until then had acted as the President's intelligence advisor, coordinated the entire intelligence community, and headed the CIA.[111] The principle responsibility of the DNI set forth by IRTPA is to "oversee and direct the implementation" of the National Intelligence Program, signaling an advisory and policy role rather than an

108. *See supra* note 97 and accompanying text.

109. *See* RICHARD A. POSNER, THE REORGANIZED U.S. INTELLIGENCE SYSTEM AFTER ONE YEAR 3 (2006) http://www.aei.org/docLib/20060411_SENSOg.pdf (arguing that the Office of the DNI "has become a new bureaucracy layered on top of the intelligence community" and that merely adding one more intelligence asset to the fifteen that already exist as part of the national intelligence community does not reorganize the intelligence community into the effective body that was envisioned by the IRTPA).

110. *See* Ignatius, *supra* note 16 (indicating that the DNI presently has at least 1,500 employees).

111. *See* NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 409 (discussing the previous burden of responsibilities that the DCI had in his capacity as head of the CIA, leader and manager of the intelligence community at large, and intelligence advisor to the President); *see also* BEST & CUMMING, *supra* note 40, at 1 (describing the primary responsibilities of the DCI).

operational one.[112] Congress's intent to separate the administrative policy and oversight aspects of intelligence coordination from the operational acts of intelligence gathering is implicit in the express prohibition against a current DNI serving concomitantly as the DCIA.[113] Additionally, the wording of statutory authority in the appointment of station chiefs indicates Congress intended the DNI to have more of an administrative, policy-based role. As mentioned before, the DNI is given the authority to "oversee" the appointment while the CIA's statutory authority is to actually "coordinate" the appointments.[114] Therefore, when attempting to rectify these statutory ambiguities and inconsistencies, a solution should reflect the congressional intent while accommodating some of the strengths of the pre-IRTPA structure.[115]

B. General Reorganization Strategies

Critics of the IRTPA reorganization, both past and present, offer various solutions to the legislation. Some draw from the legislative intent, while some use historical reorganizations as a model. However, both camps acknowledge that the intelligence reorganization is not complete. Various suggestions have been made as to the role the DNI should play within the national intelligence infrastructure. The following sections discuss several options.

1. The DNI as an Intelligence Czar

One of the more popular suggestions made to rectify the various IRTPA criticisms calls on Congress to empower the DNI to exercise greater control over the intelligence community.[116] The idea of enhancing the power of

112. IRTPA, Pub. L. No. 108-458, § 1011(a), 118 Stat. 3638, 3643–62 (codified at 50 U.S.C. § 403(b)(3) (2006)).

113. IRTPA § 1011(a) (codified at 50 U.S.C. § 403(c) (2006)) ("The individual serving in the position of Director of National Intelligence shall not, while so serving, also serve as the Director of the Central Intelligence Agency or as the head of any other element of the intelligence community.").

114. The word *oversee* denotes a supervisory or administrative role, while the task of actual coordination denotes an operational role. *See supra* note 74 and accompanying text.

115. For example, even though FOIA requests may be considered administrative tasks that would, under congressional intent, be delegated to the DNI, the DNI conceded that the system would be more efficient if the agencies continued to process FOIA requests independently. *See supra* note 96 and accompanying text.

116. This view of increased DNI authority has been extolled by members of Congress, members of the intelligence community, legal commentators at large, and even the DNI himself. *See* S. COMM. ON INTELLIGENCE, INTELLIGENCE AUTHORIZATION ACT FOR FISCAL YEAR 2010, S. REP. NO. 111-55 (1st Sess. 2009), *available at* http://fas.org/irp/congress/2009_rpt/srpt111-55.pdf (setting forth in Title III, Subtitle A,

the DNI to improve his effectiveness was first introduced in the 9/11 Commission Report. The Commission set forth various perceived intelligence community failures, such as a lack of channels for cooperation and information sharing that would have allowed the community to predict the imminence of a 9/11-style attack.[117] A powerful National Intelligence Director would presumably be this centralizing figure, overseeing all intelligence and counterterrorism collection efforts of the CIA, the FBI, and the Department of Defense.[118]

Although the 9/11 Commission Report and the legislative history of the IRTPA suggest the intent was to provide the DNI with administrative authority over the entire intelligence community, that solution has not been completely feasible, as shown by the FOIA classification issue[119] or the pay-for-performance directive.[120] The issue with this centralization of power is that with so many administrative and operational activities and missions, the sixteen individual intelligence entities are better left with current entity leadership like the DCIA, who has a better understanding of the daily budgetary and personnel needs, and not with the DNI, who has spent the first several years of his new position playing catch-up.[121]

---

and Title IV, Subtitle A, the enhanced authority sought for the DNI, such as authority to conduct accountability reviews of the various intelligence community entities, authority to use funding for information access and sharing across the community, the ability to approve interagency financing of boards, commissions, and councils, as well as providing the Office of the DNI with several new positions such as chief information officers, an enhanced inspector general, and a chief financial officer); Kaplan, *supra* note 13 (noting that the DNI's authority under the IRTPA looks "nothing like the locus of decision-making and responsibility that the 9/11 commission had in mind"); *McConnell Hearing*, *supra* note 64, at 1–2 (voicing the same concerns as DNI McConnell, Senator Rockefeller stated that "[s]ome of us worry that Congress may not have given the DNI enough authority to match his enormous responsibilities," with Senator Christopher S. Bond adding that that IRTPA denied the DNI "the full authorities required truly to direct the intelligence community, not just coordinate its activities").

117. *See* NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 357 (detailing CIA efforts to remain vigilant despite a lull in terrorist activity abroad, even though the warning was unavailable to or disregarded by intelligence entities beyond the CIA).

118. *Id.* at 411.

119. *See* 50 U.S.C. § 403-1(i)(C) (2006) (providing the DNI with exclusive authority for the "preparation of intelligence products . . . for dissemination"). *But see* Exec. Order No. 12,958, 3 C.F.R. 333 (1995), *as amended in* 70 Fed. Reg. 21,609 (Apr. 26, 2005) (allowing agencies, upon receiving a FOIA request, to respond with declassified information or a valid exemption).

120. *See supra* notes 82–87 and accompanying text.

121. The Office of the DNI is still playing catch-up within some of its own administrative functions. *See* Dennis Blair, Dir. of Nat'l Intelligence, Statement for the Record by the Director of National Intelligence Before the S. Select Committee on Intelligence on the Intelligence Authorization Proposal for FY10, p. 1–2 (May 19, 2009) [hereinafter Blair Statement], http://intelligence.senate.gov/090617/proposals.pdf

2. The DNI as an Intelligence Advisor and Policymaker

While the most popular of the publicly offered solutions to the intelligence community reorganization is to give the DNI more power and control over the intelligence agencies, the position could be more effective with more of a policy and oversight role.[122] In light of the need for increased collaboration and communication, the DNI position could focus on just that—coordinating all the various entities just as the Secretary of Defense coordinates the various military branches. A DNI with oversight power would not be an "intelligence czar" with absolute power over all the entities but would be in charge of budgetary issues, threat estimates, and other community-wide policy decisions.[123]

Focusing the DNI's authority on administrative coordination and oversight would leave the operational component of intelligence gathering to the various intelligence community entities, with the CIA director as the President's advisor for operational efforts. The CIA's preeminence in collection and analysis of human intelligence—and its long history of being a customer of the other intelligence entities like the National Security Agency and the National Reconnaissance Office[124]—position it to facilitate operational coordination of intelligence issues.

Meanwhile, the DNI should have the power to centralize administrative and policy issues through the use of committees comprised of representatives from each agency tasked with controlling issues like budget and personnel. Specifying the administrative tasks over which the DNI has undisputed authority may fix the statutory ambiguity and overlap problems exemplified by the NICCP, FOIA, and NCTC issues without disrupting operational intelligence gathering and analysis.

(requesting source and method protections under FOIA that are equal to those explicitly provided to the CIA for operational file exemptions, as well as similar exemptions under the Privacy Act "akin to [exemptions] enjoyed by the Director of the Central Intelligence Agency" and Federal Advisory Committee Act (FACA) exemptions "identical to the exemption the Director of the CIA has").

122. *See generally* POSNER, *supra* note 109, at 2–3 (acknowledging that while the effect of the IRTPA reorganization was founded on the idea of creating a DNI with the capabilities of being an "administrator," "czar," or "presiding deity" of the intelligence community bureaucracy, the DNI should not be given absolute authority but rather a role of "coordinator" or "board chairman").

123. *Id.*

124. The CIA has been a long-standing "customer" of the NRO and the NSA, two of the largest intelligence collection and analysis agencies within the Department of Defense. *See* Welcome to the NRO, http://www.nro.gov/ (last visited Apr. 27, 2010); National Security Agency, About NSA, http://www.nsa.gov/about/index.shtml (last visited Apr. 27, 2010).

3. Goldwater–Nichols Act as a Model

Another suggestion is to model the intelligence community after the Department of Defense following the Goldwater–Nichols Reorganization Act of 1986.[125] Congressional concerns in 1986 about the need to unify the military seem to parallel the 2004 concerns about the need to unify the intelligence community—both relate to improving communication and coordination among the individual entities.[126]

Both the military and the intelligence community are comprised of specialized branches. In the military, the branches are divided essentially by function: the Army occupies the land, the Navy operates in the oceans, and the Air Force concentrates on the air operations.[127] Likewise, members of the intelligence community often contribute specialty services to the national intelligence effort.[128] And as it became important that the various military services be able to unify their specific capabilities on the

125. Goldwater–Nichols Department of Defense Reorganization Act of 1986, Pub. L. No. 99-433, 100 Stat. 992 (codified as amended in scattered sections of 10 U.S.C.); *see* NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 408–09 (highlighting the structural barriers of the intelligence community prior to 9/11 by contrasting its organization with that of the United States Armed Services after the Goldwater–Nichols reorganization in 1946, which focused on creating joint commands based on field operations and not capabilities or type of service); *see also* 150 CONG. REC. S9555 (daily ed. Sept. 23, 2004) (statement of Sen. Graham) ("The key to this mission-based decentralization of intelligence, in my opinion, is that we must give the Director of National Intelligence the statutory authority to manage the community with flexibility and nimbleness so he or she can quickly establish new centers or modify existing centers as future threats emerge, just as Goldwater-Nichols has given that authority to the Secretary of Defense."); *id.* at S9556 (citing Flynt Leverett, *Force Spies to Work Together*, N.Y. TIMES, July 9, 2004, at A19 ("We need to develop a model of 'jointness' for the intelligence community, analogous to what the Goldwater-Nichols Act did for the uniformed military 18 years ago.")).

126. *See* STAFF OF S. COMM. ON ARMED SERVICES, 99TH CONG., REPORT ON DEFENSE ORGANIZATION: THE NEED FOR CHANGE 86 (Comm. Print 1985) (setting forth perceived problems with the Department of Defense's ability to cooperate and work effectively in light of the technological changes, the changing international political landscape, and the changing demands of protecting U.S. security interests); *cf.* NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 399 (setting forth the intelligence issues prior to 9/11 of entities that were facing technological challenges in the face of a new enemy much different than the enemy of the Cold War).

127. This is an oversimplified representation of military capabilities used merely to illustrate a possible solution to IRTPA's perceived shortcomings. Both the armed services and members of the intelligence community provide overlapping capabilities that further complicate the integration process but will not be discussed at length here.

128. *See* NATIONAL INTELLIGENCE: A CONSUMER'S GUIDE 32–42 (2009), http://www.dni.gov/IC_Consumers_Guide_2009.pdf (providing an overview of special intelligence capabilities, such as the CIA providing human intelligence efforts, the DIA providing intelligence on foreign military capabilities, the National Geospatial-Intelligence Agency providing geospatial intelligence, the NSA being responsible for signals intelligence, and the National Reconnaissance Office providing space reconnaissance via satellites).

battlefield, as evidenced in the Goldwater–Nichols legislation, the same desire to unify and coordinate intelligence capabilities is seen in the IRTPA legislation.[129] Essentially, IRTPA could provide the DNI with the same oversight and advisory role over the intelligence community that the Chairman of the Joint Chiefs of Staff has over the armed services.[130] Further, the DNI would relinquish operational control over intelligence capabilities to the DCIA and the heads of other intelligence entities, who would act in the operational capacity similar to a combatant commander.[131]

The most apparent problem with applying the Goldwater–Nichols model to the intelligence community is that most intelligence entities are already part of another cabinet-level department and thus already report to a cabinet-level official.[132] The reorganization may not be as successful as the original Goldwater–Nichols reorganization without supplanting the sixteen intelligence agencies from their current cabinet departments into a new one. Several independent panels suggested such recommendations in 2001, but the Bush Administration never adopted them.[133]

---

129. *See supra* note 126.

130. *See* 10 U.S.C. § 151(b)(1) (2006) (stating that the Chairman of the Joint Chiefs of Staff shall serve as the "principal military adviser to the President"); *see also* § 153(a)(1)–(6) (stating that the Chairman shall be responsible for providing strategic direction, strategic planning, training and education policies, and advising on programs and budgets of the armed forces while deferring to the combatant commanders for recommendations on operational capabilities and assessments).

131. *See* DEPARTMENT OF DEFENSE DIRECTIVE 5100.1: FUNCTIONS OF THE DEPARTMENT OF DEFENSE AND ITS MAJOR COMPONENTS (2002) http://www.dtic.mil/whs/directives/corres/pdf/510001p.pdf (providing the roles and responsibilities of each of the Department of Defense entities, including the relationship of authority between the Secretary of Defense and the Chairman of the Joint Chiefs of Staff); *see also* DEPARTMENT OF DEFENSE DIRECTIVE 5158.1: ORGANIZATION OF THE JOINT CHIEFS OF STAFF AND RELATIONSHIPS WITH THE OFFICE OF THE SECRETARY OF DEFENSE (1985), http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA272367&Location=U2&doc=GetTRDoc.pdf (setting forth "policies, procedures, and organizational relationships" necessary to accomplish the reorganization of the "defense establishment").

132. *See* Members of the Intelligence Community, *supra* note 53 (stating that all intelligence entities except the CIA fall under a cabinet-level official).

133. *See* Walter Pincus, *Intelligence Shakeup Would Boost CIA: Panel Urges Transfer of NSA, Satellites, Imagery from Pentagon*, WASH. POST, Nov. 8, 2001, at A1 (recommending that the NRO, NSA, and NGA should be removed from the Department of Defense and placed under the control of the CIA); *see also* POSNER, *supra* note 109, at 6 (noting that a commission headed by Brent Scowcroft, the Chairman of President George W. Bush's Foreign Intelligence Advisory Board, suggested that the Department of Defense's disproportionate control over the intelligence budget could be offset if the intelligence agencies within the Department of Defense were removed and placed under the control of the DNI).

4. Recommendations that Should Be Adopted

Each of these recommendations have merit and backing from DNI supporters and critics alike. The first task is to remove barriers imposed by having intelligence community entities under the direct control of cabinet-level departments outside the DNI structure.[134] Richard Posner, a judge on the U.S. Court of Appeals for the Seventh Circuit and a respected commentator on the intelligence reorganization efforts, refers to this as the "twin stars problem."[135] One of his suggestions, which is adopted as a recommendation here, is to pull the large intelligence-gathering agencies out of the Department of Defense and align them under the direct control of the DNI.[136]

Instead of making the DNI an "intelligence czar," as was often the suggestion following the 9/11 Commission Report recommendations, the operational capabilities as well as the daily administrative functions—responding to FOIA requests and implementing payment, retirement, and benefit packages—should remain with the individual entities. While it is easy to understand the value of keeping operational capabilities with the individual entities who have developed and dominated their fields in both knowledge and resources, the benefit of leaving certain administrative duties to the individual entities is that it allows the DNI time to adjust his own FOIA and personnel issues before taking on sixteen others.[137] This would leave the DNI with the larger policy and oversight issues of running an intelligence community.

If focusing the individual intelligence entities on operational intelligence gathering and analysis were a central goal of the IRTPA legislation, placing the NCTC under the DNI—an administrative coordination and oversight organ—does not necessarily adhere to that goal.[138] Rather, the NCTC should absorb all of the individualized counterterrorism efforts and be placed under the control of an operational agency like the CIA.[139]

Lastly, as each of these examples demonstrates, the statutory language of

134. *See supra* note 132.

135. *See* POSNER, *supra* note 109, at 6 (defining the "twin stars" problem as "the secretary of defense and the director of national intelligence circling warily around each other").

136. *See id.* (suggesting that removing the larger intelligence assets from the Department of Defense will do more than alleviate financial issues, as it will resolve some of the cultural clashes among military and nonmilitary intelligence activities).

137. *See* Blair Statement, *supra* note 121, at 1 (documenting the DNI's request for FOIA authority comparable to that currently afforded to the CIA).

138. *See supra* notes 97–104 and accompanying text (noting that the various counterterrorism programs are predominantly operational, having been established by the FBI and CIA).

139. *See* POSNER, *supra* note 109, at 3 ("[K]eep the analysts close to the operations officers.").

the IRPTA is not always clear and there is currently no established system to interpret inconsistencies.[140] As suggested by the 9/11 Commission Report prior to the drafting of the IRTPA, the intelligence community must have a formal channel to discuss inconsistencies and ambiguities with the heads of the intelligence entities and resolve disputes before they become national headlines.[141]

CONCLUSION

Before the enactment of the IRTPA, the DCI position entailed three jobs but lacked statutory authority to perform all of them efficiently.[142] While the IRTPA contributes important changes to the intelligence landscape and creates a new structure that holds the promise of marked improvements in communications and asset sharing among intelligence entities, the aforementioned statutory ambiguities and overlap indicate there is still room for improvement within the current IRTPA legislation.

Further DNI reorganization, modeled loosely after the Goldwater–Nichols Act, should remove the various intelligence entities from their cabinet-level shields. In doing so, the DNI should be tasked with coordinating intelligence community policies, controlling the overall intelligence budget, and setting priorities for their operational activities. Daily administrative tasks such as FOIA requests and employee compensation should remain with the individual intelligence agencies, along with all operational tasks of collecting and analyzing intelligence. A dispute system should be created to ensure that any statutory ambiguity could be addressed efficiently and privately. Any resolutions should be documented and should guide the refinement of the IRTPA legislation to prevent future disputes over similar statutory ambiguities.

Hopefully, future changes to the current IRTPA legislation will be more than cosmetic, helping our vital national security assets accomplish their administrative duties without undue friction or confusion. IRTPA brings the nation closer than it has ever been to having a unified U.S. intelligence community helping to protect our nation from any future threats.

---

140. *See supra* notes 71–74 and accompanying text.

141. This recommendation existed in the 9/11 Commission Report but was not adopted by the IRPTA legislation. *See* NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., *supra* note 38, at 414 ("Too many agencies now have an opportunity to say no to change. The National Intelligence Director should participate in an NSC executive committee that can resolve differences in priorities among the agencies and bring the major disputes to the president for decision.").

142. *See* CIA ORIGIN AND EVOLUTION REPORT, *supra* note 8, at 7 (arguing that the CIA was given contradictory mandates by being responsible for coordinating intelligence efforts across the community but not having the ability to control intelligence assets, rendering the DCI's job practically impossible).

RECENT DEVELOPMENTS

POLITICS, RULEMAKING, AND JUDICIAL REVIEW: A RESPONSE TO PROFESSOR WATTS

ENRIQUE ARMIJO*

TABLE OF CONTENTS

In an article recently published by *The Yale Law Journal* titled *Proposing a Place for Politics in Arbitrary and Capricious Review*,[1] Professor Kathryn A. Watts argues for a more robust role for politics in agencies' informal rulemaking procedures under the Administrative Procedure Act (APA),[2] as well as in arbitrary and capricious judicial review of those rules.[3] However, her proposal ignores the primary function the APA envisions for the views of regulated entities in notice-and-comment rulemaking. She also minimizes the ways in which politics sets regulatory policy before an agency commences the process of adopting, rescinding, or defending a rule. In so doing, she overvalues and undervalues the effect of politics on the agency

* Associate, Covington & Burling LLP, Washington, DC. The views expressed in this Article are mine alone. Thanks to Professor Kathryn Watts for writing an important, accessible, and thought-provoking article.

1. Kathryn A. Watts, *Proposing a Place for Politics in Arbitrary and Capricious Review*, 119 YALE L.J. 2 (2009), *available at* http://www.yalelawjournal.org/images/pdfs/824.pdf.

2. 5 U.S.C. § 553 (2006) (listing the procedural requirements of informal rulemaking).

3. 5 U.S.C. § 706(2)(A) (2006) (allowing courts to overrule agency action that is arbitrary and capricious).

rulemaking process at the same time.

In Part I of this Recent Development, I discuss why the APA intentionally insulated agency rulemaking from the political branches, show that this insulation predated "hard look" review, and demonstrate why, from the perspective of regulated entities, this needs to be so. In Part II, I consider how politics sets the regulatory agenda for the Federal Communications Commission (FCC), the agency I am most familiar with as a practitioner. In Part III, I examine the Supreme Court's decision last Term in *FCC v. Fox Television Stations, Inc.*[4] and find the case is less of an invitation for agencies and courts to rely on political influence than Professor Watts believes.

I. THE APA'S SEPARATION OF POLITICS FROM RULEMAKING

The APA "established the fundamental relationship between regulatory agencies and those whom they regulate—between government, on the one hand, and private citizens, business, and the economy, on the other hand."[5] The history of its passage shows that in order for this relationship to be a balanced one, rulemaking had to be a facts-driven process. And Congress intended the entities subject to and affected by an agency's rules to be the wellsprings for those facts.

New Deal politics permeated the 1940s-era debates that led to the APA's adoption,[6] but those debates made clear that politics should be absent from agencies' procedures for adopting generally applicable regulations. The template for administrative procedure reform, and for what became the APA, was the 1941 Attorney General Committee's Report on Administrative Procedure (Final Report).[7] No less an administrative law authority than Kenneth Culp Davis noted that the Attorney General Committee's Final Report—in particular one of the draft bills attached to the Final Report—was the basis for notice-and-comment rulemaking.[8] The Final Report argued that rulemaking procedures should insulate agencies, and by extension the regulated entities subject to those agencies' purview, from politics:

---

4. 129 S. Ct. 1800 (2009).

5. George B. Shepherd, *Fierce Compromise: The Administrative Procedure Act Emerges from New Deal Politics*, 90 NW. U. L. REV. 1557, 1558 (1996).

6. *See id.* at 1595 (discussing the emergence of the Administrative Procedure Act (APA)).

7. FINAL REPORT OF THE ATTORNEY GENERAL'S COMMITTEE ON ADMINISTRATIVE PROCEDURE (1941) [hereinafter FINAL REPORT], http://www.law.fsu.edu/library/admin/pdfdownload/apa1941.pdf.

8. K.C. Davis & Walter Gellhorn, *Present at the Creation: Regulatory Reform Before 1946*, 38 ADMIN. L. REV. 511, 520 (1986).

> An administrative agency, [unlike a legislature,] is not ordinarily a representative body. Its function is not to ascertain and register its will. . . . [I]ts members are not subject to direct political controls as are legislators. It investigates and makes discretionary choices within its field of specialization. The reason for its existence is that it is expected to bring to its task greater familiarity with the subject than legislators, dealing with many subjects, can have. But its knowledge is rarely complete, and it must always learn the frequently clashing viewpoints of those whom its regulations will affect.
>
> These differences are and should be reflected in its procedures, which should be adapted to giving adequate opportunity to all persons affected to present their views, the facts within their knowledge, and the dangers and benefits of alternative courses.[9]

Professor Watts claims that the emphasis on data over politics in agency procedure is a product of the court-made hard look doctrine,[10] but as the Final Report demonstrates, a data-driven rulemaking process predates even the APA, let alone hard look review. Pre-APA agency-specific procedural statutes compelled agencies to base their decisions upon evidence presented by regulated entities.[11] The Food and Drug Administration's and the Wage and Hour Division's enabling statutes required findings of fact to support any regulations the agencies imposed, and those findings had to be based exclusively on evidence put before the agency.[12] And as noted, the Final Report argued that rules could be legitimate only if the agencies promulgating them took serious account of input from regulated entities.[13] Participation in an agency's rulemaking procedure by "those upon whom [an agency's] authority bore" was considered "essential in order to permit administrative agencies to inform themselves and to afford adequate safeguards to private interests."[14]

A careful reader of Professor Watts's article might respond that hers is not an argument to allow agencies to go whole hog in relying on political influence over data when making rules. Rather, her claim is that courts should be more tolerant of agencies' reliance on record evidence of that influence, and that political considerations should play the same kind of role as data provided by parties potentially affected by a rule.[15] In other

---

9. FINAL REPORT, *supra* note 7, at 101–02.

10. Watts, *supra* note 1, at 16 (describing the development of hard look review by the D.C. Circuit).

11. *See* FINAL REPORT, *supra* note 7, at 106.

12. *See id.* at 109 (citing the Federal Food, Drug, and Cosmetic Act of 1938, 21 U.S.C. § 371(e) (1940), and Fair Labor Standards Act of 1938, 29 U.S.C. §§ 208, 210(a) (1940)).

13. *Id.* at 101–02.

14. *Id.* at 103.

15. *See* Watts, *supra* note 1, at 73 ("[I]f . . . the evidence would equally support the selection of either Rule A, B, *or* C, then it would be entirely rational for the agency to rely upon political influences in explaining why it chose Rule C over Rules A or B.").

words, in the rulemaking context both the agency and reviewing court should essentially treat the politician or political body as a commenter, not a policymaker. But this is not a meaningful distinction. Agencies are expert in their areas of delegated authority, not in assessing the strength and direction of political winds. And the notion that an agency would give political "evidence" a weight comparable to empirical evidence in a rulemaking—that political influence would be used only as a "tiebreaker" when an agency record would support several proposed courses of action[16]—is specious. If a client considered spending upwards of several thousand dollars to commission a feasibility analysis or white paper in support of its position in a rulemaking proceeding, and that position could be trumped or even canceled out at the agency level and upon judicial review by a simple "reference to the President's clearly expressed executive priorities"[17] or to "a group of congressmen's comments on the substance of a proposed rule,"[18] any decent administrative law attorney would have to consider advising the client to save its money—or to spend it on lobbying.[19]

So the reign of the technocrats that Professor Watts laments is not a product of courts' coarsening of arbitrary and capricious analysis via hard look review; rather, it is entirely consistent with the agency independence and fairness rationales underlying the APA. But politics' absence from rulemaking *procedure* is by no means a sign that it is also absent from rulemaking *policy*. Two recent proceedings before the FCC demonstrate this point.

II. HOW POLITICS SETS THE RULEMAKING AGENDA: TWO EXAMPLES

Agencies do not make or defend rules on a blank slate. Professor Watts accuses agencies of "failing to disclose or affirmatively hiding political

16. *Id.* at 82.

17. *Id.* at 58.

18. *Id.* at 65.

19. Professor Watts seems confident that reviewing courts could distinguish political influence that "maximize[d] the public good" from "backdoor political tactics." *Id.* at 82–84. I am not sure that this is as simple a task as she sets out. Professor Watts hypothesizes that under her conception of arbitrary and capricious review, an Environmental Protection Agency (EPA) rule that was justified as "consistent with the President's foreign policy initiatives" on global warming would likely survive judicial review, while a Food and Drug Administration (FDA) that supports a final rule by stating "[t]he President directed us to rescind the preemption regulations in order to reward the trial lawyers" would suffer a remand. *Id.* at 54, 56. The general counsel of an FDA that allowed such a statement to sneak into his agency's final order would probably be fired. To support her claim that judicial review could distinguish "public good"-maximizing political influence from "backdoor political tactics," Professor Watts has staffed her hypothetical FDA with straw men.

influences that factor into the mix,"[20] but agencies often appeal to politics when adopting or changing course. And when they do leave politics out of the mix, it is because the APA compels them to do so.

For example, the FCC is required to review its structural media ownership rules every four years to ensure they remain in the public interest.[21] After its most recent review, more than a dozen parties challenged the Commission's decision to change one rule and retain others as arbitrary and capricious (public interest groups challenged the changes as too deregulatory, while media parties claimed the changes did not go far enough) in the U.S. Court of Appeals for the Third Circuit.[22] The proceeding has been circuitous to say the least. Most relevant for the present discussion, however, was the FCC's request to the Third Circuit that any judicial review of the rules be stayed because the administration, and the FCC's makeup, had changed; the rules as adopted therefore no longer "reflect[ed] the views of a majority of the current members of the Commission."[23] In other words, the politics changed.[24] Nothing in the *rulemaking record* had changed, of course; the proceeding was closed, the rules had already been promulgated, and petitions for review had already been filed with the court. So to claim, as Professor Watts does, that agencies "sweep political influences under the rug"[25] when making rules, or even when defending them upon judicial review, is to tell a selective story.

Similarly, and at the front end of the rulemaking process rather than the back, the FCC recently published a notice of proposed rulemaking regarding net neutrality, or the general principle that Internet access providers should be barred from discriminating among the content or

---

20. *Id.* at 23.

21. 47 U.S.C. § 303 note (2006) ("The Commission shall review . . . all of its ownership rules quadrennially . . . . The Commission shall repeal or modify any regulation it determines to be no longer in the public interest.").

22. *See, e.g.*, Petition for Review at 1, Prometheus Radio Project v. FCC, No. 08-3078 (3d Cir. July 15, 2009). I represent one of the parties in this proceeding.

23. Letter from P. Michele Ellison, Acting Gen. Counsel, FCC, to Marcia M. Waldron, Clerk, U.S. Court of Appeals for the Third Circuit (May 5, 2009) (on file with author); *see also* Status Report of the Federal Communications at 3, Prometheus Radio Project v. FCC, No. 08-3078 (3d Cir. Oct. 1, 2009) (the rule under review, "[*o*]*f necessity*, . . . does not incorporate" the views of post-election appointed commissioners (emphasis added)). Indeed, one Commissioner who was in the majority when the rules under review were promulgated filed his own letter, stating he disagreed with the "alter[ation] of the agency's litigation procedural posture." Letter from Robert M. McDowell, Comm'r, FCC, to Clerk of the Court, U.S. Court of Appeals for the Third Circuit (April 3, 2009), http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-289974A1.pdf.

24. The Third Circuit rejected the FCC's argument, and the agency must soon defend the rules adopted by the previous majority. *See* Order, Prometheus Radio Project v. FCC, No. 08-3078 (3d Cir. March 23, 2010).

25. Watts, *supra* note 1, at 29.

applications accessed by their users.[26] The questions concerning the proposed rules that the agency asked of potentially affected parties were highly technical and data driven in nature.[27] But the principle of an open Internet was a primary pillar in then-candidate Obama's technology platform.[28] It was therefore only natural, indeed self-evident, that his FCC would seek to put this policy into effect via informal rulemaking. However, the fact that the agency's notice did not attribute the policy to the new President was no obfuscation. Rather, it failed to do so because administrative law disconnects policy from procedure as a matter of fairness to affected parties.[29]

The requirement that net neutrality policy must still go through the "technocratic" wringer demanded by the APA, and that the FCC or any other agency must justify any eventual rule's adoption to a reviewing court without regard to the politics from which it was birthed, insulates affected parties from the political drivers of agency decisions. This procedural requirement serves a number of due process values that Professor Watts ignores. Politics-free process and politics-free arbitrary and capricious review ensure regulated entities, to the greatest degree possible, that their positions and concerns regarding a rule-related course of action will be heard and addressed by the agency. Regulated entities participate in rulemakings—even when the political deck is stacked against them—

---

26. Preserving the Open Internet, Broadband Industry Practices, 74 Fed. Reg. 62,638 (proposed Nov. 30, 2009) (to be codified at 47 C.F.R. pt. 8). The notice of proposed rulemaking and other documents can also be found at the proceeding's home page, http://www.openinternet.gov.

27. *See, e.g.*, *id.* at 62,640–43 (requesting comment on how to "promote and protect the legitimate business needs of broadband Internet access service providers"; how to "defin[e] the scope" of Internet service-providing "entities covered by our proposals"; and on the "effects of . . . technolog[y] on the content, applications, and services being provided—or capable of being provided—over the Internet"; further, seeking "qualitative or quantitative evidence and analysis" and "specific examples" illuminating, inter alia, "economic theory" on "benefits [that] can arise from price and quality discrimination").

28. *See* Obama–Biden Technology Agenda, http://change.gov/agenda/technology_agenda/ (last visited Jan. 29, 2010) (supporting "the principle of network neutrality to preserve the benefits of open competition on the Internet"); John Eggerton, *Obama Makes Network-Neutrality Pledge*, BROADCASTING & CABLE, Oct. 29, 2007, http://www.broadcastingcable.com/article/110976-Obama_Makes_Network_Neutrality_Pledge.php.

29. The D.C. Circuit recently held that the FCC exceeded its jurisdictional authority when it FCC cited Comcast for the company's management of Internet traffic. The decision, Comcast v. FCC, No. 08-1291 (D.C. Cir. Apr. 6, 2010), has obvious implications for the Administration's policy goals in this area. However, I do not believe Professor Watts's argument could be extended to grant agencies politics-based deference when the agency asserts jurisdiction over areas where a political priority has been expressly stated since an agency's interpretation of its jurisdiction is a legal question rather than an evidentiary one.

precisely because the APA contemplates agency procedures and judicial review that are evidentiary, not political. Regulated entities also participate in notice-and-comment rulemaking to develop an appellate record in the event the agency promulgates a rule contrary to their interest or position.[30] And their increased participation leads to better rules.

III. *FCC V. FOX TELEVISION STATIONS, INC.*

Upon an initial read, Justice Scalia's opinion in *Fox* sent a tiny shiver down the collective spine of the communications bar (or at least the spine of my practice group when we discussed the decision over lunch). But the opening it creates for courts to consider politics when engaged in arbitrary and capricious review of agency rulemakings is smaller than Professor Watts makes it appear to be.

Fox involved an arbitrary and capricious review of the FCC's ratcheting up of its indecency policy to find isolated utterances of the F-word and S-word indecent. The FCC took two steps in defending its new policy that met the Court's satisfaction: (1) it "forthrightly acknowledged that its recent actions have broken new ground" (in that previously it did not find isolated utterances to be indecent) and "explicitly disavow[ed]" its prior inconsistent decisions as "no longer good law"; and (2) it gave "reasons for expanding the scope of its enforcement activity [that] were entirely rational."[31]

To be sure, *Fox* might add more branches to an agency's decision tree, but not to allow it to openly consider politics during rulemakings. First, the portion of Justice Scalia's opinion joined by the full Court analyzed the agency's reasoning in changing its indecency *policy*, not on its lack of reliance on record data in changing an agency *rule*.[32] Justice Scalia's arbitrary and capricious analysis noted that the FCC's indecency policy was an adjudication, not a rulemaking, stating, "there is no basis for

30. *See* Advocates for Highway & Auto Safety v. Fed. Motor Carrier Safety Admin., 429 F.3d 1136, 1150 (D.C. Cir. 2005) ("[A] party will normally forfeit an opportunity to challenge an agency rulemaking on a ground that was not first presented to the agency for its initial consideration.").

31. FCC v. Fox Television Stations, Inc., 129 S. Ct. 1800, 1812 (2009) (internal quotation marks omitted). Indeed, Justice Scalia paid little mind to the FCC's feeble attempt to reconcile its new policy with its prior regime, which the agency viewed as necessary under its view of arbitrary and capricious review, but which he viewed as irrelevant under the APA's text. *Id.* (characterizing the Commission's attempt to reconcile the old and new fleeting expletives policies as "superfluous," "irrelevant," an "unnecessary detour," and not "entirely convincing").

32. Not every Justice appreciated the rule–policy distinction in the case's particular context. *See id.* at 1825 (Stevens, J., dissenting) ("[T]he Court espouses the novel proposition that the Commission need not explain its decision to discard a longstanding rule in favor of a dramatically different approach to regulation.").

incorporating all of the Administrative Procedure Act's notice-and-comment procedural requirements into arbitrary-and-capricious review of adjudicatory decisions."[33] Therefore, Justice Breyer's argument that the FCC had acted arbitrarily by failing to address a scenario raised by a party to the proceeding was unavailing.[34] There was "scant empirical evidence" supporting the FCC's change to a more aggressive indecency policy, but unlike as in a rulemaking, this was no fatal flaw.[35] In other words, the FCC in indecency-regulating mode is engaged in a regulatory activity that is procedurally distinct from a carbon-emissions rulemaking before the Environmental Protection Agency (EPA).[36] In the former case, it is enough for the agency to articulate a coherent rationale for its policy, and a reviewing court will test for the "coherence of the rationale the agency gave."[37] *Fox*, therefore, is an application of arbitrary and capricious review to an adjudicatory proceeding that by definition lacked a notice-and-comment record; because there was no such record, there could be no error in the agency's neglect of record data. The case's greatest impact may be to encourage agencies to take up "soft" regulatory topics like indecency in adjudications rather than rulemakings, where the agency (1) can avoid the burden of "respond[ing] to all significant comments" by regulated entities and the public,[38] and (2) may more freely draw its own conclusions based on its reasoning and expertise, so long as it makes a rational effort to justify those conclusions.

The part of the opinion joined only by Chief Justice Roberts and Justices Thomas and Alito, and that Professor Watts focuses on, noted that the FCC's change in policy was "spurred by significant political pressure from Congress."[39] However, Justice Scalia also wrote that Congress's influence is an "extrastatutory" one.[40] Addressing Justice Stevens's claim in dissent that Congress exercises political influence over the FCC and arbitrary and

33. *Id.* at 1819 n.8 (plurality opinion).

34. *Id.* at 1837–38 (Breyer, J., dissenting).

35. *Id.* at 1813 (majority opinion) ("There are some propositions for which scant empirical evidence can be marshaled, and the harmful effect of broadcast profanity on children is one of them. . . . It is one thing to set aside agency action under the Administrative Procedure Act because of failure to adduce empirical data that can readily be obtained [such as data regarding the passive restraints in *State Farm*]. It is something else to insist upon obtaining the unobtainable." (citation omitted)).

36. *See id.* at 1824 (Kennedy, J., concurring in part and concurring in the judgment) ("The FCC did not base its prior policy on factual findings.").

37. *Id.* at 1817 (plurality opinion).

38. *Id.* at 1837 (Breyer, J., dissenting) (quoting ACLU v. FCC, 823 F.2d 1554, 1581 (D.C. Cir. 1987)) (emphasis omitted).

39. *Id.* at 1815–16 & 1816 n.4 (plurality opinion).

40. *Id.* at 1816 n.5.

capricious review should therefore operate as a check upon that influence, Justice Scalia argued that "[*i*]*f the FCC is indeed an agent of Congress*, it would seem an adequate explanation of its change of position that Congress made clear its wishes for stricter enforcement" because "[*t*]*he Administrative Procedure Act . . . does not apply to Congress and its agencies.*"[41] Political pressure or political branch policy concerns are thus matters exogenous not only to arbitrary and capricious review of agency rulemakings, but to the APA itself.

So at most, *Fox* means that (1) the APA does not require an agency to harmonize its past policies when undertaking a new policy direction outside of notice-and-comment rulemaking and that (2) four Justices noted that Congress sometimes exerts extrastatutory influence that can bear upon agency policy. Even under Justice Scalia and his three concurring colleagues' interpretation of arbitrary and capricious review, an agency must still show that action taken pursuant to a new policy is "permissible under the statute" and "that there are good reasons for it."[42] Nothing indicates that these "reasons" can or should include considerations of political influence. If *Fox* were applied to a rulemaking, the relevant record need reach back only to the record of the proceeding adopting the change, not to every Administration-spanning step the agency has taken in a particular area.[43] This may lower the burden on agency rule changes, but it does not necessarily open the door for the agency to cite political branch influence as justification for the change. An Obama FTC need only justify its own rule, not its departure from the Bush FTC's rule on the same issue—but it is still insufficient for arbitrary and capricious review purposes for the agency to justify the new rule by declaring, "This is no longer the Bush FTC."

CONCLUSION

None of us—judges included—are willfully ignorant of the fact that political branch priorities play a dominant, and much of the time dispositive, role in agency policymaking. But the rulemaking process, as contemplated by the APA and as recognized by arbitrary and capricious judicial review, envisions a procedure that insulates affected parties from that role. If, on the other hand, expert-based decisionmaking is, as Professor Watts posits, a product of judicial preference as expressed through arbitrary and capricious review, then let that preference serve as

41. *Id.* at 1816–17 (emphases added).

42. *Id.* at 1811 (majority opinion).

43. *See id.* ("The statute makes no distinction . . . between initial agency action and subsequent agency action undoing or revising that action.").

the last line of defense protecting regulated entities from political caprice.

PROXIMITY, PRESUMPTIONS, AND PUBLIC PARTICIPATION: REFORMING STANDING AT THE NUCLEAR REGULATORY COMMISSION

DAVID A. REPKA* AND TYSON R. SMITH**

TABLE OF CONTENTS

* Partner, Winston & Strawn LLP. J.D., 1981, Georgetown University Law Center; B.A., 1978, Northwestern University, Physics.

** Associate, Winston & Strawn LLP. J.D., 2003, Lewis & Clark School of Law, Certificate in Environmental and Natural Resource Law; M.S., 1999, Stanford University, Civil and Environmental Engineering; B.S., 1995, Vanderbilt University, Civil and Environmental Engineering.

INTRODUCTION

Various federal statutes provide opportunities for members of the public to participate in agency administrative proceedings if their "interests" may be affected by a licensing or permitting action. An agency's selection of criteria to assess the adequacy of the interests can have significant legal, financial, and regulatory consequences for the public, the regulators, and the regulated community. However, the nature of the issues at stake in some administrative hearings poses challenges to the application of traditional Article III judicial standing principles as the threshold test for participation.

For some agencies, such as the U.S. Nuclear Regulatory Commission (NRC or Commission), the decision to grant a request for a hearing by a member of the public permits active participation by public stakeholders in the hearing process but can also cost millions of dollars and add years of delay to the licensing process.[1] NRC licensing was the subject of substantial controversy in the 1980s.[2] Now, as the industry seems poised for rebirth in the United States, the NRC licensing process is again a focus of attention. It is therefore appropriate to reconsider the criteria and processes by which the NRC determines whether a person does or does not have standing to

1. Among the twenty-five most recently licensed plants, the length of time from filing of the application for a construction permit through issuance of the operating license ranged from 11.5 to 24.7 years. Letter from Nils J. Diaz, Chairman, Nuclear Regulatory Comm'n, to Rep. Joe Barton (R-Tex.), U.S. House of Representatives (Feb. 20, 2006), http://www.nrc.gov/reading-rm/doc-collections/congress-docs/correspondence/2006/barton-02-20-2006.pdf.

2. Although initial applications for many plants were filed in the early 1970s, operating licenses were not issued until the late 1980s or early 1990s. These licensing delays contributed to large construction cost overruns and additional delays. The cancellation of the Shoreham nuclear power plant, which had been completed but never operated after expenditures of $5.5 billion, epitomizes the regulatory gridlock of the era. CHARLES KOMANOFF & CORA ROELOFS, KOMANOFF ENERGY ASSOCS., FISCAL FISSION: THE ECONOMIC FAILURE OF NUCLEAR POWER; A GREENPEACE REPORT ON THE HISTORICAL COSTS OF NUCLEAR POWER IN THE UNITED STATES 23 (1992), http://www.earthtrack.net/earthtrack/library/FiscalFission.pdf. Many other plants faced similar delays and cost overruns. Comanche Peak Units 1 and 2 cost more than $9 billion to complete. Jack Z. Smith, *Another Shot for Nuclear*, FORT WORTH STAR-TELEGRAM, Dec. 28, 2007, at B13. Vogtle Units 1 and 2 cost $8.87 billion. Jon Gertner, *Atomic Balm?*, N.Y. TIMES, July 16, 2006, (Magazine), at 36, 38. Seabrook Unit 1 alone cost $6.2 billion. Matthew L. Wald, *N.R.C. Panel Supports a License for Seabrook*, N.Y. TIMES, Nov. 14, 1989, at D2. Watts Bar Unit 1 cost $6.8 billion. U.S. GOV'T ACCOUNTABILITY OFFICE, TENNESSEE VALLEY AUTHORITY: FINANCIAL PROBLEMS RAISE QUESTIONS ABOUT LONG-TERM VIABILITY 5 (1995).

participate in a particular NRC proceeding. Below, we discuss the history of standing at the NRC, describe the difficulties involved in applying Article III case law to NRC adjudications, and explore options for reforming standing at the NRC in a way that balances the public's right to participate in the hearing process with the applicant's right to an efficient and timely licensing decision.

I. BACKGROUND

A. Standing to Participate in NRC Hearings

As set forth in the Atomic Energy Act (AEA), the NRC must offer an opportunity for a hearing on many licensing actions involving a facility that produces or uses nuclear material, including an application for a license to construct and operate a nuclear facility.[3] Administrative judges from the Atomic Safety and Licensing Board (ASLB) conduct these hearings, typically in three-judge panels (one legal judge and two technical judges).[4] The judges are employees of the NRC but are independent from the NRC staff and have no stake in the outcome of a proceeding. The Commission entertains appeals and petitions for review of the decisions of the ASLB.[5]

According to the NRC, "[a] petitioner's standing, or right to participate in a Commission licensing proceeding, is grounded in section 189a of the [AEA], 42 U.S.C. § 2239(a)(1)(A), which requires the NRC to provide a hearing 'upon the request of any person whose interest may be affected by the proceeding.'"[6] Any person who requests a hearing or seeks to intervene in a Commission proceeding must demonstrate that he or she has a sufficient interest, or standing.[7] "Standing is not a mere legal technicality"; it is a necessary and vital part of our legal system that serves to ensure that litigation is limited to real disputes that are appropriate for judicial resolution.[8]

---

3. For example, the NRC offers an opportunity to request a hearing on applications to construct and operate new nuclear power plants. 42 U.S.C. § 2239 (a)(1)(A) (2006).

4. 10 C.F.R. § 2.313 (2009).

5. *Id.*

6. *In re* Duke Energy Corp. (McGuire Nuclear Station, Units 1 & 2; Catawba Nuclear Station, Units 1 & 2), LBP-02-4, 55 N.R.C. 49, 61 (2002); *see also In re* Me. Yankee Atomic Power Co. (Me. Yankee Atomic Power Station), CLI-04-5, 59 N.R.C. 52, 56 n.14 (2004) (citing 42 U.S.C. § 2239(a)(1) (2000)).

7. 10 C.F.R. § 2.341 (2009).

8. *In re* Westinghouse Elec. Corp. (Nuclear Fuel Export License for Czech Republic —Temelin Nuclear Power Plants), CLI-94-7, 39 N.R.C. 322, 331–32 (1994) (citation omitted).

B. Application of Judicial Concepts of Standing in NRC Proceedings

Because agencies are neither constrained by Article III[9] nor governed by judge-made standing doctrines limiting access to the federal courts, "administrative standing" may be easier to attain than "judicial standing."[10] While judicial proceedings are intended to resolve genuine controversies, administrative tribunals were created to "uphold the public interest."[11] Agencies may therefore wish to encourage greater public participation than that permitted by Article III in order to enhance the quality and transparency of their decisionmaking. Agencies may also seek different perspectives than those of the typical participants in administrative proceedings (i.e., regulated entities). Nonetheless, the NRC has, as a matter of choice, long applied contemporaneous judicial concepts of standing to determine whether a party has a sufficient interest to intervene as a matter of right.[12]

In *In re Portland General Electric Co.* (*Pebble Springs I*), the NRC's Appeal Board certified a question to the Commission: Should standing in NRC proceedings be governed by "judicial" standards?[13] The Commission responded to the certified question in *Pebble Springs II* by ruling that judicial concepts of standing should be applied by adjudicatory boards in determining whether a petitioner is entitled to intervene as of right under § 189a of the AEA.[14] This continues to be current Commission practice.[15]

The Commission in *Pebble Springs II* also held that licensing boards may, as a matter of discretion, grant intervention in licensing cases to petitioners

---

9. U.S. CONST. art. III.

10. Envirocare of Utah, Inc. v. NRC, 194 F.3d 72, 74 (D.C. Cir. 1999) (citation omitted); *see also* HENRY J. FRIENDLY, FEDERAL JURISDICTION: A GENERAL VIEW 118 (1973) (citing 3 KENNETH CULP DAVIS, ADMINISTRATIVE LAW TREATISE § 22.08, at 241 (1st ed. 1958)) (asserting that the differences between judicial standing and administrative standing include "[t]he need for a 'case or controversy' to seek judicial review but not to intervene in an administrative hearing; the differences between statutes and agency rules controlling intervention and statutes controlling judicial review; and the differing characters of administrative and judicial proceedings").

11. *See, e.g.*, Tex. Indus. Traffic League v. R.R. Comm'n, 628 S.W.2d 187, 197 (Tex. App. 1982), *rev'd on other grounds*, 633 S.W.2d 821 (Tex. 1982) (adding that administrative tribunals accomplish this purpose "through the exercise of their investigative, rulemaking and quasi-judicial powers").

12. *In re* Yankee Atomic Elec. Co. (Yankee Nuclear Power Station), CLI-98-21, 48 N.R.C. 185, 195 (1998).

13. *In re* Portland Gen. Elec. Co. (Pebble Springs Nuclear Plant, Units 1 & 2) (*Pebble Springs I*), ALAB-333, 3 N.R.C. 804, 807 (1976).

14. *In re* Portland Gen. Elec. Co. (Pebble Springs Nuclear Plant, Units 1 & 2) (*Pebble Springs II*), CLI-76-27, 4 N.R.C. 610, 613–14 (1976).

15. *In re* Calvert Cliffs 3 Nuclear Project, L.L.C. (Combined License Application for Calvert Cliffs, Unit 3) (*Calvert Cliffs 3*), (No. 52-016-COL) CLI-09-20, slip op. at 6–7 (N.R.C. Oct. 13, 2009) (applying *Lujan v. Defenders of Wildlife*, 504 U.S. 555 (1992)).

who are not entitled to intervene as of right under judicial standing doctrines but who may, nevertheless, make some contribution to the proceeding.[16] This is referred to as "discretionary standing," and the criteria for assessing discretionary standing are now codified in NRC regulations.[17]

C. Proximity Presumption

Under Article III, the Supreme Court has established the now-familiar three-prong test for standing.[18] Ostensibly in furtherance of its application of this judicial test, the NRC has established a "shortcut" that obviates the need for a petitioner to provide information addressing each of the three prongs of traditional standing concepts (injury in fact, causation, and redressability).[19] In proceedings involving proposed nuclear power reactors, the Commission has adopted a presumption whereby a petitioner can base its standing upon a showing that his or her residence, or—in the case of an organization—that of its members, is within the geographical proximity (usually taken to be fifty miles) of the proposed nuclear unit. The presumption is that individuals within the radius might be affected by a potential accidental release of fission products from a nuclear power plant.[20] For other lesser NRC approvals, such as license amendments, the geographic scope of the presumption is more limited.[21]

---

16. *Pebble Springs II*, CLI-76-27, 4 N.R.C. at 616.

17. 10 C.F.R. § 2.309(e) (2009).

18. *See, e.g.*, Lujan v. Defenders of Wildlife, 504 U.S. 555, 560–61 (1992) (reciting the bases for the injury in fact, causation, and redressability elements of the Article III standing inquiry).

19. *See Calvert Cliffs 3*, CLI-09-20, slip op. at 6–7. In this case the Commission focused on its ability to deal with standing issues generically. That principle seems uncontroversial. The Commission, however, did not deal as effectively with the issue of whether standing can be based on a risk widely shared by all persons living in the vicinity of the proposed plant. The Commission relied on its technical expertise and the generic conclusion that off-site risks may be significant. The issue of standing based on risk is discussed further below.

20. *In re* Houston Lighting & Power Co. (S. Tex. Project, Units 1 & 2), LBP-79-10, 9 N.R.C. 439, 443 (1979); *In re* Detroit Edison Co. (Enrico Fermi Atomic Power Plant, Unit 2), LBP-79-1, 9 N.R.C. 73, 78 (1979); *see also In re* Phila. Elec. Co. (Limerick Generating Station, Units 1 & 2), LBP-82-43A, 15 N.R.C. 1423, 1447 (1982) (holding that a residence more than seventy-five miles from a plant will not "alone . . . establish an interest sufficient for standing as a matter of right").

21. The Commission will apply the proximity presumption to licensing actions if the party shows that a particular licensing action raises an "obvious potential for offsite consequences." *In re* Exelon Generation Co. (Peach Bottom Atomic Power Station, Units 2 & 3), CLI-05-26, 62 N.R.C. 577, 581 (2005); *see id.* (concluding that the risks associated with transferring a non-operating, 50% ownership interest in a power reactor were *de minimis* and therefore did not justify proximity standing); *In re* U.S. Dep't of the Army (Army Research Lab.), LBP-00-21, 52 N.R.C. 107, 107–08 (2000) (declining to apply the proximity

According to the Commission, a petitioner residing near a nuclear facility need not personally show a causal relationship between injury to its interest and the licensing action being sought in order to establish standing.[22] Instead, mere proximity is deemed sufficient—standing alone—to establish the requisite interest for intervention on the basis that "in construction permit and operating license cases . . . persons living within the roughly 50-mile radius of the facility 'face a realistic threat of harm' if a release from the facility of radioactive material were to occur."[23] Thus, this "proximity presumption" purports to reflect a generic determination and application of judicial concepts.[24] Petitioners invoking the presumption need not show any other injury beyond mere risk, such as injury from planned construction activities or from routine operations of the plant.

The proximity presumption used in reactor construction and operating license proceedings also applies to reactor license renewal proceedings. The Commission determined that reactor license extension cases should be treated similarly because they allow operation of a reactor over an additional period of time during which the reactor could be subject to some of the same equipment failures and personnel errors as during operations over the original period of the license.[25] According to the Commission, "the incremental risk of reactor operation for an additional 12–15 years is sufficient to invoke the presumption of injury in fact for persons residing within 10 to 20 miles from the facility."[26] In such a case the petitioner is

---

presumption in a proceeding for an amendment to a materials license based on a person's residence twenty miles from a site); *In re* Entergy Nuclear Vt. Yankee, L.L.C. (Vt. Yankee Nuclear Power Station), LBP-04-28, 60 N.R.C. 548, 553–54 (2004) (applying the proximity presumption to an extended power uprate application based on representative members living within fifteen miles of the plant).

22. *In re* Armed Forces Radiobiology Research Inst. (Cobalt-60 Storage Facility), ALAB-682, 16 N.R.C. 150, 153 (1982), (citing *In re* Va. Elec. & Power Co. (N. Anna Nuclear Power Station, Units 1 & 2), ALAB-522, 9 N.R.C. 54, 57 n.5 (1979)); *In re* Ga. Inst. of Tech. (Ga. Tech Research Reactor, Atlanta, Ga.), LBP-95-6, 41 N.R.C. 281, 287 (1995).

23. *Calvert Cliffs 3*, CLI-09-20, slip op. at 7.

24. *Id.* Although the Commission asserted in *Calvert Cliffs 3* that the Supreme Court in *Lujan* created a similar presumption for persons living adjacent to the site for a proposed federal dam, the Commission does not recognize that *Lujan* was referring to a procedural rather than a substantive injury. In footnote 7 in *Lujan*, the Supreme Court distinguished a procedural injury (e.g., the failure to prepare an environmental impact statement) from a person who lacks a concrete interest, such as a person living far from the proposed dam site. Lujan v. Defenders of Wildlife, 504 U.S. 555, 572 n.7 (1992). The Commission's proximity presumption presumes a concrete harm even in the absence of an alleged procedural harm. The Commission therefore has effectively eliminated the requirement that a petitioner specifically demonstrate a concrete injury in fact.

25. *In re* Duke Energy Corp. (Oconee Nuclear Station, Units 1, 2, & 3), LBP-98-33, 48 N.R.C. 381, 385 n.1 (1998).

26. *See, e.g.*, *In re* Pac. Gas & Elec. Co. (Diablo Canyon Nuclear Power Plant, Units 1 & 2), LBP-93-1, 37 N.R.C. 5, 6 (1993).

not required to show "that his concerns are well-founded in fact."[27]

II. DISCUSSION

A. Changes in Federal Standing Jurisprudence

The Commission's proximity presumption has remained relatively unchanged since it was first adopted in the late 1970s. However, judicial concepts of standing have been clarified since that time, effectively refuting the basis for a presumption based on hypothetical accident risk. In *Lujan v. Defenders of Wildlife*, the Supreme Court made clear that plaintiffs must suffer a concrete, discernible injury to be able to bring suit.[28] This injury in fact requirement is case specific, "turn[ing] on the nature and source of the claim asserted"[29] and "whether the complainant has personally suffered the harm."[30] Moreover, the alleged harm must be "concrete and particularized" and "actual or imminent, not conjectural or hypothetical."[31] These qualifiers ensure that courts address only cases and controversies in which the plaintiff is "in a personal and individual way"[32] "immediately in danger of sustaining some direct injury,"[33] thus avoiding advisory opinions on matters "in which no injury would have occurred at all."[34]

By requiring plaintiffs to demonstrate an injury in a concrete factual context, courts also avoid claims involving only "generalized grievances" shared by other members of the public.[35] When a party's "asserted injury arises from the government's allegedly unlawful regulation (or lack of regulation) of *someone else*"—such as when a petitioner challenges a license application but is not itself regulated by the NRC—"standing . . . is

27. *In re* Va. Elec. & Power Co. (N. Anna Nuclear Power Station, Units 1 & 2), ALAB-522, 9 N.R.C. 54, 56 (1979); *see also In re* Duquesne Light Co. (Beaver Valley Power Station, Unit 2), LBP-84-6, 19 N.R.C. 393, 410, 429 (1984).

28. *Lujan*, 504 U.S. at 560.

29. Raines v. Byrd, 521 U.S. 811, 818 (1997) (quoting Warth v. Seldin, 422 U.S. 490, 500 (1975)).

30. Wilderness Soc'y v. Alcock, 83 F.3d 386, 390 (11th Cir. 1996).

31. *Lujan*, 504 U.S. at 560 (quoting Whitmore v. Arkansas, 495 U.S. 149, 155 (1990)) (internal quotation marks omitted); *see also* Summers v. Earth Island Inst., 129 S. Ct. 1142, 1152 (2009) ("Standing, we have said, is not an ingenious academic exercise in the conceivable . . . [but] requires . . . a factual showing of perceptible harm." (quoting *Lujan*, 504 U.S. at 566) (internal quotation marks omitted) (alterations in original)); *id.* at 1151–52 (declining to rely on a "statistical probability" or a "realistic threat" to establish that individuals are threatened with concrete injury).

32. *Lujan*, 504 U.S. at 560 n.1.

33. City of Los Angeles v. Lyons, 461 U.S. 95, 102 (1983).

34. *Lujan*, 504 U.S. at 564 n.2.

35. Valley Forge Christian Coll. v. Ams. United for Separation of Church & State, 454 U.S. 464, 475 (1982).

ordinarily 'substantially more difficult' to establish."[36] Indeed, the Supreme Court has held that "much more is needed" in terms of the "nature and extent of facts . . . averred" to show that the petitioner will be affected by the alleged injury "in such a manner as to produce causation."[37] The Supreme Court's standing test is plainly more demanding than the Commission's now outdated and overly simplified proximity presumption, which is based on no more than the speculative, hypothetical possibility of a reactor accident in the future that will somehow injure any and all off-site residents within a fifty-mile radius.[38]

Recently, the Supreme Court issued a decision on standing that directly undermines the basis for the NRC's proximity presumption.[39] The Court began by reiterating the traditional standing principles—that is, that standing requires a concrete injury in fact that is actual and imminent and not hypothetical or conjectural. The Court then found that a plaintiff's "intention" to visit the National Forests in the future, without showing that the challenged regulations would affect a specific forest visited by the plaintiff, "would be tantamount to eliminating the requirement of concrete, particularized injury in fact."[40] The Court rejected a standing test that would have accepted a statistical probability that some of an organization's members would be threatened with concrete injury.[41] The Court also declined to substitute the requirement for "imminent" harm with a requirement of a "realistic threat."[42] In doing so, the Supreme Court rejected a standing test that is substantially similar to the test embedded in the NRC's proximity presumption, which is based on hypothetical accidents or risk rather than concrete injury in fact.[43]

---

36. *Lujan*, 504 U.S. at 562 (quoting Allen v. Wright, 468 U.S. 737, 758 (1984)).

37. *Id.* at 560–62.

38. In the absence of an actual injury from plant construction or from an ongoing discharge from the plant, there could be no standing based on an unsupported claim regarding the risk of an accidental release or the fear of an accidental release. *See generally* Metro. Edison Co. v. People Against Nuclear Energy, 460 U.S. 766 (1983) (holding that fear of an accident is not a cognizable injury under the National Environmental Policy Act (NEPA)).

39. *See* Summers v. Earth Island Inst., 129 S. Ct. 1142 (2009).

40. *Id*. at 1150.

41. *Id*. at 1151. The Court also declined to reduce the threshold for standing because the case involved a procedural injury (such as a claim under NEPA). Specifically, the Court concluded that "deprivation of a procedural right without some concrete interest that is affected by the deprivation—a procedural right *in vacuo*—is insufficient to create . . . standing." *Id.*

42. *Id*. at 1152 (emphasis omitted).

43. *Summers* would also appear to call into question the types of standing analyses that have recently been used by the D.C. Circuit to permit a finding of injury in fact based on a showing that harm was "substantially probable." *See* Fla. Audubon Soc'y v. Bentsen, 94 F.3d 658, 665 (D.C. Cir. 1996); Natural Res. Def. Council, Inc. v. EPA, 464 F.3d 1, 6 (D.C. Cir.

B. Other Issues with Proximity Presumption

The NRC's proximity presumption creates additional issues for an orderly administrative process. As discussed above, the proximity presumption presupposes harm from an accidental release from a plant. A petitioner, therefore, can raise issues of accident risk for hearing. But can a party less than fifty miles away who is only affected by a prospective accident raise other issues, or "contentions," for hearing (e.g., construction impacts, wetland destruction, or occupational exposures)? The NRC has said yes, but this also does not appear to be a defensible construction of judicial standing.

1. Concrete and Particularized

In the recent *Calvert Cliffs* case, the affidavits accompanying the request for hearing noted the location of the individuals' residences from the proposed facility (e.g., forty-five miles away) and the affiants expressed "concern" that the proposed new unit could affect their health and safety and the integrity of the environment.[44] Specifically, for standing, each individual stated only that he or she was concerned about the risk of accidental releases to the environment and the potential harm to groundwater and surface water supplies. That, however, was the extent of the alleged injury. The petitioners provided no information regarding the potential for an accident, how it might occur, the quantitative risk, or methods by which they personally might be harmed by an accident. Based on that showing of standing, the petitioners offered their contentions for hearing. The specific contentions they presented also had nothing to do with accidents or accidental releases.[45]

2006).

44. *In re* Calvert Cliffs 3 Nuclear Project, L.L.C., (Combined License Application for Calvert Cliffs Unit 3), LBP-09-04, slip op. at 7–9 (N.R.C. Mar. 24, 2009).

45. In contrast to the petitioners' focus on the risk of an accident as the basis for standing, the admitted contentions had little to no bearing on the potential for or causes of accidental releases. For example, one contention related to prospective foreign participation in the project and compliance with the Atomic Energy Act's foreign ownership and control restrictions. *Id.* at 24–31; *see also* 42 U.S.C. § 2133(d) (2006). This contention related primarily to security and control of special nuclear material, not accident risk. Other contentions related to the applicants' satisfaction of the financial test to provide decommissioning funding assurance through a parent guarantee or to on-site storage of *low-level* radioactive waste. Certainly, neither the timing of financial tests for decommissioning funding nor low-level waste management relate to the risk of accidents. *See, e.g.*, *In re* Calvert Cliffs 3 Nuclear Project, L.L.C. (Combined License Application for Calvert Cliffs Unit 3) LBP-09-04, slip op. at 31–33.

2. Speculative

Another issue arises in connection with the speculative nature of an accident. Judicial standing would require a concrete or threatened injury. However, presuming that an accident will occur at some unspecified point in the future from some undetermined cause is, by its nature, speculative and hypothetical. The probabilities of an accident occurring are projected to be very low (on the order of 1E-06/year).[46] The probability of an accident resulting in an actual injury to a person within fifty miles is much smaller still. And, in the absence of a posited accident mechanism, it cannot be said that the injury is "fairly traced" to the NRC's licensing of the facility.

3. Imminence

Using proximity as a surrogate for injury also undermines the temporal aspect of standing. In *Lujan*, the Court's standing analysis crystallizes and focuses on two aspects of the injury-in-fact requirement: the particularity (or specificity) aspect, which requires that the injury be to the party seeking review; and the temporal aspect, which requires that the injury be impending (or "soon").[47] As to the former aspect, it is an irreducible constitutional minimum of standing that a person suffer an injury-in-fact.[48] According to the Supreme Court, in order for injury to be "particularized," it must affect the plaintiff in a *personal and individual* way, such that "the party seeking review be himself among the injured."[49] As to the latter, the Court recognizes that the timing of injury may be flexible, but at the very least, "imminent" means sooner than "in this lifetime."[50]

The problems that the NRC creates by relying on judicial tests are highlighted by the proposed high-level waste repository at Yucca Mountain. The spent fuel from nuclear plants is proposed to be placed into canisters and stored within tunnels carved into the mountain. Even assuming canister failures and releases to the environment, the releases would not occur for tens, if not thousands, of years—well beyond the lifetime of any person alive today.[51] There is no suggestion—by anyone—

46. The total core damage frequency (CDF) for the design of Calvert Cliffs Unit 3 is 5.3E-07/year. AREVA, U.S. EPR FINAL SAFETY ANALYSIS REPORT 19.1.8.1, http://adamswebswearch2.nrc.gov/idmws/ViewDocByAccession.asp?AccessionNumber=ML091671748. The large early release frequency (LERF) from internal events is 2.6E-08/year. *Id.*

47. Lujan v. Defenders of Wildlife, 504 U.S. 555, 564 n.2 (1992).

48. Sierra Club v. Morton, 405 U.S. 727, 734–35 (1972).

49. *Id.*

50. *Lujan*, 504 U.S. at 564 n.2.

51. Potentially imminent injuries might include impacts due to construction (e.g.,

that such releases would occur in the near future. Under these circumstances, judicial standing could not be demonstrated based on the hypothetical, unintended releases because no person currently alive would be personally injured. Yet, a petitioner would have standing under the NRC's proximity presumption.[52]

4. Causation and Redressability

The NRC's analysis also seems plainly inconsistent with the causation and redressability elements of standing. Consider the situation where a petitioner is concerned with the impact of the facility on a nearby water body (e.g., harm to a particular aquatic species). In such circumstances, the NRC would permit a party to participate based on a speculative, hypothetical future injury from an accident. However, the speculative "injury" (harm from an accident) would not be caused by the aquatic species' impacts that the petitioners seek to litigate. Moreover, addressing the harm to aquatic species would not redress an injury caused by the hypothetical accident.

Under judicial standing precedent, the petition would fail at least two, and possibly all three, of the elements of standing. Yet, under NRC precedent, the petitioner would be allowed to participate in the proceeding, triggering automatic disclosure requirements and (potentially) adjudicatory hearings. In light of the tenuous relationship between the purported injury and the issues subject to the proceeding, it is far from clear that the proximity presumption and a lack of a tie between standing and the claims involved comport with efficiency of the process (time and expertise required to address the point).

In this regard, it is an important factor that the NRC also permits (quite voluntarily it would seem)[53] parties to litigate National Environmental Policy Act (NEPA) issues in its hearing process. It is a fundamental tenet of NEPA that the statute demands only "disclosure" and not a particular course of action. If the remedy for a NEPA violation in an NRC proceeding is mere disclosure, then how can additional disclosure redress

clearing land or construction dust). Injuries due to routine operations might also arguably be imminent if they could be identified.

52. The NRC's regulations provide that state and local governments are excused from demonstrating standing to participate in a proceeding for a repository within their borders and need only satisfy the admissible contention requirements of 10 C.F.R. § 2.309(f). 10 C.F.R. § 2.309(d)(2)(iii) (2009).

53. The AEA does not require litigation of National Environmental Policy Act (NEPA)-related contentions (as opposed to contentions involving issues of radiological health and safety). NEPA has its own public participation process and will be discussed in greater detail below.

an injury (i.e., eliminate "risk") from a future accident? At bottom, the proximity presumption may be fairly straightforward to apply and certainly increases public participation. But the presumption also yields results that are inconsistent with judicial standing principles and potentially inconsistent with the NRC's own policy considerations related to efficiency and timely processing of applications.

C. Possible Solutions for Improving Standing Assessments

Although we have highlighted some of the apparent inconsistencies between the NRC's proximity presumption and traditional concepts of judicial standing, we can appreciate the challenges that an agency such as the NRC faces in attempting to satisfy the AEA "interest" requirement and, in so doing, balancing the need for public participation in its processes and the rights of applicants to fair, efficient, and timely reviews of license applications. Below, we explore several possible approaches to improve the test for demonstrating an adequate interest in NRC or other administrative proceedings under the AEA or the Administrative Procedure Act (APA).

1. Revert to Strict Application of Judicial Concepts in Agency Proceedings

One approach to resolving the conflict between NRC practice and judicial concepts is also one that would be simple to implement. Rather than carve out exceptions from judicial concepts where there is a remote possibility of an accident, or awkwardly attempt to justify a results-driven application of judicial concepts, the Commission could simply require a petitioner to satisfy the judicial Article III test. This would require petitioners (or members of petitioning organizations) to do more than merely provide their addresses and the distance from their homes to the proposed reactor. Such a test would undoubtedly increase the showing required to participate but would not be a prohibitive barrier to participation.

Petitioners regularly challenge environmental rules, permits, and licenses in federal courts where they are required to establish injury, causation, and redressability. At the NRC, for environmental contentions, a petitioner would need to demonstrate that he or she would be injured by the construction or operation of the proposed plant and that a favorable outcome to the challenge would redress that harm. The injuries would need to involve concrete impacts from the project (e.g., excavation, land clearing, or routine effluents). For radiological safety issues, the petitioners would need to show some realistic nexus to off-site harm. Consistent with the case law, however, merely speculating that there might be an accident one day would not be enough. This approach also has the advantage of

providing the NRC with an existing body of cases (in the form of federal court decisions) that it could look to in evaluating standing.

2. Develop Regulations with Clear Criteria for Sufficient Interest

As noted above, agencies are neither constrained by Article III nor governed by judge-made standing doctrines.[54] Agencies therefore have wide discretion to craft their regulations governing participation in administrative hearings.[55] The Commission could therefore avoid the vexing legal issues of the judicial-standing inquiry entirely. It could establish, by rule, a balance between the need to permit public participation and the objectives of the hearing process. For example, the Commission could permit litigation on issues where a petitioner is likely to contribute something of value to the process and decline to litigate issues that have no bearing on the ultimate outcome of the licensing review or that could easily be remedied through the licensing review process (e.g., inadvertent omissions). No party benefits from the need to brief arcane legal concepts of standing, and the effort increases the cost, delay, and regulatory burden associated with a hearing. The Commission could establish a set of clear, objective criteria that would be sufficient to establish the requisite interest.

The Commission already has in place criteria for evaluating discretionary intervention.[56] Other criteria might also be transparent and easily applied. Some criteria might confer standing as of right. Others might require a case-by-case assessment by the presiding licensing board. For example, the right to participate could be based on

(i) distance to the proposed reactor (e.g., within ten miles);

(ii) participation in the NEPA process (e.g., attending meetings or submitting comments);

(iii) the extent to which the requestor's/petitioner's participation may reasonably be expected to assist in developing a sound record;[57]

---

54. Envirocare of Utah, Inc. v. NRC, 194 F.3d 72, 74 (D.C. Cir. 1999).

55. Vt. Yankee Nuclear Power Corp. v. Natural Res. Def. Council, Inc., 435 U.S. 519, 543–49 (1978).

56. *See* 10 C.F.R. § 2.309(e)(1)–(2) (outlining factors for standing consideration); *Pebble Springs II*, 4 N.R.C. 610, 616 (1976) (presenting factors both in favor and against intervention).

57. Considerations in determining the petitioner's ability to contribute to development of a sound record include the following:

> (1) a petitioner's showing of significant ability to contribute on substantial issues of law or fact which will not be otherwise properly raised or presented; (2) the specificity of such ability to contribute on those substantial issues of law or fact; (3) justification of time spent on considering the substantial issues of law or fact; (4) provision of additional testimony, particular expertise, or expert assistance; and (5) specialized education or pertinent experience.

(iv) the nature and extent of the requestor's/petitioner's property, financial, or other interest in the proceeding;
(v) the availability of other means whereby requestor's/petitioner's interest will be protected;
(vi) the extent to which the requestor's/petitioner's interest will be represented by existing parties; or
(vii) the extent to which requestor's/petitioner's participation will inappropriately broaden or delay the proceeding.

Permitting intervention could therefore be based upon a petitioner's demonstration of the potential significant contribution it could make on substantial issues of law and fact not otherwise raised or presented and a showing of the importance and immediacy of those issues.

3. Require Standing for Each Contention

The Commission could continue to use a proximity presumption for standing but limit its applicability to contentions (i.e., claims or issues) that relate to accidents. Under this formulation, the Commission could decide to use the proximity presumption for a limited set of accident-related contentions. For contentions that relate to other safety or environmental concerns, a petitioner would need to establish standing through the traditional standing inquiry (injury in fact, causation, and redressability). This would eliminate the situation described above whereby a petitioner has standing (based solely on speculative risk of an accident) to raise claims relating to foreign ownership, low-level waste disposal, or impacts to aquatic species.

One example of this would be emergency planning issues. A petitioner may have difficulty demonstrating an injury in fact from a future, hypothetical accident. There is a very low probability that an accident would ever occur and the risk of an accident that would actually harm the specific petitioner is lower still. Yet the NRC could presumptively grant standing to persons living within a ten-mile or a fifty-mile radius for contentions involving emergency planning issues that arise in connection with the specific area in question. This approach would recognize the public's interest in participating in the hearing on significant issues where an individual might otherwise have difficulty in establishing standing under judicial standing principles. For the typical environmental or safety issue, however, the person would need to demonstrate injury in fact, causation, and redressability.

This approach would also be broadly consistent with the judicial

In re Duke Power Co. (Catawba Nuclear Station, Units 1 & 2), LBP-81-1, 13 N.R.C. 27, 33 (1981).

approach to standing. The Supreme Court recently reaffirmed the principle that standing must be shown for every single claim in *Davis v. Federal Election Commission*.[58] Precisely relevant to the current situation, the *Davis* Court reiterated that "standing is not dispensed in gross" and remarked that a party "must demonstrate standing for each claim he seeks to press" and "for each form of relief that is sought."[59] According to the Court, standing for one claim does not suffice for all claims even where those claims arise from the same nucleus of operative facts.[60]

Because standing is rooted in the need for an actual "case" or "controversy," holding otherwise, the Court noted, would undermine other important judicial principles and permit, for example, adjudication of moot or unripe claims.[61] The Court explained that the actual injury requirement would not ensure that there is a legitimate role for an agency adjudicatory body in dealing with a particular grievance if, once a party "demonstrated harm from one particular inadequacy in government administration," the adjudicatory body was "authorized to remedy *all* inadequacies in that administration."[62] As the Court emphasized in *Lewis*, "The remedy must of course be limited to the inadequacy that produced the injury in fact that the [party] has established."[63]

In *Calvert Cliffs 3*,[64] the Commission incorrectly distinguished *Lewis* and *DaimlerChrysler*. The Commission defined a claim as an issue that could result in the agency denying the license.[65] However, the NRC does not require that a claim (or contention) refer to some articulated form of relief, and this issue is often overlooked. For a NEPA-based contention, an applicant's failure to fully discuss impacts on the environment would not result in denial of the license. NEPA only compels disclosure; NEPA does not mandate a substantive outcome. Moreover, the responsibility to comply with NEPA actually lies with the NRC, not the applicant.

---

58. *See* 128 S. Ct. 2759, 2769 (2008).

59. *Id.* (citing DaimlerChrysler Corp. v. Cuno, 547 U.S. 332, 352 (2006), and Friends of the Earth, Inc. v. Laidlaw Envtl. Servs., Inc., 528 U.S. 167, 185 (2000)); *see also* Rosen v. Tenn. Comm'r of Fin. & Admin., 288 F.3d 918, 928 (6th Cir. 2002) ("It is black-letter law that standing is a claim-by-claim issue.").

60. DaimlerChrysler Corp. v. Cuno, 547 U.S. 332, 352 (2006).

61. *Id.*

62. Lewis v. Casey, 518 U.S. 343, 357 (1996).

63. *Id.*

64. *In re* Calvert Cliffs 3 Nuclear Project, L.L.C., & UniStar Nuclear Operating Servs., L.L.C. (Combined License Application for Calvert Cliffs, Unit 3) (*Calvert Cliffs 3*), No. 52-016-COL CLI-09-20, slip op. (N.R.C. Oct. 13, 2009).

65. *See id.* at 8 n.28 ("[S]o long as either denial of a license or issuance of a decision mandating compliance with legal requirements would alleviate a petitioner's potential injury, then under longstanding NRC jurisprudence the petitioner may prosecute any admissible contention that could result in the denial or in the compliance decision.").

Presuming that all contentions could lead to denial of a license is as flawed an approach as the proximity presumption.

By adopting an approach that would link interests and contentions, the Commission could maximize public participation while focusing on real issues and available relief. A petitioner who would have standing on accident risk would be required to demonstrate a contention that relates to accident risk. Petitioners who would raise other issues must show that they would personally suffer some injury related to the contention. And petitioners could not invoke generalized accident risk for standing on NEPA claims that cannot relieve or eliminate that risk. This approach would allow participation on those aspects of licensing with the greatest potential for significant environmental harm (accidents), while otherwise limiting the time and expense of a hearing to those issues where a petitioner can demonstrate an actual concrete harm to his or her interest with relief available in the proceeding.

4. Eliminate Hearings on NEPA Issues

Similar public policy objectives (fairness, efficiency, and public participation) might be achieved by focusing NRC hearings on issues of radiological health and safety. Under such an approach, public participation is not eliminated; the existing public scoping and comment process for environmental reviews would be used to resolve NEPA-related concerns. This would obviate the need for separate NRC hearings on NEPA issues. And in so doing, this approach would resolve some of the clearest inconsistencies between NRC practice and judicial standing concepts.

The Atomic Energy Commission (AEC), the predecessor to the NRC, initially elected to permit hearings on environmental issues but did not consider such hearings to be required by the AEA.[66] In 1971, the U.S. Court of Appeals for the D.C. Circuit rendered its decision in *Calvert Cliffs' Coordinating Committee v. Atomic Energy Commission*.[67] The court concluded that several aspects of the AEC's NEPA policy statement failed to comply with the NEPA statute. In the court's view, NEPA established environmental protection as an integral part of the AEC's basic mandate, and the court therefore concluded that the AEC must itself take the initiative of

66. *See* Implementation of the National Environmental Policy Act of 1969, 35 Fed. Reg. 5463 (Apr. 2, 1970) (codified at 10 C.F.R. pt. 50 app. D). The policy statement addressed preparation of the "detailed statement" (i.e., the Environmental Impact Statement) required by NEPA but also noted that the statement should not be construed as "extending the licensing or regulatory jurisdiction of the Commission." *Id.* at 5464.

67. 449 F.2d 1109 (D.C. Cir. 1971).

considering environmental values at every stage of the process beyond the staff's evaluation and recommendation.[68] The AEC subsequently revised its regulations to provide a hearing opportunity on environmental matters following the NEPA review.[69]

Subsequent judicial decisions have altered the conclusions underlying the 1971 *Calvert Cliffs* decision. By its terms, NEPA imposes procedural requirements on agencies, not substantive ones. "The statute requires only that an agency undertake an appropriate assessment of the environmental impacts of its action without mandating that the agency reach any particular result concerning that action."[70] The statute also "does not require agencies to adopt any particular internal decisionmaking structure."[71] And "[w]hile NEPA clearly mandates that an agency fully consider environmental issues, it does not itself provide for a hearing on those issues."[72]

The AEC interpreted the agency's jurisdiction under the AEA as limited to protecting against radiological hazards.[73] Courts have agreed with the AEC, recognizing that the Commission has jurisdiction under the AEA only to the extent necessary to provide adequate protection to "the health and safety of the public with respect to the special hazards" of radiological impacts.[74] Moreover, the right of interested persons to intervene as a party in a licensing proceeding stems from the AEA, not from NEPA, and is covered in AEA § 189 and 42 U.S.C. § 2239(a)(1)(A). In this context, the

68. *Id.* at 1117–19.

69. 36 Fed. Reg. 18,071 (Sept. 9, 1971). The hearing requirements in 10 C.F.R. Part 50, Appendix D, were incorporated into 10 C.F.R Part 2. *See* Restructuring of Facility License Application Review and Hearing Processes and Consideration of Environmental Statements, 37 Fed. Reg. 9331 (May 9, 1972). Appendix D to Part 50 eventually became 10 C.F.R. Part 51. *See* Licensing and Regulatory Policy and Procedures, 39 Fed. Reg. 26,279 (July 18, 1974).

70. *See, e.g.*, Robertson v. Methow Valley Citizens Council, 490 U.S. 332, 350 (1989); *In re* Babcock & Wilcox (Apollo, Pa. Fuel Fabrication Facility), LBP-93-4, 37 N.R.C. 72, 93 (1993); *In re* La. Energy Servs., L.P. (Claiborne Enrichment Ctr.), LBP-96-25, 44 N.R.C. 331, 341–42 (1996); *In re* Ne. Nuclear Energy Co. (Millstone Nuclear Power Station, Unit 3), CLI-01-3, 53 N.R.C. 22, 44 (2001).

71. Balt. Gas & Elec. Co. v. Natural Res. Def. Council, Inc., 462 U.S. 87, 100 (1983).

72. Kelley v. Selin, 42 F.3d 1501, 1512 (6th Cir. 1995) (quoting Union of Concerned Scientists v. NRC, 920 F.2d 50, 56 (D.C. Cir. 1990)). The Council on Environmental Quality has stated that "[p]ublic hearings or meetings, although often held, are not required; instead the manner in which public input will be sought is left to the discretion of the agency." Guidance Regarding NEPA Regulations, 48 Fed. Reg. 34,263 (July 28, 1983) (codified at 40 C.F.R. pt. 1500).

73. *See* New Hampshire v. AEC, 406 F.2d 170, 174–75 (1st Cir. 1969) (noting that "[t]he Commission has been consistent in confining itself to [radiological] hazards").

74. *Id.* at 174–75; *see also* Gage v. AEC, 479 F.2d 1214, 1220 n.19 (D.C. Cir. 1973) (asserting that the Commission lacks the authority to mandate that an applicant take certain actions that are unrelated to radiological considerations).

AEA hearing requirement only extends to those determinations made under the AEA related to "radiological consequences." The adequacy of the environmental impact statement is not a matter within the scope of the AEA.[75]

Under this approach, standing to raise AEA safety issues could be based on proximity (i.e., accident risk). However, other environmental concerns would not be addressed through the AEA hearing process but rather would be dealt with through a separate and independent process. Environmental issues not material to the adequacy of the license application under the AEA from a radiological health and safety standpoint would be handled through the NEPA scoping and comment process.

CONCLUSION

In light of the renewed interest in licensing new reactors and the continued focus on renewing the licenses of existing reactors, the NRC's hearing processes are again a focal point of attention from public stakeholders. Recent Commission decisions have focused on questions of the proper application of judicial standing principles to complex administrative matters. While the Commission's approach to standing may have once been consistent with judicial standing principles, the agency's long-standing proximity presumption is no longer aligned with those principles. Given the enormous potential for delay and the time and expense inherent in the NRC hearing processes, the Commission has an obligation to the public and its licensees to use its hearing powers wisely and in the pursuit of significant health and safety concerns. A consistent and defensible requirement for standing is an important part of that obligation.

Any reform must balance the public's right to participate in NRC licensing proceedings if petitioners have an adequate "interest" with the public interest in efficient and timely adjudicatory proceedings. We have outlined several approaches that could form the basis for a potential rulemaking to address the issue on a generic basis, avoiding recurring legal arguments and judicial review. The approaches range from minor refinements in the current processes to a radical departure from long-standing but outdated requirements to conduct hearings on environmental issues. At a minimum, we hope to spark a conversation as to the proper role of NRC adjudicatory authority in the pursuit of public participation

75. Some issues discussed in the environmental impact statements may also have radiological health and safety components and therefore could not be excluded entirely from consideration in hearings. For example, radiological dose consequences, severe accidents, and decommissioning strategies have both radiological and environmental components.

and, ultimately, protection of the public health and safety. The safety of our nuclear infrastructure is an overriding concern, but process merely for the sake of process does not promote public confidence in the NRC, its regulatory programs, or its licensees.

* * *

THE SUPREME COURT MAKES IT HARDER TO CONTEST ADMINISTRATIVE AGENCY POLICY SHIFTS IN *FCC V. FOX TELEVISION STATIONS, INC.*

CHARLES CHRISTOPHER DAVIS*

TABLE OF CONTENTS

> Unless we make the requirements for administrative action strict and demanding, *expertise*, the strength of modern government, can become a monster which rules with no practical limits on its discretion. Absolute discretion, like corruption, marks the beginning of the end of liberty.[1]

Nearly sixty years ago, Justice William O. Douglas dissented from a decision in which the Supreme Court upheld an action, taken by the Interstate Commerce Commission, raising intrastate railroad fare prices to comparable interstate levels.[2] The Court did not issue an opinion with its decision; Justice Douglas did. Justice Douglas was troubled by what he saw as the Commission's failure to "justify its action."[3] The details are

* J.D. Candidate, May 2011, American University Washington College of Law.

1. New York v. United States, 342 U.S. 882, 884 (1951) (Douglas, J., dissenting).
2. *Id.* at 882.
3. *Id.* at 883.

unimportant; as Justice Douglas himself remarked, "This case is perhaps insignificant in the annals."[4] Though the facts were arguably trivial, the potential ramifications of the decision worried Justice Douglas. Justice Douglas's warning seems, in hindsight, strangely alarmist. Still, while his words may have been menacing, the relevancy of his message remains true today. The administrative state wields an extraordinary amount of power and influence.[5] When courts fail to insist that administrative agencies supply thoroughly reasoned and rational explanations for their decisions, the social and economic liberties of all citizens may become implicated.[6] In a recent case, the Supreme Court arguably validated Justice Douglas's fears by washing away a judicial gloss on the Administrative Procedure Act's (APA's) arbitrary and capricious standard of judicial review, making it much easier for agencies to reverse themselves in the future.[7]

INTRODUCTION

In *FCC v. Fox Television Stations, Inc.* (*Fox II*),[8] the Federal Communications Commission (FCC or Commission) asked the Court to reconsider a ruling of the U.S. Court of Appeals for the Second Circuit holding the FCC's recent decision (that so-called fleeting expletives[9] may be found indecent) an arbitrary and capricious exercise of agency discretion under the APA.[10] Since the Court upheld the constitutionality of the FCC's enforcement powers in the late 1970s,[11] the FCC had followed a policy of forgoing indecency findings when only a single, isolated expletive was at issue.[12]

---

4. *Id.* at 884.

5. *See, e.g.*, WILLIAM O. DOUGLAS, POINTS OF REBELLION 79 (1969) ("The examples are legion and they cover a wide range of subjects from food stamps, to highway locations, to spraying of forests or grasslands to eliminate certain species of trees or shrubs, to the location of missile bases, to the disposal of sewage or industrial wastes, to the granting of off-shore oil leases.").

6. *See* ANDREW F. POPPER & GWENDOLYN M. MCKEE, ADMINISTRATIVE LAW 2 (1st ed. 2009) ("Unelected administrative officials can announce standards that interpret statutes and shift significantly interests and entitlements . . . .").

7. The Administrative Procedure Act (APA) instructs courts to "set aside agency action" which is "arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law." 5 U.S.C. § 706(2)(A) (2006).

8. 129 S. Ct. 1800 (2009).

9. That is, an "isolated use of an offensive expletive." Dave E. Hutchinson, *"Fleeting Expletives" Are the Tip of the Iceberg: Fallout from Exposing the Arbitrary and Capricious Nature of Indecency Regulation*, 61 FED. COMM. L.J. 229, 231 (2008).

10. 5 U.S.C. §§ 551–559.

11. *See* FCC v. Pacifica Found., 438 U.S. 726, 729 (1978) (declaring that the Federal Communications Commission (FCC) holds the "power to regulate a radio broadcast that is indecent but not obscene").

12. *See, e.g.*, *In re* Indus. Guidance on the Comm'n's Case Law Interpreting 18 U.S.C. § 1464 & Enforcement Policies Regarding Broad. Indecency, 16 F.C.C.R. 7999, 8008 (2001)

This regime changed in 2004 when the FCC decided repetition would no longer be a requisite factor leading to an indecency finding.[13] Fox Television challenged the FCC's new enforcement regime.[14] The network raised numerous issues with the policy,[15] but the Second Circuit reached only one of them: finding the FCC's change in policy arbitrary and capricious under the APA, the court instructed the FCC to proffer a "reasoned analysis" that could survive APA review.[16]

The Supreme Court reversed the Second Circuit in a 5–4 ruling.[17] Whereas the Second Circuit felt the FCC failed to supply a "reasoned basis"[18] for its shift, the Supreme Court disagreed: "The Commission could rationally decide it needed to step away from its old regime where nonrepetitive use of an expletive was *per se* nonactionable because that was 'at odds with the Commission's overall enforcement policy.'"[19] Like *New York v. United States*,[20] the facts and underlying dispute in *Fox II* may well fade into obscurity; nonetheless, the central holding of this case—setting a very low threshold for agencies to clear before reversing or rescinding existing policies—will remain on the books, perhaps waiting to be picked up and trumpeted by an overzealous administrative body.[21]

Arguably, the Court granted certiorari to clear up some uncertainty among the lower courts with respect to the appropriate standard of review to be utilized when an agency reverses itself.[22] Section 706(2)(A) of the APA states that a court may set aside agency action that is arbitrary and capricious.[23] Just what level of scrutiny the test entails has confused some courts—the Supreme Court itself was accused of sending "conflicting signals" for a number of years.[24] In a prior opinion, the Court defined the

("[W]here sexual or excretory references have been made once or have been passing or fleeting in nature, this characteristic has tended to weigh against a finding of indecency.").

13. *Fox II*, 129 S. Ct. at 1807.

14. Fox Television Stations, Inc. v. FCC (*Fox I*), 489 F.3d 444 (2d Cir. 2007), *rev'd*, *Fox II*, 129 S. Ct. 1800 (2009).

15. *See* 489 F.3d at 454 (listing seven arguments against the validity of the policy).

16. The court made its offer with the caveat that no matter how reasoned the new rationale might be, it was nevertheless unlikely to survive *constitutional* review. *Id.* at 462, 467.

17. *Fox II*, 129 S. Ct. at 1805.

18. *Fox I*, 489 F.3d at 447.

19. *Fox II*, 129 S. Ct. at 1813 (quoting *In re* Complaints Regarding Various Television Broads. Between Feb. 2, 2002 & Mar. 8, 2005, 21 F.C.C.R. 2664 (2006)).

20. 342 U.S. 882 (1951).

21. *See Fox II*, 129 S. Ct. at 1832 (Breyer, J., dissenting) (warning that the Court's ruling would "change judicial review . . . and not in a healthy direction").

22. *See* Hutchinson, *supra* note 9, at 240 (discussing the "apparent confusion regarding the scope and standard of arbitrary and capricious review").

23. 5 U.S.C. § 706(2)(A) (2006).

24. Lisa Schultz Bressman, *Judicial Review of Agency Discretion*, *in* A GUIDE TO JUDICIAL AND POLITICAL REVIEW OF FEDERAL AGENCIES 177, 178 (John F. Duffy & Michael Herz

inquiry as "searching and careful,"[25] yet also cautioned that the test is a "narrow one."[26] Some courts, in turn, have taken advantage of the inconsistency by favoring one approach over the other (i.e., by adhering to the narrow approach or, on the other hand, by engaging in a broader review).[27] In any event, if there was any doubt as to what an agency had to show before it could change extant regulations, the Supreme Court has now provided an answer: Not very much at all.

I. FACTUAL BACKGROUND

While accepting a 2003 Golden Globe Award, Bono, the lead singer of the rock band U2, declared "[t]his is really, really, f***ing brilliant" in front of over twenty million television viewers.[28] This incident ultimately resulted in the FCC concluding—for the first time ever—that a so-called "nonliteral . . . use of the F- and S-Words could be actionably indecent, even when the word is used only once."[29]

The FCC determined that henceforth, any use of the F-word—because the word "inherently has a sexual connotation"—would fall within its regulatory reach.[30] Thus, even when the F-word is used as nothing more than an "intensifier" (as Bono supposedly employed it), the material would nonetheless qualify as indecent.[31] Further, that the word was used only once (i.e., in a fleeting way) would no longer be dispositive.[32] As the FCC noted, "The mere fact that specific words or phrases are not sustained or repeated does not mandate a finding that material that is otherwise patently offensive to the broadcast medium is not indecent."[33] Because this new approach was a departure from its previous policy, the FCC declined to impose any forfeiture penalties as it recognized "existing precedent would have permitted this broadcast."[34] Nevertheless, the networks were now on

---

eds., 2005).

25. Citizens to Pres. Overton Park, Inc. v. Volpe, 401 U.S. 402, 416 (1971).

26. *Id.*

27. *See generally* Bressman, *supra* note 24, at 178 (describing the "tension" arising from the Court's seemingly bipolar treatment of the matter).

28. *Fox II*, 129 S. Ct. 1800, 1807 (2009) (internal quotation marks omitted).

29. *Id.*

30. *In re* Complaints Against Various Broad. Licensees Regarding Their Airing of the "Golden Globe Awards" Program, 19 F.C.C.R. 4975, 4978 (2004).

31. *Id.*

32. *See id.* at 4980 (announcing that those "cases holding that [the] isolated or fleeting use of the 'F-Word' or a variant thereof . . . is not indecent" were no longer "good law").

33. *Id.*

34. *Fox I*, 489 F.3d 444, 452 (2d Cir. 2007) (citing *In re* Complaints Against Various Broad. Licensees Regarding Their Airing of the "Golden Globe Awards" Program, 19 F.C.C.R. 4975, 4981 (2004)).

notice.

On March 15, 2006, the FCC released an order responding to the "concerns" of broadcasters and viewers.[35] Seeking to "provide substantial guidance to broadcasters and the public," the FCC's report examined a number of "factual patterns" (i.e., examples of broadcasts possibly constituting indecency).[36] One of the incidents examined in the order involved another entertainer receiving an award who also expressed her emotions through colorful language. On December 9, 2002, over nine million people watched on television as Cher received an "Artist Achievement Award."[37] During her acceptance speech, Cher took a shot at her detractors, saying, "[s]o f[***] 'em. I still have a job and they don't."[38] The FCC again declared that "any use of [the F-word] inherently has a sexual connotation."[39] The order referred to the Bono incident when it noted that lack of repetition no longer shielded "otherwise patently offensive" material from an indecency finding.[40] No sanctions were imposed as a result of this incident, however, because the event took place prior to the Bono incident.[41]

One year after the Cher incident, the 2003 Billboard Music Awards broadcast to ten million viewers. This time, it would not be the entertainers receiving the awards who would incur the wrath of the FCC but the entertainers presenting them. Paris Hilton and Nicole Richie, who "play themselves as two spoiled, rich young women"[42] on "reality" television, were selected to present an award during the ceremony.[43] The women were supposed to follow a scripted monologue; however, Richie took some artistic liberties with her lines. Where Richie was supposed to rhetorically ask the audience, "Have you ever tried to get cow manure out of a Prada purse? It's not so freaking simple," she instead queried, "Have you ever tried to get cow s[***] out of a Prada purse? It's not so f[***]ing simple."[44]

---

35. *In re* Complaints Regarding Various Television Broads. Between Feb. 2, 2002 & Mar. 8, 2005, 21 F.C.C.R. 2664, 2665 (2006).

36. *Id.*

37. Brief for Federal Communications Commission & United States at 11, *Fox I*, 489 F.3d 444 (2d Cir. 2007) (No. 06-1760-ag).

38. *Id.*

39. *In re* Complaints Regarding Various Television Broads. Between Feb. 2, 2002 & Mar. 8, 2005, 21 F.C.C.R. at 2691.

40. *Id.* (quoting *In re* Complaints Against Various Broad. Licensees Regarding Their Airing of the "Golden Globe Awards" Program, 19 F.C.C.R. 4975, 4980 (2004)).

41. *Id.* at 2692.

42. *Fox I*, 489 F.3d 444, 468 (2d Cir. 2007) (Leval, J., dissenting).

43. Brief for Federal Communications Commission & United States at 12, *Fox I*, 489 F.3d 444 (2d Cir. 2007) (No. 06-1760-ag) (citing *In re* Complaints Regarding Various Television Broads. Between Feb. 2, 2002 & Mar. 8, 2005, 21 F.C.C.R. at 13,303).

44. *Id.* at 12–13 (citing *In re* Complaints Regarding Various Television Broads. Between

This incident was also analyzed in the 2006 report.[45] Once again, the FCC reiterated its earlier statement as to the F-word's supposedly inherent sexual connotation. Similarly, the order viewed Richie's use of the S-word as "invariably invok[ing] a coarse excretory image."[46] And, as before, the lack of repetition did not weigh against a finding of indecency.[47] Finally, the order noted the "shocking and gratuitous" nature of Richie's dialogue.[48] The FCC again declined to impose sanctions.[49]

The Second Circuit determined that the validity of the FCC's new indecency regime was arbitrary and capricious because the FCC's new direction "represent[ed] a dramatic change in agency policy without adequate explanation."[50] The Second Circuit singled out three reasons why the FCC's stated rationale for its new approach did not pass muster under arbitrary and capricious review. First, the court rejected the FCC's argument that were the FCC to maintain the previous policy, viewers would be forced to take a "harmful first blow."[51] Second, the court also dismissed the FCC's first-blow argument because the theory had no "rational connection" to the revised enforcement approach.[52] Finally, the court described the FCC's prediction that, without the new approach, broadcasters would inevitably begin to barrage the airwaves with fleeting expletives "so long as they did so one at a time" as "divorced from reality."[53] Thus, the court concluded that the FCC's proffered reasons failed to comprise the necessary "reasoned analysis justifying its departure."[54] The Supreme Court, however, reversed the Second Circuit's decision.

II. THE SUPREME COURT'S ANALYSIS

In reversing the Second Circuit, the Court held that the FCC's decision to go against its previous fleeting expletives regime and sanction as indecent "offensive words" that "are not repeated" was not an arbitrary and

---

Feb. 2, 2002 & Mar. 8, 2005, 21 F.C.C.R. at 13,303, 13,311).

45. *In re* Complaints Regarding Various Television Broads. Between Feb. 2, 2002 & Mar. 8, 2005, 21 F.C.C.R. at 2692.

46. *Id.* at 2693.

47. *Id.*

48. *Id.* at 2694.

49. *See id.* at 2695 (declining penalties because the precedent at the time of the broadcast was to the contrary).

50. *Fox I*, 489 F.3d 444, 454 (2d Cir. 2007) (majority opinion).

51. *Id.* at 458.

52. *Id.*

53. *Id.* at 460 (quoting *In re* Complaints Regarding Various Television Broads. Between Feb. 2, 2002 & Mar. 8, 2005, 21 F.C.C.R. at 13,309).

54. *Id.* at 462.

capricious change in policy.[55] The Court concluded that the FCC acted in accordance with the APA because, first, it acknowledged the change,[56] and second, the FCC's supplied rationale for changing its policy was "entirely rational."[57] Declining to extend its review beyond what was appealed, the Court passed on the serious constitutional questions surrounding the FCC's new direction.[58]

As this case turned on an administrative law question, the Court looked to the language of the APA. Additionally, because the administrative law issue revolved specifically around the reversal of an agency's policy, the Court turned to the leading case on that topic, *Motor Vehicle Manufacturers Ass'n of the U.S., Inc. v. State Farm Mutual Automobile Insurance Co.*[59] Noting that neither the APA nor *State Farm* calls for heightened scrutiny above what the normal arbitrary and capricious standard entails, the majority dismissed any notion that courts should engage in a "more searching review" of agency changes.[60] Whether a court is examining original agency action or agency change, all that is required is a "satisfactory explanation for [that] action."[61] Justice Scalia, writing for the majority, explained what would constitute such an explanation:

> [T]he requirement that an agency provide reasoned explanation for its action would ordinarily demand that it display awareness that it *is* changing position. . . . And of course the agency must show that there are good reasons for the new policy. But it need not demonstrate to a court's satisfaction that the reasons for the new policy are *better* than the reasons for the old one; it suffices that the new policy is permissible under the statute, that there are good reasons for it, and that the agency *believes* it to be better, which the conscious change of course adequately indicates.[62]

In other words, the Court's test for scrutinizing agency changes under arbitrary and capricious review reads something like the following:

---

55. *Fox II*, 129 S. Ct. 1800, 1805, 1812 (2009).

56. *Id.* at 1812.

57. *Id.*

58. *See id.* at 1819 ("This court . . . is one of final review, 'not of first view.'" (quoting Cutter v. Wilkinson, 544 U.S. 709, 718 n.7 (2005)).

59. 463 U.S. 29 (1983).

60. *Fox II*, 129 S. Ct. at 1810. *But see* Verizon Commc'ns, Inc. v. FCC, 535 U.S. 467, 502 n.20 (2002) (distinguishing *State Farm* in that it "may be read as prescribing more searching judicial review").

61. *Fox II*, 129 S. Ct. at 1810 (quoting *State Farm*, 463 U.S. at 43).

62. *Id.* at 1811.

(1) The agency must explicitly acknowledge its change in policy; and

(2) The agency must give good reasons to support the change, which turns on whether—

(A) The change is in accordance with the agency's organic statute; and

(B) The agency believes it to be better than the prior approach.

As the majority notes, Step (2)(B) is basically self-fulfilling, so its consideration seems irrelevant to the final analysis. Essentially, an agency attempting to show that its policy change satisfies arbitrary and capricious review has a fairly easy task. Step 1 is easily accomplished—either the agency acknowledges its change or it does not. Step 2(A) is also fairly straightforward—the agency cannot violate existing law. Step 2(B), again, is apparently automatically satisfied by the change. Thus, the only portion of the Court's standard that seems open to discussion is the requirement that the agency supply "good reasons." If any of the steps might occasion litigation, this is probably it; indeed, what is good enough for a good reason?

A good reason, first of all, does not equate to "good enough" in the eyes of the reviewing judge.[63] And a good reason may require no more than the justification necessary for an original agency action.[64] Applying its newly delineated test to the FCC's revised approach, the Court accepted the FCC's view that any version (i.e., literal or nonliteral) of the F-word inherently possesses a "sexual meaning" as a good reason for its change in policy.[65] Additionally, the Court noted it was "surely rational" for the FCC to predict that continuing its prior policy—which, according to the Court, essentially amounted to a safe harbor for fleeting expletives—would lead to an increased presence of such language over the airwaves.[66] The majority observed that the FCC's overall approach to indecency regulation turned primarily on context; thus, an automatic exemption for a single expletive was "at odds with the Commission's overall enforcement policy."[67] Finally, the Court saw "technological advances" as weighing in favor of tougher enforcement; with the relative ease broadcasters today have in blocking

63. *See id.* ("[The agency] need not demonstrate to a court's satisfaction that the reasons for the new policy are *better* than the reasons for the old one . . . .").

64. On the other hand, the majority does note that when an agency turns its back on previous "factual findings," or if its new approach jeopardizes "serious reliance interests," these factors must be considered by the agency; however, the Court reiterates that this does not mean any "further justification" is needed, just "reasoned explanation." *Id.*

65. *Id.* at 1812.

66. *Id.* at 1812–13.

67. *Id.* at 1813 (quoting *In re* Complaints Regarding Various Television Broads. Between Feb. 2, 2002 and Mar. 8, 2005, 21 F.C.C.R. 2664, 13,308, ¶ 23 (2006)) (internal quotation marks omitted).

offensive language, any expletive—fleeting or not—should be a rare occurrence on the airwaves.[68]

Notably, even if the change is of questionable constitutional validity, this will not upset the Court's arbitrary and capricious analysis.[69] *FCC v. Pacifica Foundation*[70] delineated the constitutional scope of the FCC's enforcement regime; while that scope expanded slightly over time to include words other than only those uttered in the George Carlin monologue at issue in the case, judicial decisions from *Pacifica* forward exhibited an expectation that the FCC would tread cautiously and that the occasional, isolated use of a single expletive would not incur liability.[71] The FCC's new indecency enforcement approach deviates from these expectations, so the approach arguably deviates from the constitutionally acceptable to perhaps the unconstitutional.[72] This fact, however, did not influence the Court's arbitrary and capricious review. The Court, while acknowledging that the constitutionality of the new policy may be open to challenge, did not see that fact as having any bearing on the arbitrary and capricious analysis.[73]

After establishing the governing principles and applying them to the FCC's action, Justice Scalia devoted the rest of the opinion to scrutinizing

---

68. *Id.*

69. Traditionally, agency statutory interpretations and policy rationales are entitled to judicial deference. *See, e.g.*, Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837 (1984). Fox argued that this situation should not be entitled to the usual level of deference (*Chevron* deference) because it was a special case with constitutional issues inextricably intertwined with the administrative law question. *See* Brief for Respondent Fox Television Stations, Inc. at 19, *Fox II*, 129 S. Ct. 1800 (2009) (No. 07-582) ("Simply put, the First Amendment trumps *Chevron*."). Yet contrary to what Fox suggested, the plurality cabined its opinion very tightly. *Fox II*, 129 S. Ct. at 1817–18 (plurality opinion). Apparently the First Amendment does not "trump *Chevron*."

70. 438 U.S. 726 (1978).

71. *See, e.g.*, *id.* at 761 n.4 (Powell, J., concurring) ("[S]ince the Commission may be expected to proceed cautiously, as it has in the past, I do not foresee an undue 'chilling' effect on broadcasters' exercise of their rights." (citation omitted)); *see also, e.g.*, Action for Children's Television v. FCC, 852 F.2d 1332, 1340 n.14 (D.C. Cir. 1988) ("[T]he FCC has assured this court, at oral argument, that it will continue to give weight to reasonable licensee judgments when deciding whether to impose sanctions in a particular case. Thus, the potential chilling effect of the FCC's generic definition of indecency will be tempered by the Commission's *restrained enforcement policy*." (citation omitted) (emphasis added)).

72. *Cf. Fox II*, 129 S. Ct. at 1827 (Stevens, J., dissenting) ("The narrow treatment of the term 'indecent' in *Pacifica* defined the outer boundaries of the enforcement policies adopted by the FCC in the ensuing years.").

73. *See id.* at 1812 (majority opinion) (noting that the APA provides a separate section for unlawful agency action, including unconstitutional action). *But cf.* Raoul Berger, *Administrative Arbitrariness and Judicial Review*, 65 COLUM. L. REV. 55, 83 (1965) ("The fact that arbitrariness tinged with racial or religious factors offends still other constitutional guarantees may make courts more alert to the slightest trace of arbitrariness in that area.").

both the Second Circuit's reasoning and the dissent's arguments.[74] The Court's repudiation of the Second Circuit's reasoning is helpful primarily in that it provides further explanation for what would pass for "good reasons." Recall that the Second Circuit found the first-blow theory lacking because the FCC "fail[ed] to explain why it had not previously banned fleeting expletives as 'harmful first blow[s].'"[75] The FCC did not proffer any evidence showing that a fleeting expletive is harmful enough to "warrant government regulation."[76] The majority countered that not every subject of an administrative agency's regulation will be as easily quantifiable and reducible to a proper analysis as, say, the effect of airbags on the rate of traffic accident fatalities.[77] Additionally, the majority argued that the Second Circuit was wrong to demand evidentiary support on a topic "for which scant empirical evidence can be marshaled." Indeed, for the majority, it seemed satisfactory to allow the FCC to rely on the common perception "that children mimic the behavior they observe."[78]

Again, the Second Circuit's biggest problem with the first-blow theory was not so much that the FCC failed to adopt it in the past but that it seemed to be in conflict with the FCC's insistence that context was all-important.[79] If the FCC's first-blow theory is taken to its logical extreme, then a per se ban on fleeting expletives would seem to be called for.[80] Yet throughout the proceedings the FCC stressed the importance of context in the indecency analysis. The agency did not, for instance, find the word "bulls[**]tter"—uttered during a live news broadcast—to be indecent under its revised approach.[81] The Second Circuit found this fatally contradictory.[82] The Supreme Court, on the other hand, saw the Second

74. *Fox II*, 129 S. Ct. at 1813–15 (majority opinion); *id.* at 1815–19 (plurality opinion).

75. *Id.* at 1813 (majority opinion) (quoting *Fox I*, 489 F.3d 444, 458 (2d Cir. 2007)) (alteration in original).

76. *See Fox I*, 489 F.3d at 461 (emphasizing that when an agency changes a previously settled view, the agency must provide a reasoned basis for that change).

77. *See Fox II*, 129 S. Ct. at 1813 ("One cannot demand a multiyear controlled study, in which some children are intentionally exposed to indecent broadcasts (and insulated from all other indecency), and others are shielded from all indecency.").

78. *See id.* (reiterating that Congress has decided to let the FCC enforce the ban on indecent material that is harmful to children).

79. *Fox I*, 489 F.3d at 458.

80. *Cf.* Brief for Respondent Fox Television Stations, Inc. at 13, *Fox II*, 129 S. Ct. 1800 (2009) (No. 07-582) ("The first blow theory ma[kes] sense only if the FCC presumed that mere exposure to potentially offensive language harmed the broadcast audience.").

81. *See id.* ("The FCC . . . permitted some isolated and fleeting expletives if, for example, they occurred during a '*bona fide* news interview' . . . .").

82. *See Fox I*, 489 F.3d at 459 n.9 (expressing doubt as to the logical consistency of the FCC's new approach).

Circuit's reasoning as misdirected.[83] The agency has to maintain some amount of discretion with its enforcement responsibilities, according to the majority. Indeed, they had the same level of discretion under the previous policy.[84]

Finally, the Second Circuit balked at the FCC's prediction that, absent the new approach, networks might begin to barrage the airwaves with fleeting expletives.[85] For the Supreme Court majority, however, deduction and past experience apparently play negligible roles in determining an agency's "good reasons." If the agency's estimation is *theoretically* possible, this is, presumably, perfectly acceptable as a good reason in support of a change.[86] In closing, the majority remarked that both the "pervasiveness of foul language" and the growth in alternative forums where indecency restrictions play no part (i.e., cable television and the Internet) provided further good and rational reasons why the agency needed to change.[87]

III. IMPLICATIONS

Fox II is an important case for several reasons. The facts underlying the dispute involve potentially substantial constitutional law implications. This case stands a very good chance of coming back to the Court, possibly as a vehicle for revisiting *Pacifica*.[88] For now, though, the ruling's primary impact would seem to be in the field of administrative law. The opinion could equate, over time, to greater judicial deference toward agencies when they act to upset long-existing policies.[89] Whether one agrees with the majority's ruling or not, the opinion is helpful in at least one respect: There should no longer be much ambiguity surrounding the appropriate standard of review when an agency's decision to change its policies is challenged.[90]

---

83. *See Fox II*, 129 S. Ct. at 1814 ("Any complaint about the Commission's failure to ban only some fleeting expletives is better directed at the agency's context-based system generally rather than its inclusion of isolated expletives.").

84. *Id.*

85. *See Fox I*, 489 F.3d at 460 n.11 (observing that the theory "is both unsupported by any evidence and directly contradicted by prior experience").

86. *Fox II*, 129 S. Ct. at 1814 ("Even in the absence of evidence, the agency's predictive judgment (which merits deference) makes entire sense.").

87. *Id.* at 1819.

88. *See id.* ("It is conceivable that the Commission's orders . . . [are] beyond the Commission's reach under the Constitution. Whether . . . it is unconstitutional [] will be determined soon enough, perhaps in this very case.").

89. *See, e.g.*, Brief of Federal Appellees at 44 n.16, Humane Soc'y v. Gutierrez, 558 F.3d 896 (9th Cir. 2009) (No. 08-36038) (noting that "the proposition that a change in agency interpretation must be supported by a reasoned analysis over and above that required for an interpretation in the first instance" has ceased to be "good law." (internal quotation marks omitted)).

90. *See* David L. Hudson, Jr., *Was the FCC's Change in Policy Regarding Broadcast Expletives*

The Court has now made clear that an administrative agency changing its policies should face little resistance from the Judiciary.[91]

Judicial review in this area is now very deferential; agencies must act in accordance with law and acknowledge the change, but once these requirements are satisfied the only thing left for the agency to do is provide good reasons for changing.[92] And these reasons—it would seem from the opinion—can be contradictory to experience and unsupported by empirical evidence yet still be acceptable as adequate justification when an agency turns its back on prior policies.

A. Judicial Deference to Agency Predictions

For instance, the Court accepted the agency's prediction that, without the new indecency regime, networks might begin exploiting the exemption for fleeting expletives by "barrag[ing] the airwaves" with isolated incidents involving indecent words.[93] The Court conceded that in nearly thirty years this had not happened.[94] Yet this did not weigh in favor of a finding of arbitrary change. Because the agency's prediction is *theoretically* possible, the rationale is acceptable.[95]

In other words, deference borders on ignorance.[96] Courts are instructed to ignore whether the agency's fear has been realized in the past and defer to agency predictions *if* the prediction *could* occur. It is understandable that the Court thinks it best to defer to an agency's "predictive judgment."[97] After all, some agencies are involved in regulating complex subject matter

---

Arbitrary and Capricious?, 36 PREVIEW U.S. SUP. CT. CASES 141 (2008) (suggesting that *Fox II* would be helpful because it would give "more guidance in this post-*Chevron* era as to just how much deference administrative agencies receive when their policies impact constitutional rights").

91. *See* Posting of Jonathan Adler to the Volokh Conspiracy, http://volokh.com/posts/1240966018.shtml (Apr. 28, 2009, 20:46 EST) ("Some courts have read a prior Supreme Court case to require more evidence and explanation when an agency is shifting policy. The Court rejected this view. . . . Justice Scalia's opinion . . . make[s] such shifts by agencies easier and at least at the margins should improve the agency's chances of surviving judicial review.").

92. *See Fox II*, 129 S. Ct. at 1811 (making the additional point that the agency is not required to show the court that the rationale underlying the new policy change is better than that for the previous one).

93. *Id.* at 1814 (internal quotation marks omitted).

94. *See id.* (postulating that this might have been due to the fact that "its prior permissive policy had been confirmed (save in dicta) only at the staff level").

95. *Id.*

96. *Cf.* Volkswagenwerk Aktiengesellschaft v. Fed. Mar. Comm'n, 390 U.S. 261, 272 (1968) ("The deference owed to an expert tribunal cannot be allowed to slip into a judicial inertia . . . ." (alteration in original)).

97. *Fox II*, 129 S. Ct. at 1814.

which many judges (and even Justices) may not fully understand. Few would find it appropriate for a court to thoroughly scrutinize an agency's decision on a "scientific determination," for instance.[98] Were a court to do so, the chance that it might impermissibly "substitute its judgment for that of the agency"[99] seems great because of the possibility that the court is, frankly, not qualified to make the necessary determinations. Where, however, the issue is not complex and simply involves the use of deduction,[100] the chance that a court would substitute its judgment seems far less likely. For instance, in *State Farm*, while it was perhaps theoretically possible that the inclusion of automatic safety belts would not lead to a decrease in traffic fatalities, the agency's predictive judgment on this matter (which is at least equally if not more complex than indecency) did not warrant the Court's deference.[101]

B. Does Arbitrary and Capricious Review Equate to Rational Basis Review?

The Court's treatment of the agency's predictions and assumptions seems incompatible with *State Farm* for another reason as well. Upholding agency predictions on the grounds that they theoretically could occur (despite the fact that the converse actually occurred) and allowing an agency to rely on assumptions (despite the fact that contrary evidence exists) seems to imply similarities to the rational basis analysis courts utilize to review legislation.[102] But *State Farm* dismissed the idea that rational basis review and arbitrary and capricious analysis were alike.[103] Arguably, when the *State Farm* Court distinguished rational basis review from arbitrary and capricious review, it meant to suggest the former affords more deference than the latter; it would seem odd to imply that arbitrary and capricious review entails *greater* deference than rational basis review.[104] Now, though,

98. Baltimore Gas & Elec. Co. v. Natural Res. Def. Council, Inc., 462 U.S. 87, 103 (1983).

99. *Fox II*, 129 S. Ct. at 1810 (quoting Motor Vehicle Mfrs. Ass'n of the U.S., Inc. v. State Farm Mutual Auto. Ins. Co., 463 U.S. 29, 43 (1983)).

100. For example, in the absence of *X*, *Y* does not occur for thirty years. Thus, *Y* probably will not occur even if agency *A* fails to promulgate *X*.

101. *See, e.g.*, *State Farm*, 463 U.S. at 54 ("[S]tatements that passive belts will not yield substantial increases in seatbelt usage apparently take no account of the critical difference between detachable automatic belts and current manual belts.").

102. *See, e.g.*, Williamson v. Lee Optical of Oklahoma, Inc., 348 U.S. 483, 487–88 (1955) ("[T]he law need not be in every respect logically consistent with its aims to be constitutional. It is enough that there is an evil at hand for correction, and that *it might be thought* that the particular legislative measure was a rational way to correct it." (emphasis added)).

103. *State Farm*, 463 U.S. at 43 n.9.

104. *See* Alan B. Morrison, *Administrative Agencies Are Just Like Legislatures and Courts—Except*

the two seem roughly equivalent; under the deferential approach adopted by the Court, agencies are closer in stature to legislatures than ever before.

CONCLUSION

Practitioners should understand what *Fox II* means for administrative law. Where a person or an entity—either of whom is subject to an administrative agency's regulatory reach—seeks to challenge the reversal or rescission of an agency's policies, the prospects for successful prosecution of such a claim have been lessened.[105] This is not to say that a challenge to a regulatory change will never be successful; however, the requirements agencies must meet are few and easily satisfied.

When They're Not, 59 ADMIN. L. REV. 79, 96 (2007) ("[Agencies] are not forbidden from changing their minds, but if they do so, they must explain why—or at least how they reconcile the disparate treatment of similar matters. *Congress is under fewer restraints in terms of consistency* . . . ." (emphasis added)).

105. Already, some do seem aware of the case's impact. On the website for the law firm Wiley Rein LLP, the firm suggests that a better approach to contesting agency change in the future would be to challenge "flaws in statutory interpretation" rather than attempting to persuade a court to "second-guess agency policy judgments." *See* Bert W. Reign & Thomas W. Queen, Wiley Rein LLP, Administrative Law Bulletin: *FCC v. Fox*—The Supreme Court Gives the Green Light to Regulatory Change (May 7, 2009), http://www.wileyrein.com/publications.cfm?sp=articles&id=5135.