YALE JOURNAL OF HEALTH POLICY, LAW, AND ETHICS

VOLUME XVIII Summer 2018

TABLE OF CONTENTS

The Problem of Intra-Personal Cost

Brian Galle*

Abstract:

"Externalities," or harms to others, provide a standard justification for government intervention in the private market. There is less agreement over whether government is justified in correcting "internalities," or harms we inflict on our own health or well-being. While some of the internality dispute is philosophical, some is practical. Critics suggest government lacks information to regulate internalities, and that any intervention would inefficiently distort a private market for self-help. This Article argues that these critiques of regulation overlook well-established tools of externality regulation, as well as a burgeoning literature on the measurement of internalities.

Having answered the "should" question, the Article moves on to "how?" It examines the established tools of externality regulation and considers to what extent the standard advice of the externality literature extends to internality regulation. In departures from earlier consensus, the analysis suggests that "carrots" may at times be an attractive alternative to "sticks," and that even large taxes on internalities can produce a so-called "double dividend." The Article also compares traditional regulatory options to "nudges" and other forms of cognitively-informed government interventions. It identifies a set of cases in which nudges may be preferable to either taxes or command and control regulation.

Thus, this Article's analysis also helps to resolve a second, related, debate over the propriety of nudges. The nudge debate has almost exclusively revolved around whether nudges avoid philosophical objections to paternalistic government regulation. This Article offers instead a new reason to employ nudges in some cases: they are more efficient.

---

* Professor, Georgetown University Law Center. I am grateful for helpful comments and suggestions from Gregg Bloche, Jacob Goldin, Jim Hines, Louis Kaplow, Saul Levmore, Katie Pratt, Chris Robertson, Ben Roin, Carol Rose, Darien Shanske, Peter Siegelman, Jessica Silbey, and attendees of presentations at Arizona University Law School, Boston College Law School, Loyola-L.A. Law School, the University of Connecticut Law School, the U.C. Berkeley Burch Center Colloquium on Public Finance, the Murphy Institute for Public Finance at Tulane University, the Boston Area Summer Research Group, and the annual meeting of the National Tax Association. The editorial staff of the YJHPLE provided excellent substantive advice and copy editing.

I.

INTRODUCTION

Governments properly act to protect people from one another, or so political philosophers have long agreed.[1] Increasingly, the modern regulatory state also steps in to protect us from ourselves. Obviously, regulation of tobacco, opiates, and other addictive drugs falls into this category, but so too can "fat taxes,"[2] social security and retirement savings policy,[3] the regulation of consumer financial products,[4] internet privacy rules,[5] the design of crop and flood insurance programs,[6] government oversight of workplace safety and health,[7] mandatory vaccinations (which protect not only children and their classmates but also parents who would otherwise suffer when their child contracts a terrible disease), and many others. Evidence suggests waiting periods to purchase firearms may best be justified as a policy to reduce suicide,[8] and one might say much the same about laws requiring motorcyclists to wear helmets.[9]

To be sure, "paternalism" is not new. The possibility that government might help us avoid these kinds of regrettable decisions dates back at least to Aristotle and possibly Homer, depending on how metaphorically one wants to read the Odyssey.[10]

But as vast and culturally pervasive as the paternalism literature has become, it has tended toward the philosophical, lingering on the propriety of government intervention to correct self-harms.[11] A decade, for instance, after Sunstein &

---

1 *E.g.*, JOHN LOCKE, SECOND TREATISE ON GOVERNMENT Ch.9 §§ 123–31 (1696).

2 Jeff Strnad, *Conceptualizing the "Fat Tax": The Role of Food Taxes in Developed Economies*, 78 S. CAL. L. REV. 1221, 1244–58 (2005).

3 Deborah M. Weiss, *Paternalistic Pension Policy: Psychological Evidence and Economic Theory*, 58 U. CHI. L. REV. 1275, 1278–82 (1991).

4 John Y. Campbell, Howell Jackson, & Bridget Madrian, *Consumer Financial Protection*, 25 J. ECON. PERSPECTIVES 91, 92–106 (2011); Oren Bar-Gill & Elizabeth Warren, *Making Credit Safer*, 157 U. PA. L. REV. 1, 7– 25(2008).

5 Katherine J. Strandburg, *Privacy, Rationality, and Temptation: A Theory of Willpower Norms*, 57 RUTGERS L.J. 1235, 1260–68, 1282–99 (2005).

6 HOWARD C. KUNREUTHER, MARK V. PAULY & STACEY MCMORROW, INSURANCE AND BEHAVIORAL ECONOMICS 114–26 (2013).

7 Christine Jolls, *Employment Law*, *in* 2 RESEARCH HANDBOOK ON LAW & ECONOMICS 1349, 1354–56 (Mitchell Polinsky & Steven Shavell eds. 2007).

8 *See* Jens Ludwig & Philip Cook, *Homicide and Suicide Rates Associated with Implementation of the Brady Handgun Violence Prevention Act*, 284 JAMA 585, 586–91 (2000) (finding that waiting periods reduced suicide but not homicide).

9 David J. Houston & Lilliard E. Richardson, *Motorcyclist Fatality Rates and Mandatory Helmet-Use Laws*, 40 ACCIDENT ANAL. & PREVENTION 200, 201–08 (2008) (finding fatality rates up to 33% lower in mandatory-helmet states).

10 *See* JON ELSTER, ULYSSES UNBOUND: STUDIES IN RATIONALITY, PRECOMMITMENT, AND CONSTRAINTS 3, 8 (2000).

11 *E.g.*, JULIAN LE GRAND & BILL NEW, GOVERNMENT PATERNALISM: NANNY STATE OR HELPFUL FRIEND? 105–82 (2015).

Thaler first introduced the idea that government might "nudge" us toward better decisions,[12] the nudge debate remains caught up in cycles of argument over whether nudging is consistent with libertarian values.[13]

This is a frustrating state of affairs for those who are relatively comfortable with government intervention in the marketplace.[14] Human failings and new developments in how they can be addressed raise difficult and important questions, none so far addressed comprehensively in the existing literature.[15] Many governments already are deeply committed to helping consumers overcome what the governments perceive to be poor choices.[16] Canada, Australia, and many other countries around the world regulate tobacco with a complex regime in which manufacturers cannot display brand information, and instead must print disturbing images illustrating the health consequences of smoking.[17] The United States has proposed a similar policy, which currently is

---

12 Cass R. Sunstein & Richard Thaler, *Libertarian Paternalism is Not an Oxymoron*, 70 U. CHI. L. REV. 1159, 1190–95 (2003); *see generally* RICHARD THALER & CASS SUNSTEIN, NUDGE (rev. & expanded ed. 2009).

13 Cass R. Sunstein, *Behavioral Economics and Paternalism*, 122 YALE L.J. 1826, 1867–97 (2013); *Symposium, The Ethics of Nudging: Evaluating Libertarian Paternalism*, 14 GEORGETOWN J.L. & PUB. POL'Y 645 et seq. (2016); Christian Coons & Michael Weber, *Introduction, in* PATERNALISM: THEORY AND PRACTICE 1, 15–23 (Christian Coons & Michael Weber eds., 2013).

14 On Amir & Orly Lobel, *Stumble, Predict, Nudge: How Behavioral Economics Informs Law and Policy*, 108 COLUM. L. REV. 2098, 2127–32 (2008); Ryan Bubb & Richard H. Pildes, *How Behavioral Economics Trims Its Sails and Why*, 127 HARV. L. REV. 1593, 1596–98 (2014); Lauren E. Willis, *When Nudges Fail: Slippery Defaults*, 80 U. CHI. L. REV. 1155, 1227–29 (2013). Bubb & Pildes argue that policy makers should "analyze [nudges] much as we would analyze explicit mandates," and make a "full comparison of the advantages and disadvantages of different regulatory instruments," *supra* at 1601, but they do not engage in that analysis themselves. This Article does.

15 The notable major exception to the absence of substantive analysis of internality regulation is a short recent policy brief focused on energy use by Sunstein and Hunt Allcott, an NYU economist. Hunt Allcott & Cass R. Sunstein, *Regulating Internalities* (Nat'l Bureau of Econ. Res. Working Paper No. 21187, Feb. 2015), https://dash.harvard.edu/handle/1/16150609 [hereinafter Sunstein & Allcott, Working Paper]; for a condensed published version, see Hunt Allcott & Cass R. Sunstein, *Regulating Internalities*, 34 J. POL'Y ANAL. & MGMT. 698 (2015). Allcott and Sunstein briefly consider some of the issues I address, including the choice between different approaches to regulation. Working Paper at 7–9. But their treatment is cursory, omits most of the analysis offered here, and as a result goes astray at one or two points. *See infra* notes 227, 271.
Another partial analysis is Jacob Goldin & Nicholas Lawson, Defaults, *Mandates, and Taxes: Policy Design with Active and Passive Decision-Makers*, 18 AM. L. & ECON. REV. 438 (2016). Goldin & Lawson show persuasively that the combination of taxes and nudges can be superior to a flat ban on harmful choices along some dimensions, *id.* at 441–42, but they presume that nudges will always be used for passive actors, taxes for active. *Id.* at 450. My central question is what extent these instruments may be preferable for either set of actors. It also is unclear whether their framework extends to other irrationality settings. Finally, they do not consider income, revenue, or distributive effects, among other considerations examined here. *See infra* Part IV.

16 Coons & Weber, *supra* note 13, at 1.

17 AUSTRALIAN GOVERNMENT DEPARTMENT OF HEALTH AND AGING, EVALUATION OF THE

tied up in litigation.[18] Are these "graphic images" the best way to regulate smoking, or would something else, like a higher tobacco tax, be the best choice? Critics' qualms do not relieve courts and other actors in these regimes from having to confront the question of regulatory design.

In addition to neglecting the practical urgency of advancing the debate, the paternalism debaters overlook a standard argument for regulation, tracing all the way back to Ronald Coase's classic essay "The Problem of Social Cost."[19] The typical critique of government intervention rests on government's supposed inability to know better than the individual what will satisfy that person's preferences.[20] For many self-harms, however, we can observe that individuals want to behave otherwise, but struggle to overcome their own worst impulses. We join Weight Watchers, enlist our employers to help us save, buy gym memberships we know it will be costly to escape. In essence, what we are seeing is two close neighbors, sharing the same piece of property, at war over how best to live.

Self-harms, that is, closely resemble Coase's framework for thinking about regulation. In a perfectly functioning market with effective property rights, Coase notes, neighbors can negotiate with each other to come to agreements on harms and benefits that cross property lines.[21] More realistically, in many cases the transaction costs of negotiating make these deals impossible or prohibitively expensive, so that government may need to step in to reproduce the bargain the parties might otherwise have struck.[22] I'll show here that this same analysis can often be readily applied to self-harms—neighbors in the same body, struggling to agree—though I also note that in some cases the existence of private markets for self-correction complicate the story.

Transaction costs alone are not a complete justification for regulation. Even if government is right that there is a problem, compliance and enforcement carry

---

EFFECTIVENESS OF THE GRAPHIC HEALTH WARNINGS ON TOBACCO PRODUCT PACKAGING 12–15, http://webarchive.nla.gov.au/gov/20140801094920/http://www.health.gov.au/internet/main/publishing.nsf/Content/phd-tobacco-eval-graphic-health-warnings-full-report; David Hammond, *Health Warning Messages on Tobacco Products: A Review*, 20 TOBACCO CONTROL 327, 327 (2011).

18 *R.J. Reynolds Tobacco Co. v. Food & Drug Admin.*, 696 F.3d 1205, 1208 (D.C. Cir. 2012). The underlying First Amendment basis for the D.C. Circuit's initial rejection of the graphic images rules was later overruled by the Court sitting en banc in a different case, *Am. Meat Inst. v. U.S. Dep't of Agric.*, 760 F.3d 18, 22–23 (D.C. Cir. 2014) (en banc), leaving the future of the rules unclear. Similar rules for smokeless tobacco are still in effect. *See Smokeless Tobacco Labeling and Warning Statement Requirements*, U.S. FOOD & DRUG. ADMIN. (Jan. 1, 2018), http://www.fda.gov/TobaccoProducts/Labeling/Labeling/ucm2023662.htm.

19 *See generally* Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960).

20 Claire Hill, *Anti-Anti-Anti-Paternalism*, 2 N.Y.U. J. L. & LIB. 444, 445–48 (2007). Most of these arguments trace back to JOHN STUART MILL, ON LIBERTY 8 (Kathy Casey ed., 2002) (1859).

21 Coase, *supra* note 19, at 5, 11.

22 *Id.* at 13–16.

costs that might make regulation wasteful overall. [23] But law and economics has already grappled with a similar set of problems in a similar context. As Coase suggests, a basic economic rationale for government regulation is the presence of externalities—harms or benefits that one of us creates for the others, in settings where the producer of the harms or benefits has limited incentives to care about the well-being of those affected by the spillovers.[24] Here, too, government faces the problem of incomplete information: how much does it cost to remedy an externality problem, and what is the value to society of the remedy?[25]

Over the past forty years or more, law and economics and related fields, such as environmental economics, have developed an elaborate set of answers to the challenges of limited information and transaction costs. Contemporary debates focus on a handful of key design questions about the structure of regulation, and to a surprising degree have reached something like consensus on many points.[26]

My focus here will therefore be on to what extent the lessons of the externality-regulation literature apply to government efforts to protect us from ourselves.[27] Following the terminology of some leading economic commentators, I will call these failures of self-regard "internalities."[28] I first show how many lessons of the externality literature help to resolve practical complaints about whether we should even be engaged in paternalistic regulation, and document how more recent scholarly innovations go even further to resolve critics' concerns. I then move on to more concrete design questions.

To preview briefly my results, I find that many settled lessons of the externality literature are likely to be different, often profoundly different, in the internality context. "Command and control" regulation or its contemporary cousin, the "nudge," could dominate corrective taxation; rewards might be better than punishments, and legal rules can be important tools of redistribution. These

---

23 Jeff Rachlinski, *The Uncertain Psychological Case for Paternalism*, 97 N'WESTERN UNIV. L. REV. 1165, 1219–25 (2003).

24 JONATHAN GRUBER, PUBLIC FINANCE AND PUBLIC POLICY 4 (3d ed. 2011).

25 *Id.* at 137–39.

26 For an overview, see Brian Galle, *Tax, Command... or Nudge? Evaluating the New Regulation*, 92 TEX. L. REV. 837, 848–53 (2014).

27 My focus on choice of instruments distinguishes this Article from the small handful of earlier efforts at analyzing paternalistic regulation through economic tools. In Eyal Zamir, *The Efficiency of Paternalism*, 84 VA. L. REV. 229 (1998), Professor Zamir offers a model for deciding when regulation of individual failings will on net increase welfare, with the main factors being a balance of individual benefit against the "frustration" and administrative costs of regulating. *Id.* at 263–65. Zamir does not attempt to distinguish between different regulatory options.

28 *E.g.*, Jonathan Gruber, *Tobacco at the Crossroads: The Past and Future of Smoking Regulation in the United States*, 15 J. ECON. PERSPECTIVES 193, 206 (2001). The term is generally attributed to Herrnstein et al., *Utility Maximization and Melioration: Internalities in Individual Choice*, 6 J. BEHAV. DECISION MAKING 149, 149 (1993).

points probably need some unpacking for those who are not already deeply immersed in externality regulation.

Let me first try to be clear at the outset what I mean by an internality. What I have in mind is an outcome that the individual, if they deliberated about their choice in a coolly reflective, objective moment, would reject.[29] I wish I had gone to the gym last week, and that I had not eaten that second slice of pecan pie, and that I had saved more for retirement.

I want to distinguish these kinds of regretted outcomes from simple ignorance. Sometimes, we go wrong because we do not have all the information to make the right choice. Often, though, it's rational for us not to gather all the data ourselves, since information acquisition is costly.[30] In these cases it is relatively straightforward that government should just provide the information, or subsidize its production by others,[31] although to be sure the design of the best information-sharing regime is not always obvious.[32] Our challenge here is different. What should we do about people who might have the necessary information available, but still act—or fail to act—in ways that are wrong for them?

As we'll see, human decision making can go wrong in a number of different ways. I'll argue that the best regulatory design for a given problem often will be different, depending on what kind of error individuals are making. Thus, one of the contributions of the paper will be to group and categorize these errors in ways that are analytically useful.

One other preliminary point to make is that my analysis is aimed at what might be called "unforced errors." Many individuals make decisions that do not maximize their own preferences because they have been tricked or misled by others, usually for profit.[33] Assuming we're confident that trickery is in fact happening, the case for regulation in those cases is little different than the case

---

29 Zamir, *supra* note 27, at 237; see Allcott & Sunstein, Working Paper, *supra* note 14, at 12 ("[P]aternalistic regulation can be limited to situations in which individuals' choices are demonstrably inconsistent . . . .").

30 John Conlisk, *Why Bounded Rationality?*, 34 J. ECON. LITERATURE 669, 671 (1996); Roy Radner, *Bounded Rationality, Indeterminacy, and the Theory of the Firm*, 106 ECON. J. 1360, 1363 (1996).

31 Alan Schwartz, *Regulating for Rationality*, 67 STAN. L. REV. 1373, 1375–76 (2015).

32 For example, recent commentators critique most efforts to cure information market failures, OMRI BEN-SHAHAR & CARL SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE 59–118 (2014), and propose extensive (if pessimistic) design solutions for some of these problems, *id.* at 121–37. I do not mean to claim that there is always a clear-cut difference between simple ignorance and more complex cognitive problems. *See* Zamir, *supra* note 27, at 254. I only intend to rule out those cases of information failure that indeed are straightforwardly rational.

33 Willis, *supra* note 14, at 1170–73.

for prohibiting robbery or fraud.[34] This is not to say that the choices regulators must make are simple, as private actors can respond to each move the regulator makes to protect consumers. Those are interesting challenges, but they have been well addressed by others.

With that definitional work out of the way for the moment, let's turn back to the externality-regulation literature. A central issue for any would-be regulator is to choose what tools or "instruments" the government will employ.[35] Should government regulate using "prices" or more traditional "command and control" regulation? If it's a price, should the price be a penalty (stick) or reward (carrot)? A carbon tax, a subsidy for going green, or a hard cap on the tons of carbon emitted? In recent work I suggest that behaviorally-informed policies with "surprising" impact, including what Sunstein & Thaler term "nudges," also can be fit into this framework.[36]

While there is some nuance in the literature's answers, as with any important and complex question, the general consensus is that sticks are the best choice.[37] The government can design a price to capture most of the features of a hard cap, and in addition sticks bring in money and elicit more information from the public.[38] Carrots can reveal private information, but also cost money, and even worse they may induce moral hazard: bad actors will commit bad deeds simply in order to be paid to stop.

I'll argue here that often none of these traditional advantages of sticks apply to internality regulation. Contrary to a celebrated result from Ronald Coase, moral hazard is a small concern for internalities, because it is difficult to credibly threaten to injure oneself for gain.[39] Prices induce rational actors to reveal their private costs of compliance with the government's preference, but of course the problem for internality sufferers is that their responsiveness does not necessarily

---

34 Steven Shavell, *Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent*, 85 COLUM. L. REV. 1232, 1239 n.28 (1985).

35 Jonathan Baert Wiener, *Global Environmental Regulation: Instrument Choice in Legal Context*, 108 YALE L.J. 677, 755–60 (1999); Brian Galle, *The Tragedy of the Carrots*, 64 STANFORD L. REV. 797, 813–40 (2012).

36 Galle, *supra* note 26, at 854–59. The main difference between that earlier paper and this is the earlier work focuses almost entirely on externalities, or in a few instances externalities mixed with internalities. The category of "surprising" regulation is broader than nudges. According to Thaler and Sunstein, a nudge is an instrument that, for a fully rational actor, would have no or minimal welfare effects. THALER & SUNSTEIN, *supra* note 12, at 12. A surprising regulation, in my systematization, is any whose effects are larger than rational-choice theory would predict, and so can include instruments whose impact is non-negligible. *See* Ian Ayres, *Regulating Opt-Out: An Economic Theory of Altering Rules*, 121 YALE L.J. 2032, 2087 (2012) (noting that "sticky defaults" may have "moderate" cost). For more discussion of the significance of the difference, see Brian Galle, *What's In a Nudge?*, 3 ADMIN. L. REV. ACCORD 1 (2017).

37 *See infra* Part I.A.

38 *Id.*

39 Coase, *supra* note 19, at 42.

reflect the real long-term costs and benefits they face. Government must use other methods, such as experiments, menus, and self-targeting regulation, to reach the right actors when it regulates internalities, and these methods work roughly the same whether the instrument is denominated in dollars or not.

Taxes and other sticks still bring in more money than other options, of course, but in some instances the revenues are not worth the extra costs they bring. Building on my earlier work in the externality context, I show that "nudges" can potentially be a better choice on net, despite the absence of revenue. On this front I depart from earlier work by economists arguing that "small" internality-correcting taxes on soda or other tempting foods can be highly efficient.[40] I argue that this claim does not always hold for taxes large enough to affect consumers' labor-supply decisions.

Overall, it will turn out that the best choice of instrument depends on what kind of mistake individuals are making. For those of us who struggle with willpower or impatience, taxes may dominate. For those of us who go wrong through failures of attention—for instance, by neglecting our retirement savings—nudges look, at least on current available evidence, to be a more promising alternative.

In sum, while the lessons of the externality literature may not apply fully in the internality context, the analytical tools of that literature remain powerful. Even at this early stage of academic study of internalities, we can form some good hypotheses about what an efficient internality-regulation regime might look like. Whether that is enough to satisfy critics of internality regulation I cannot say. But many governments are already embarked on fairly extensive internality regulation. My analysis here offers a first, tentative, glance at how those regulations should take shape in the future.

II. BACKGROUND AND PRIOR LITERATURE

This Part provides context for readers who may be unfamiliar with aspects of my argument. Part I.A. offers a general overview of economic approaches to externality regulation.[41] Part I.B. then briefly summarizes the concept of "internalities." Finally, Part I.C. offers some definitions and caveats going forward; even readers familiar with the concepts in Parts I.A. and I.B. may want to briefly visit I.C.

---

40 Ted O'Donoghue & Matthew Rabin, *Optimal Sin Taxes*, 90 J. PUB. ECON. 1825, 1827 (2006).

41 Part I.A. repeats, in essentially identical form, my earlier summary of this topic. Galle, *supra* note 26, at 843–46.

A. Regulating Externalities

Modern economic theories of government regulation begin with the premise that markets sometimes fail.[42] Externalities are a classic example.[43] An externality, simply put, is a harm ("negative externality") or benefit ("positive externality") that affects someone other than the actor making an economic decision.[44]

In general, the goal of regulation is neither to eliminate negative nor to produce boundless quantities of positive externalities, but rather to achieve what might be called the optimal level of externality.[45] Eliminating even the worst pollutants is costly. Should government bankrupt coal producers, or is there a way to balance clean air against the costs of achieving it? On the positive externality side, everyone might agree that charity is beneficial. But should government spend millions to clothe or educate one more child?

Economists typically answer these kinds of balancing questions using marginal analysis.[46] Under this approach, the policy maker asks herself, "on the margin—that is, for the very next unit of good or bad produced—what is the harm or benefit of that one unit for *everyone in society*?" We might therefore call this "marginal *social* damage," in the case of a negative externality, or "marginal social benefit" for a positive one. She then compares this harm or benefit against the marginal costs to the producer. If the producer's private marginal cost is greater than the marginal social damage, it does not pay, on net, to prevent the damage: counting the producer's losses, society would lose by forcing the producer to avoid the externality.[47]

The point at which these two quantities are equal is known as the optimal point, the point at which there are no social gains from either more or less externality correction.[48] With greater externality correction, the costs of charity

---

42 GRUBER, *supra* note 24, at 3.

43 *Id.* at 4.

44 *Id.* at 122–23.

45 *Id.* at 137–39; Gloria E. Helfand et al., *The Theory of Pollution Control*, in 1 HANDBOOK OF ENVTL. ECON. 249, 253 (Karl-Goran Maher & Jeffrey R. Vincent eds., 2003).

46 GRUBER, *supra* note 42, at 126.

47 Note, importantly, that for simplicity we are assuming here that we should count the costs and benefits for the producer and everyone else equally. That is a controversial proposition, but I'll leave it aside here for ease of exposition.

48 I'm simplifying here for the sake of exposition. A more rigorous approach to setting the optimal quantity would also account for other factors that might affect the efficiency of the regulation. For example, if the regulation imposes costs, and the expectation of those costs changes behaviors other than the production of the externality—for example, distorts consumer choices among products—the ideal regulation might balance disruption of these expectations against pollution control. Helmuth Cremer et al., *Externalities and Optimal Taxation*, 70 J. PUB. ECON. 343, 346 (1998).

or pollution reduction outweigh the benefits. With less, we have left cost-effective improvements on the table.

We could imagine a few ways of achieving production at this optimal level. If government knew the shapes of the two curves, it could calculate the optimal quantity and simply mandate that producers achieve it, with jail for those who refuse.

Another approach is to set a price for producers. In the case of pollution, government could impose a fee or tax on each unit of carbon, in an amount equal to the producer's marginal cost at the optimum. Call this price "tau". For producers whose costs of eliminating the next unit of carbon are below tau, they will eliminate it, saving themselves tau minus their cost. For producers whose costs are above tau, they will simply emit the carbon and pay the tax. Thus, just as with the mandate, rational producers should produce exactly the optimal amount of carbon. Or, similarly, government could pay producers to eliminate carbon or produce charity. Once more, if the government offers a price tau, only producers who can fill a shelter bed for less than tau will take the offer. Economists often call the first of these approaches "quantity regulation,"[49] and the second two "price instruments."[50]

Most commentators strongly favor price instruments over quantity regulation, except in settings where special administrative considerations make prices impractical.[51] As Kaplow & Shavell show, prices can be used to duplicate most of the features of mandates.[52] Prices provide vital information to the government that regulation supposedly does not.[53] Further, prices are said to provide for revenues that the government can use for other projects.[54]

---

49 GRUBER, *supra* note 24, at 137.

50 THOMAS STERNER, POLICY INSTRUMENTS FOR ENVIRONMENTAL AND NATURAL RESOURCE MANAGEMENT 214–15 (2003).

51 GRUBER, *supra* note 24, at 140; Don Fullerton et al., *Environmental Taxes*, *in* DIMENSIONS OF TAX DESIGN: THE MIRRLEES REVIEW 231 (James Mirrlees ed. 2011); Maureen L. Cropper & Wallace E. Oates, *Environmental Economics: A Survey*, 30 J. ECON. LIT. 675, 686 (1992); Cameron Hepburn, *Regulation by Prices, Quantities, or Both: A Review of Instrument Choice*, 22 OXFORD REV. ECON. POL'Y 226, 228–29 (2006). As an example of a "special consideration," price instruments may be riskier than quantity regulation when the marginal social damage curve is steep but its exact shape is uncertain, GRUBER, *supra* note 42, at 143–46, and the policy maker cannot sharply vary the tax rate to account for this risk.

52 Louis Kaplow & Steven Shavell, *On the Superiority of Corrective Taxes to Quantity Regulation*, 4 AM. L. & ECON. REV. 1, 7–10 (2002).

53 *Id.* at 4.

54 *E.g.*, Helfand et al., *supra* note 45, at 287; Ian Parry et al., *When Can Carbon Abatement Policies Increase Welfare? The Fundamental Role of Distorted Factor Markets*, 37 J. ENVTL. ECON. & MGMT. 52, 52 (1999).

B. Internalities

Harms done to others are a classic economic rationale for government regulation, but what about harms done to self? Most readers likely know that a large body of literature now suggests that individuals make decisions—or fail to make them—in ways that in the long run likely do not maximize their own subjective well-being. [55] Some commenters, seizing on the externality analogy, have dubbed these kinds of mistakes "internalities": costs that the deciding self inflicts on its temporal successors.[56]

Because a good deal of my later discussion will turn on the details of how humans go wrong, it's worth highlighting some aspects of the empirical literature here. One key finding is that we are overwhelmingly creatures of the present, and only through exercises of our limited pool of willpower can we force ourselves to take sufficient account of the future.[57] Relatedly, we tend to focus our attention on facts that are readily available to us or on items in plain sight, reacting automatically and emotionally to those immediate stimuli.[58] The Nobelist Daniel Kahneman calls these two modes of reasoning, the unconscious and the deliberative, "system one" and "system two," respectively. Only with some effort do we turn our attention to the distant and the hidden, and engage our system two reasoning powers to reach better decisions.[59] We "anchor" on information we have already received, and interpret new data selectively to fit with what we already know or want to be true.[60] In all of these areas evidence suggests that individuals vary considerably in their susceptibility to the behavior.[61]

The consequences of these human tendencies can be seen all around us. Few human institutions, from families up through the federal government, make adequate plans for their financial future.[62] We procrastinate or give in to

---

55 For reviews, see B. Douglas Bernheim & Antonio Rangel, *Behavioral Public Economics: Welfare and Policy Analysis with Nonstandard Decision Makers*, *in* BEHAVIORAL ECONOMICS AND ITS APPLICATIONS 7, 10–65 (Peter Diamond & Hannu Vartiainen eds., 2008); Stefano DellaVigna, *Psychology and Economics: Evidence from the Field*, 47 J. ECON. LITERATURE 315 (2009).

56 *See supra* note 28.

57 *See generally*, Lee Anne Fennell, *Willpower Taxes*, 99 GEO. L.J. 1371, 1375–94 (2011); Shane Frederick et al., *Time Discounting and Time Preference: A Critical Review*, *in* ADVANCES IN BEHAVIORAL ECONOMICS 162, 166–79 (Colin F. Camerer et al. eds., 2007) (providing an overview of the literature).

58 Daniel Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics*, 93 AM. ECON. REV. 1449, 1451–57 (2003).

59 *Id.* at 1467–69.

60 JONATHAN BARON, THINKING AND DECIDING 203–24, 263–70 (4th ed. 2008)

61 Ted O'Donoghue & Matthew Rabin, *Self-Awareness and Self-Control*, *in* TIME AND DECISION: ECONOMIC AND PSYCHOLOGICAL PERSPECTIVES ON INTERTEMPORAL CHOICE 217, 219–20 (George Loewenstein et al. eds., 2003).

62 Shlomo Benartzi & Richard Thaler, *Heuristics and Biases in Retirement Savings Behavior*, 21 J. ECON. PERSP. 81, 82–84 (2007).

temptation, then build costly structures to overcome our tendencies, and then incur even more costs to unwind them.[63] People smoke too much, do not exercise enough, eat to excess. Many of us, even trained experts, make decisions based on only a fraction of the information available to us, choosing poor investments and neglecting "hidden" costs that in actuality are easily calculable.[64]

Importantly for my later analysis, evidence so far suggests that some of us are more self-aware of these failings than others.[65]

C. Philosophical Foundations & Other Clarifications

Although there now is extensive evidence that individuals make decisions that do not satisfy their own long-run preferences, there has been little scholarly analysis of how best to remedy that problem. Debate instead is stuck at a more fundamental question: should government be in the business of correcting internalities at all? Critics assert that government intervention is unwarranted "paternalism."[66] Behind the paternalism label are two deeper critiques: that humans should have the autonomy to make their own mistakes, and that governments lack the capacity to regulate in ways that will lead to better outcomes.[67]

The autonomy argument poses difficult philosophical problems that law & economics lacks the tools to resolve. For me, law that helps individuals achieve their true goals furthers autonomy, rather than undermining it. That, after all, is the structure of constitutions: they protect bodies politic from momentary whims and passions, and preserve the capacity for long-run self-determination.[68] But I recognize that some readers will have philosophical commitments that make it hard for them to accept this claim.

---

63 Frederick et al., *supra* note 57, at 172–79.

64 John R. Graham & Campbell R. Harvey, *The Theory and Practice of Corporate Finance: Evidence from the Field*, 60 J. FIN. ECON. 187, 188–243 (2001); Aradhna Krishna et al., *A Meta-Analysis of the Impact of Price Presentation on Perceived Savings*, 78 J. RETAILING 101, 101–18 (2002).

65 *E.g.*, Michael S. Barr & Jane K. Dokko, *Third-Party Tax Administration: The Case of Low- and Moderate-Income Households*, 5 J. EMPIRICAL L. STUDIES 963 (2008); Ryan Bubb & Alex Kaufman, *Consumer Biases and Mutual Ownership*, 105 J. PUB. ECON. 39, 53 (2013).

66 Richard A. Epstein, Exchange, *The Neoclassical Economics of Consumer Contracts*, 92 MINN. L. REV. 803, 806–07 (2008); *see generally* Heidi M. Hurd, *Fudging Nudging: Why Libertarian Paternalism is the Contradiction It Claims It's Not*, 14 GEORGETOWN J.L. & PUB. POL'Y 703 (2016), manuscript at 8.

67 Epstein, *supra* note 66, at 806–07; Joshua D. Wright & Douglas Ginsburg, *Behavioral Law & Economics: Its Origins, Fatal Flaws, and Implications for Liberty*, 106 NORTHWESTERN L. REV. 1033, 1065–74 (2012); Coons & Weber, *supra* note 16, at 7–9.

68 Frederick Schauer, *Judicial Supremacy and the Modest Constitution*, 92 CAL. L. REV. 1045, 1054–55 (2004); *see* NORMAN DANIELS, JUST HEALTH CARE 159 (1985) (offering this rationale as a justification for so-called paternalistic regulation, despite alleged autonomy concerns).

Economics has more to offer in the debate over whether government has the capacity to regulate internalities. I will try to resolve these questions in Part III. The answers, we'll see, often depend on just how individuals are failing themselves. Therefore, we will first need a brief taxonomy of internalities; Part II takes up that task.

Before that, it is worth offering a few clarifications to the scope of my task. One is that, following the prevailing law & economics literature approach to cost-benefit analysis, my approach is essentially welfarist.[69] I am interested in which set of rules maximizes total social welfare, assuming diminishing marginal utility and some degree of popular preferences for distributive fairness.[70] Of course, there could be alternative consequentialist approaches to these same questions, such as the suggestion by Sen and Nussbaum that we maximize basic human capabilities,[71] or perhaps a Rawlsian-inspired approach that would maximize health over other outcomes.[72] We also could consider deontological approaches, such as one that prioritized autonomy or dignity.[73] I do not mean to suggest these approaches are invalid, but they are not in common use in the externality literature. My goal here is to focus first on translating what we already know about externalities to internalities.

Second, it might be objected that some or even most internality problems actually present a mix of internalities and externalities.[74] What, then, does an internality analysis add? My answer is that we can think of the internality as offering a reason for more extensive regulation. We just saw that the optimal level of regulation depends on the marginal social damage of a product.[75] We should include both externalities and internalities in calculating the marginal social damage.[76] This can make a dramatic difference in the regulator's choices.

---

69 Zamir, *supra* note 27, at 233–35.

70 Part III.A. discusses the challenge of measuring welfare when individuals' observable choices do not necessarily reflect their long-term preferences.

71 MARTHA C. NUSSBAUM, FRONTIERS OF JUSTICE: DISABILITY, NATIONALITY, SPECIES MEMBERSHIP 69–81 (2006); AMARTYA SEN, INEQUALITY REEXAMINED 39–55 (1992).

72 DANIELS, *supra* note 68, at 42–47; Lawrence O. Gostin, *Securing Health or Just Health Care? The Effect of the Health Care System on the Health of America*, 39 ST. LOUIS U. L.J. 7, 13 (1994).

73 Wright & Ginsburg, *supra* note 67, at 1068–75; Hurd, *supra* note 66, at 2.

74 *See* Wendy Mariner, *Paternalism, Public Health, and Behavioral Economics: A Problematic Combination*, 46 CONN. L. REV. 1817, 1833 (2014); *cf.* Katherine Pratt, *A Constructive Critique of Public Health Arguments for Antiobesity Soda Taxes and Food Taxes*, 87 TULANE L. REV. 73, 77–103 (2012) (examining both externality and internality rationales for obesity regulation).

75 *See supra* note 46–36.

76 Zamir, *supra* note 27, at 278. To be precise, the government should include the gap between the cost of the amount of internality the individual will consume on their own, and the cost of the optimal level of consumption, in the social marginal cost. In other words, if some consumers can avoid the internality on their own, but not completely avoid it, the price can be lower.

For instance, Gruber and Koszegi estimate that the back of the envelope externality cost of a pack of cigarettes is less than $1, while the internality cost is more like $30.[77] If government currently imposes only a small per-pack cost on smoking, its decisions about whether and how to correct internalities may determine if it will now impose a massive tax hike.

Finally, I emphasize that my analysis is aimed at genuine failures of decision, not simply ignorance of the best choice. Because information often has many of the features of a public good—that is, my investment in acquiring information produces positive externalities for others I usually cannot charge them for—the economic case for government support of information creation is straightforward.[78] Likewise, when individuals can rely on informed others to act for them, we have "rational ignorance," supplying at least a basic argument in favor of regulation.[79] Having said that, human limits in absorbing and processing information can form an obstacle to good decisions even when the underlying information is freely available and individuals have incentives to employ it. [80] These kinds of failures are within the scope of my analysis here.

III. MAPPING INTERNALITIES

To simplify exposition, we can think of individuals' attitudes towards internalities as the product of a two-by-two grid. To distinguish internality regulation from the (in my view) easier case of government information production, let us assume that full information about the costs and benefits of a given choice are freely available to all decision makers.

Let the first dimension of the grid, *C*, represent the marginal compliance cost perceived by the individual at the time of compliance. I will assume for present purposes that this subjective cost of compliance should properly be included in the social welfare function. Thus *C* is the equivalent, in the internality context, of

---

Admittedly, this approach does lead to potential difficulties if a given activity produces *positive* externalities and negative internalities. I reserve that case for future development.

77 Jonathan Gruber & Botond Koszegi, *Is Addiction "Rational"? Theory and Evidence*, Q.J. ECON. 1261, 1291 (2001).

78 Janusz Ordover & William Baumol, *Antitrust Policy and High-Technology Industries*, 4 OXFORD REV. ECON. POL'Y 13, 14 (1988). On the general theory of public goods, see RICHARD CORNES & TODD SANDLER, THE THEORY OF EXTERNALITIES, PUBLIC GOODS, AND CLUB GOODS 8 (2d ed. 1996).

79 Howard Beales et al., *The Efficient Regulation of Consumer Information*, 24 J.L. & ECON. 491, 495–501 (1981); *see* Frank H. Easterbrook & Daniel R. Fischel, *Mandatory Disclosure and the Protection of Investors*, 70 VA. L. REV. 669, 681–85 (1984) (explaining this point, but also arguing that private market can substitute for government encouragement in some cases).

80 *See, e.g.*, Julie S. Downs et al., *Strategies for Promoting Healthier Food Choices*, 99 AM. ECON. REV. (Papers & Proceedings) 159, 160—62 (2009) (summarizing two studies in which disclosing health information had no or perverse effects).

the private marginal cost faced by an externality producer for each unit of externality.

The second dimension of the grid can be represented by *B*, or the subjective benefit the individual perceives at the time she must make the decision to comply or not. Since compliance or not determines whether the benefit occurs, *B* is also the value the individual assigns to a given government-preferred outcome at the time she makes the decision that contributes to that outcome. Also for simplicity, for now we will assume that the government's preference in fact would improve the individual's welfare relative to non-compliance. Under that assumption, *B* measures an individual's ability to perceive that her current decisions may not maximize her overall well-being. For example, she may recognize that she has a propensity to undervalue the future.

Figure One depicts the resulting possibilities, along with some illustrative examples.

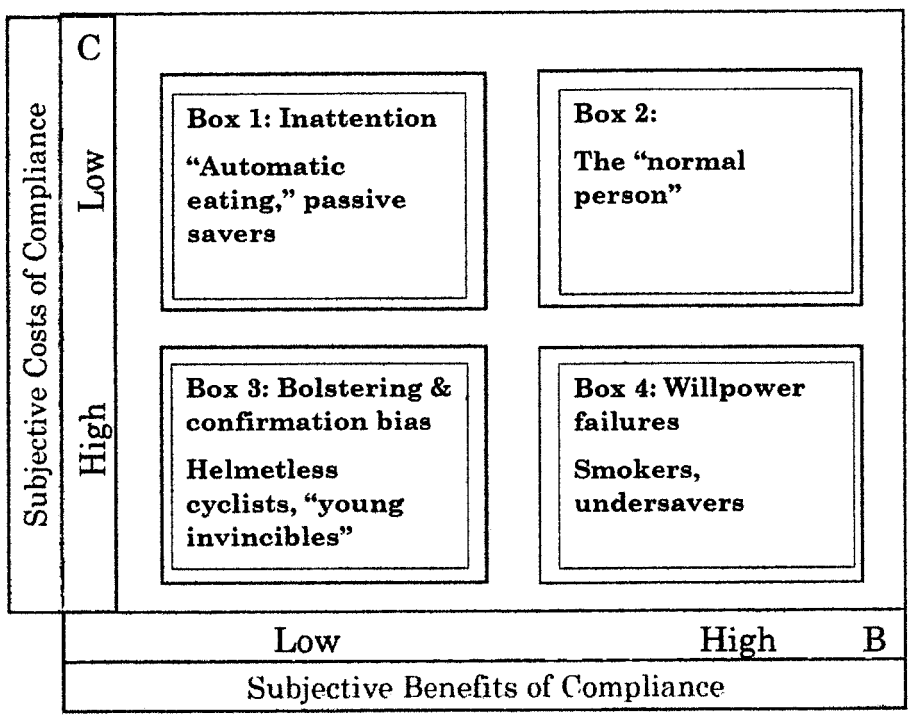

Figure 1. Cognitive Failures in Two Dimensions

Box 2, the upper right square of the grid, is the baseline case, the "normal" person.[81] She perceives the costs of complying with the government's policy as low and the benefits as high. In all likelihood, she needs no additional incentive to follow the government's course.

Individuals in the upper left box, Box 1, do not view the government's suggestion as burdensome, but also do not recognize that its goals are worthwhile. This box might capture the empirically well-documented phenomena of "inattentive" actors and "reference dependence," or the importance of framing and presentation on how we decide.[82] For example, Wansink and others find that portion size strongly influences many people's consumption of food and beverages; we eat what is in front of us, without really paying attention to how much we're eating.[83] Chetty et al. report evidence that 85% of Danish working households were unresponsive to tax incentives for savings, and also did not respond to changes in the default amount of savings the government chose for them.[84] In Kahneman's terminology, these are "decisions" that are made using system one alone.[85]

Box 4, the lower-right box, may capture willpower failures, another set of well-known behaviors.[86] Individuals know what is good for them, but in the moment they must make their decision, they find the bad choice too difficult to resist.[87] We should expect that high-C , high-B individuals will seek out "commitment devices," or tools to help them obtain the beneficial outcome.[88] Whether government intervention is justified to assist these households may depend on the extent to which commitment devices are unobtainable, create costly side-effects, or have unwanted distributive impacts.[89]

---

81 *But see* Arcade Fire, Normal Person, on *Reflektor* (Arista 2013) ("I've never really ever met a normal person.").

82 Kahneman, *supra* note 58, at 1451–57. For more in-depth reviews, see B. Douglas Bernheim & Antonio Rangel, *Behavioral Public Economics: Welfare and Policy Analysis with Nonstandard Decision Makers*, *in* BEHAVIORAL ECONOMICS AND ITS APPLICATIONS 7, 10–5 (Peter Diamond & Hannu Vartiainen eds., 2008); Stefano DellaVigna, *Psychology and Economics: Evidence from the Field*, 47 J. ECON. LITERATURE 315, 324–36, 347–56 (2009).

83 BRIAN WANSINK, MINDLESS EATING 17–19, 47–52 (2006); Pierre Chandon, *How Package Design and Packaged-Based Marketing Claims Lead to Overating*, 35 APPLIED ECON. PERSPECTIVES & POL'Y 7, 13–18 (2013).

84 Raj Chetty et al., *Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark*, 129 Q.J. ECON. 1141 (2014).

85 Kahneman, *supra* note 58, at 1451–57.

86 For surveys, see Lee Anne Fennell, *Willpower Taxes*, 99 GEO. L.J. 1371,1375–94, and Shane Frederick et al., *Time Discounting and Time Preference: A Critical Review*, *in* ADVANCES IN BEHAVIORAL ECONOMICS 162, 166–79 (Colin F. Camerer et al. eds., 2004).

87 *Id.*

88 Gruber & Koszegi, *supra* note 77, at 1278; Ted O'Donoghue & Matthew Rabin, *Doing It Now or Later*, 89 AM. ECON. REV. 103, 105 (1999).

89 Brian Galle & Manuel Utset, *Is Cap-and-Trade Fair to the Poor? Shortsighted Households*

Individuals who suffer from various forms of "bounded rationality" may fall somewhere near the southeast corner of Box 1 and northwest corner of Box 4.[90] Oftentimes we face problems we lack the cognitive capacity to absorb.[91] To economize on time and brainpower, we may take mental shortcuts that lead us to imperfect answers.[92] Accepting the government's choice might save us from having to make our own decision, but if so we do not have a good way to know if it's the best choice for us. Verifying that the government's suggestion is a good one would be somewhat costly, but also establish it as somewhat valuable. Also, in some cases it appears that the use of shortcuts is itself motivated by procrastination, making these cases a true hybrid.[93]

The last, lower-left, box is perhaps the most puzzling. Despite the presence of (by hypothesis) full information, individuals here largely ignore the benefits of the government's choice. Remember that we are assuming for now that the government's choice is correct; it is not simply that individuals in this corner know their own tastes better. Nonetheless, since they perceive the cost of changing their own decision as high, they resist government proposals. This might describe the so-called "naïve hyperbolic discounters," those who are impatient but fail to recognize their own impatience.[94]

This set of behaviors could also reflect what is sometimes called "bolstering," or "cultural cognition."[95] That is, we tend to selectively filter our understanding of the world in ways that reinforce our preferred outcome.[96] Since the present self wants to feel the wind in its hair, it screens out or rejects as "biased" evidence that helmetless motorcyclists die in remarkable numbers.[97] Often the illusion of personal control is an important tool in the process of self-deception, allowing the actor to distinguish her own case from the statistical

---

and the Timing of Consumption Taxes, 79 GEO. WASH. UNIV. L. REV. 33, 76–80 (2010).

90 Herbert Simon, *Invariants of Human Behavior*, 41 ANN. REV. PSYCHOL. 1, 6 (1990).

91 Conlisk, *supra* note 30, at 672. Bubb & Pildes, *supra* note 14, at 1613–14, summarize the evidence in the context of retirement savings.

92 RICHARD THALER, QUASI-RATIONAL ECONOMICS 3–5 (1994); Kahneman, *supra* note 58, at 1458–59.

93 Ted O'Donoghue & Matthew Rabin, *Procrastination in Preparing for Retirement, in* BEHAVIORAL DIMENSIONS OF RETIREMENT ECONOMICS 125, 125–26 (Henry J. Aaron ed., 1999); Kahneman, *supra* note 58, at 1468.

94 *See generally* Ted O'Donoghue & Matthew Rabin, *Choice and Procrastination*, 116 Q.J. ECON. 121 (2001) (developing a model of partially naïve households)

95 Dan M. Kahan, Foreword, *Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law*, 125 HARV. L. REV. 1, 19–30 (2011); Philip E. Tetlock et al., *Social and Cognitive Strategies for Coping with Accountability: Conformity, Complexity, and Bolstering*, 57 J. PERSONALITY & SOC. PSYCH. 632, 632 (1989).

96 BARON, *supra* note 60, at 208–11; George A. Akerlof & William T. Dickens, *The Economic Consequences of Cognitive Dissonance*, 72 AM. ECON. REV. 307, 307-09 (1982).

97 Neil D. Weinstein & William M. Klein, *Resistance of Personal Risk Perceptions to Debiasing Interventions*, 14 HEALTH PSYCHOL. 132, 139 (1995).

evidence: *other* people might die of lung cancer, but *I* can quit whenever I want.[98] A number of commentators point to these kinds of processes as also explaining why many households do not adequately insure against flood and other disasters.[99]

Again, the taxonomy is meant to be simplifying, and so it likely misses some relevant nuance. The boxes are not meant to represent points, but continua. Nonetheless, it is a starting place, and I will now argue that it at least in part reflects important differences between different potential sets of internalities.

IV. SHOULD WE REGULATE INTERNALITIES?

Before we proceed to design issues, I expect that many readers likely want some additional convincing that government should regulate internalities at all. Again, basic philosophical objections are beyond my scope here. The core of many putatively philosophical claims, however, are actually quite practical. For example, many self-identified libertarian thinkers object to paternalistic regulation because, they claim, the government cannot know individual preferences accurately enough to regulate effectively.[100] Thus in Part III.A. I will address the core of this informational argument against regulating internalities. A second common libertarian objection to any regulation is that the possibility of private ordering solutions makes government action unnecessary. Part III.B. therefore considers whether government regulation of internalities would interfere with an efficient private market in internality correction.

A. The Information Problem

We saw in Part I.A. that regulators usually need two sets of information in order to optimally correct market failures.[101] First, they must know the marginal harm inflicted by each additional unit of the regulated good. For externality goods, this is the damage done to others, while for internalities it is the damage to the actor herself. At the same time, government needs to know how socially costly it will be to correct the problem: what is the actor's cost of keeping that next ton of carbon out of the air, that Twinkie out of our mouths?

Early criticisms of "paternalistic" regulation claimed that identifying the "harm" of an internality was inconsistent with basic economic methods, [102] but

98 Suzanne C. Thompson et al., *Illusions of Control, Underestimations, and Accuracy: A Control Heuristic Explanation*, 123 PSYCH. BULLETIN 143, 144–61 (1998).

99 Howard Kunreuther & Mark Pauly, *Rules Rather Than Discretion: Lessons from Hurricane Katrina*, 33 J . RISK & UNCERTAINTY 101, 105-06 (2006).

100 *See* sources cited *supra* note 67.

101 *See supra* notes 29–45.

102 Zamir, *supra* note 27, at 237–38 (attributing this view to John Stuart Mill).

the literature has largely rejected that argument. That is, we usually infer marginal benefit from revealed preferences: the consumer faces a price, and if she is willing to pay that price, we conclude that her subjective welfare is greater than the price she paid.[103]

How, then, are we to second-guess consumers' choices? The answer is time.[104] By observing individual behavior over time, we can see whether people regret some of their own decisions, or take steps (commitments) to prevent themselves from making bad choices.[105] In this way, we can still rely on revealed preferences.[106] The concept of an "internality" does not necessarily privilege long-run over short-run preferences; we can treat them equally by simply adding them up, in effect balancing the revealed value of long-term preferences, such as regret and commitment, against the revealed value of momentary, System 1 preferences.[107] But since long-run preferences last much longer, they often will greatly outweigh those that last only fleetingly.[108]

We can, in other words, think of internalities as simply conflicts between internally-conflicting sets of impulses and preferences. Ronald Coase' transaction-cost framework then readily justifies regulation for many kinds of internality. Coase argued that government could resolve bargaining problems, such as hold-ups and collective action failures, that would otherwise prevent private-market solutions for externalities.[109] Similarly, we might say that internal-

---

103 Paul Burrows, *Rationality and the Instrumentalist Case for Free Choice*, 15 INT'L REV. L & ECON. 489, 491 (1995); Daniel M. Haybron & Anna Alexandrova, *Paternalism in Economics*, *in* Coons & Weber, *supra* note 13, at 157, 159–60.

104 ELSTER, *supra* note 10, at 5–7, 22; O'Donoghue & Rabin, *supra* note 40, at 1829 n.12.

105 *Id.* ("[F]or any tax policy that takes effect in the future . . . the agent agrees that [the long-run perspective] is the appropriate welfare function.").

106 *Id.*; Sunstein, *supra* note 13, at 1875–76; Weiss, *supra* note 3, at 1305–06; Zamir, *supra* note 27, at 247, 253. Although this point has been well-established for a decade, critics of internality correction continue to assert that any internality regulation unfairly privileges long- over short-term preferences. *E.g.*, Riccardo Rebonato, *A Critical Assessment of Libertarian Paternalism*, 37 J. CONSUMER POL'Y 357, 370 (2014).

107 Zamir, *supra* note 27, at 246–47.

108 Markus Haavio & Kaisa Kotakorpi, *The Political Economy of Sin Taxes*, 55 EUR. ECON. REV. 575, 578 (2011); B. Douglas Bernheim & Antonio Rangel, *Toward Choice-Theoretic Foundations for Behavioral Welfare Economics*, 97 AM. ECON. REV. 464, 467 (2007). Of course, it remains possible that some impulses might be so intense and so frequent that on net the individual would be better off if those impulses were not restrained. Gary S. Becker & Kevin M. Murphy, *A Theory of Rational Addiction*, 96 J. POL. ECON. 675 (1988). The revealed preference argument does not relieve policy-makers of the burden of attempting to measure which outcome increases overall welfare. Complicating matters, some short-term preferences might reduce lifespan. If utility during the lost period is not accounted for in some way, this might tend to mechanically favor the life-reducing impulsive conduct. *Cf.* I. Glenn Cohen, *Regulating Reproduction: The Problem with Best Interests*, 96 MINN. L. REV. 424, 437–45 (2011) (discussing difficulties of welfare assessments for rules that will affect identity of individuals alive in the future).

109 Coase, *supra* note 19, at 13–16.

bargaining breakdowns could justify regulation of individuals in Box 1 or Box 3. Inattentive individuals do not notice that they are creating conflicts with their long-term preferences, making bargaining difficult. Individuals who "bolster" using motivated logic refuse to recognize, in the moment, the claims of their long-term preferences, much as the problem of double monopoly creates bargaining impasses in the externality context.[110] Willpower deficiencies present a more nuanced case, since there exist private markets that appear to allow for intra-personal bargaining. I return to this question in Part III.B.

Recent critics, such as Yale's Alan Schwartz, raise the more subtle problem that it can be challenging to identify which choices are bad decisions, failures of the internal bargaining process.[111] Suppose that the same action may be taken both by biased actors and the unbiased. Think of an auto insurance policy with a high deductible.[112] That policy is a good choice for safe drivers, who do not need expensive coverage. It's a bad choice for bad but overconfident drivers: those who believe wrongly that they do not need a lot of coverage. Schwartz's claim is that, since we can usually only observe the consumer's choice to buy a particular policy, we cannot know whether that choice produces internalities (the overconfident driver) or not.[113]

i. Solutions from the Externality Literature

The information problem is not unique to internality regulation. As we have seen, optimal regulation of an externality demands information about both the marginal damage or benefit produced by the good as well as the cost curves of private actors involved in producing it. Many key choices in modern theories of regulation turn on how best to reveal these necessary data.[114]

---

110 *See* Herbert Hovenkamp, *Rationality in Law & Economics*, 60 GEO. WASH. UNIV. L. REV. 293, 306–10 (1992) (describing problems and informational demands of bilateral monopoly).

111 PETER CSERNE, FREEDOM OF CONTRACT AND PATERNALISM 55 (2013); Hill, *supra* note 20, at 445–48; Schwartz, *supra* note 31, at 1377–78.

112 I owe this example to Goldin & Lawson, *supra* note 14, at 441.

113 *Id.* A related claim sometimes raised in the anti-paternalism literature is that government regulators, too, can suffer from cognitive biases, or will be subject to market capture in ways that reduce their ability to process information. CSERNE, *supra* note 111, at 52, 54; Wright & Ginsburg, *supra* note 67, at 1063–64. These same difficulties arise in the standard externality setting, Zamir, *supra* note 27, at 280–81, and have standard solutions. For example, notice and comment rulemaking in its modern form helps to expose, and incentivizes agencies to limit, these kinds of failures. Mark Seidenfeld, *Cognitive Loafing, Social Conformity, and Judicial Review of Agency Rulemaking*, 87 CORNELL L. REV. 486, 508–46 (2002). For a more comprehensive empirical response to the bias and capture arguments, see Jeremy Blumenthal, *Expert Paternalism*, 64 FLA. L. REV. 721, 730–50 (2012).

114 Helfand et al., *supra* note 45, at 251, 287; Steven Shavell, *Corrective Taxation Versus Liability as a Solution to the Problem of Harmful Externalities*, 54 J.L. & ECON. S249, S258–59 (2011).

To be sure, not all the tools of externality regulation can be translated seamlessly to internalities. Price instruments serve as a favorite tool for revealing preferences and private costs,[115] but it is uncertain how well they operate for irrational actors. If government taxed high-deductible plans, that would screen out marginal over-confident drivers, but those that were especially over-confident might still make the same choice. That is, the change in price does not necessarily tell us whether the choice the consumer reveals is their "true" preference or a mistake.[116]

Time, another standard component of the externality tool kit, will often still perform well, however. Schwartz's credit contract examples are instances of so-called "ex ante" regulation—an effort to set policy and enforce it before consumers make their choice.[117] Many commentators argue that, when information is scarce, the better option is an "ex post" regulation.[118] That is, we wait until after the choice is made, measure the resulting harm or benefit, and apply a corresponding price change. Torts are the classic example.[119] We cannot easily predict which soda drinkers will be most prone to diabetes, and ban or limit their soda consumption.[120] But we can simply observe which consumers develop diabetes later, and allow them to sue to collect compensatory damages.[121] A rational, forward-looking beverage producer will anticipate the possibility of tort liability and act accordingly—although, as I have argued, limited liability and other factors may sometimes interfere with that process.[122]

We could take an analogous approach to the high-deductible example. For example, we might restrict high-deductible plans for individuals who have already demonstrated a history of risky driving, or even behaviors we know to be highly correlated with risky driving, such as regular alcohol consumption. More generally, we can observe what portion of the population that selects high deductible plans turns out to be relatively high risk. If the share of those who demonstrably made bad choices is high enough, it may still be welfare-improving on net to limit those plans, even if the plan would have improved well-being for

115 Kaplow & Shavell, *supra* note 52, at 4; Fullerton et al., *supra* note 51, at 430.

116 *See* Galle, *supra* note 26, at 863–64 (discussing cognitive problems with drawing inferences from price instruments).

117 Donald Wittman, *Prior Regulation vs. Post Liability: The Choice Between Input and Output Monitoring*, 6 J. LEGAL STUD. 93, 93 (1977)

118 RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 490–91 (8th ed. 2011); Jon D. Hanson & Kyle D. Logue, *The Cost of Cigarettes: The Economic Case for Ex Post Incentive-Based Regulation*, 107 YALE L.J. 1163, 1278 (1998).

119 *Id.*

120 Victor Fleischer, *Curb Your Enthusiasm for Pigouvian Taxes*, 68 VAND. L. REV. 1673, 1704–05 (2015).

121 *Cf. id.* at 1705 (suggesting that cure for informational problem is to impose costs on obesity, not its predictors).

122 Brian Galle, *In Praise of Ex Ante Regulation*, 68 VAND. L. REV. 1715, 1734–48 (2015).

some drivers.[123]

Admittedly, the expedient of switching from *ex ante* to *ex post* regulation may not work for all kinds of internalities. Many forms of *ex post* regulation, such as the tort suit, require the internality sufferer to recognize, at some point, that they have incurred harms.[124] This is plausible for Box 1 and Box 4 consumers—those who were not attentive enough to notice their errors at the time, or lacked the willpower to avoid them. These individuals are likely to experience regret later. In contrast, Box 3 consumers may deny that their choices were wrong, and indeed may even harden their viewpoint further in order to avoid the cognitive dissonance that would come with acknowledging contrary evidence.[125]

In any event, externality theory also shows us that the informational demands of regulation for all kinds of internality sufferers can be much lower than Schwartz and other critics seem to assume. As William Baumol famously argued, sometimes all that we need to know is where we are relative to the optimal level of regulation.[126] Whatever might be the optimal number of tons of greenhouse gas a coal factory can emit, we know that the current level authorized in the U.S. is too high.[127] We may not be able to identify the level at which there would no longer be additional social returns from further reduction. But we know we are far from that point and can safely lower emissions somewhat from the current, unregulated, level.[128]

Similarly, there are many internalities where we have at least this level of certainty. Again, economists estimate the internality cost of a pack of cigarettes at more than $30, for example.[129] Typically, the methodology is to measure the

---

123 *See* Peter Diamond, *Consumption Externalities and Imperfect Corrective Pricing*, 4 BELL J. ECON. & MGMT. SCI. 526, 528–30 (1973) (deriving optimal pigouvian tax when externalities vary by consumer); *see* Zamir, *supra* note 27, at 266–67 (applying this principle to internalities).

124 *See* William L.F. Felstiner et al., *The Emergence and Transformation of Disputes: Naming, Blaming, and Claiming*, 15 L. & SOC'Y REV. 631, 632–36 (1980–81).

125 Eva Jones et al., *Confirmation Bias in Sequential Information Search: An Expansion of Dissonance Theoretical Research on Selective Exposure to Information*, 80 J. PERSONALITY & SOC. PSYCH. 557, 557 (2001). Dan M. Kahan, *The Cognitively Illiberal State*, 60 STAN. L. REV. 115, 145–50 (2007), draws on primary psychological literature to argue that self-defensive bolstering can be minimized if new information is introduced within a frame that allows the listener to identify some elements that affirm her existing worldview.

126 William J. Baumol, *On Taxation and the Control of Externalities*, 62 AM. ECON. REV. 307, 307–08 (1972).

127 *See* James R. Hines, Jr., *Taxing Consumption and Other Sins*, 21 J. ECON. PERSPECTIVES 49, 53, 64 (2007).

128 *Id.* at 57; *see also* Louis Kaplow, *Optimal Control of Externalities in the Presence of Income Taxation*, 53 INT'L ECON. REV. 487, 488 (2012) (arguing that this proposition is always true "if a distribution-neutral income tax adjustment is employed" together with the externality correction).

129 Gruber & Koszegi, *supra* note 77, at 1291.

unbiased value of "good" outcomes in the general (and presumably unbiased) population, and assume that biased actors share that value. Thus Gruber and Koszegi, for instance, calculate the cost of smoking by looking at medical costs and the average person's value of additional years of life.[130] This assumes, of course, that smokers place equal value on long life.[131] Maybe that is a plausible assumption, but maybe it is not. Still, even if smokers placed only half the value on life as others, Gruber's numbers at least tell us that current cigarette taxes are far too low.[132]

In my work on the choice between *ex ante* and *ex post* regulation, I also show that we can reduce the amount of information government needs in order to regulate, even *ex ante*, by using multiple prices or policies.[133] The social cost of a mistaken policy grows exponentially with the size of the mistake.[134] Through some simple mathematical simulations, I show that sorting actors into high, medium, and low risk categories can be just about as good as having perfect information about them.[135] Assuming government assigns a regulated party to the right category, the size of the error it's making—the distance, say, between the optimal price for that party and the actual price imposed—is smaller than if there were only one category, and the social cost accordingly declines exponentially.[136] We do not have to get policy choices exactly right in order for them to be good policies.

This same analysis also undermines Schwartz's suggestion that law should default to an assumption that actors are rational.[137] In effect, Schwartz is proposing that we set the pigouvian price on internalities to zero, unless we have compelling evidence otherwise. But that is a disastrous policy, because it greatly increases the average expected distance between the government's price and the optimal. Say that there is a 50% chance that the optimal price is $100. This is no different, statistically, than saying that half the population has an optimal price of $0 and half $100. In that case, optimal price should be set at $50.[138] Even better,

---

130 *Id.* at 1290–91.

131 *Cf. Thomas Kneisner et al., Policy Relevant Heterogeneity in the Value of Statistical Life: New Evidence from Panel Data Quantile Regressions*, 40 J. RISK & UNCERTAINTY 15, 17 (2009) (reporting variations in estimates across income levels).

132 Gruber & Koszegi, *supra* note 77, at 1292; JONATHAN GRUBER & BOTOND KŐSZEGI, A MODERN ECONOMIC VIEW OF TOBACCO TAXATION 17 (2008) (estimating internality-correcting price of about $14 per pack).

133 Galle, *supra* note 122, at 1730–34.

134 Kaplow & Shavell, *supra* note 52, at 775–79.

135 Galle, *supra* note 122, at 1731–34.

136 *Id.*

137 Schwartz, *supra* note 31, at 1403–04.

138 Allcott & Sunstein, Working Paper, *supra* note 14, at 17; *see* Hunt Allcott et al., *Energy Policy with Externalities and Internalities*, 112 J. PUB. ECON. 72, 76 (2014) (modeling argument that optimal internality tax is always above zero if any consumer is biased, assuming biases are in

if the government has any information about which consumers are more likely to need a $100 correction, it should impose a $100 tax on those consumers, and a $0 tax on others.

ii. New Solutions

In addition to these familiar tools of externality regulation, there also are a host of new techniques, many still in development, aimed at the added data problems raised by internalities. While these tactics may not cure every informational shortfall, they at least free many internality problems from the most serious informational obstacles regulators might otherwise face.

The most familiar of these tools is asymmetric regulation.[139] Asymmetric regulations are more stringent for those who are most likely to make mistakes.[140] The now-classic example is default savings plans, under which employees must actively opt out of making retirement contributions.[141] Under-saving for retirement seems mostly to be caused by inattention and procrastination.[142] These are the same individuals who are the least likely to take the time to fill out the forms needed to opt out of the default savings plan.[143] Meanwhile, active savers who prefer lower or different retirement savings will readily fill out the one-page form, and so bear little cost from the default.[144] Unless even active savers are making mistakes, it follows that asymmetric instruments will always be preferable to an outright mandate.[145] Plan designs that force individuals to decide whether to opt in or out can also be asymmetric; for those for whom decisions are not burdensome, the cost of deliberation is trivial.[146]

Because asymmetric regulation is self-targeting, regulators do not have to be able to identify biased consumers.[147] The regulation applies to everyone. Since

---

one direction).

139 Colin Camerer et al., *Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism,"* 151 U. PA. L. REV. 1211, 1230–37 (2003).

140 *Id.* at 1222, 1225–26; Allcott et al., *supra* note 138, at 74–75.

141 John Beshears et al., *The Importance of Default Options for Retirement Savings Outcomes: Evidence from the United States*, *in* SOCIAL SECURITY POLICY IN A CHANGING ENVIRONMENT 167, 187–92 (Jeffrey Brown et al. eds., 2009).

142 *Id.* at 183–84.

143 *Id.* at 188.

144 *Id.*; *see* Chetty et al., *supra* note 84, at (reporting that "active savers" were highly responsive to changes in incentive savings, implying low transaction costs).

145 Goldin & Lawson, *supra* note 14, at 440.

146 *Cf.* Beshears et al., *supra* note 141, at 188 (advocating "active choice" savings in some settings).

147 Camerer et al., *supra* note 139, at 1222. Bubb and Pildes emphasize that there may be more than one dimension of heterogeneity among a regulated population. Bubb & Pildes, *supra* note 14, at 1621–26. The group of passive savers may include individuals who would have (eventually) saved more than the default savings rate, such that the default actually lowers savings

the cost of non-compliance is so minor for fully rational actors, though, it does not change their behavior.[148] Therefore regulators can apply it to the whole population without worrying much about the risk that it will be misapplied.[149] This same analysis still largely holds if the regulation is more than just a minor inconvenience for the rational.[150] As I explained, when government can set multiple prices or policies, the social cost of imperfectly-informed regulation is much lower. Asymmetric regulations are essentially just regulations with multiple, built-in prices: a high price for irrational actors, a lower price for the fully rational.[151]

Other regulatory techniques work by inducing consumers to reveal their potential biases, enabling regulators then to target the right policy to the right person. A simple example of this "separating equilibrium" approach, discovered accidentally by tax officials, is over-withholding.[152] Millions of taxpayers each year voluntarily allow their employers to withhold more in taxes each pay period than required, or opt to receive their tax refund in one lump sum rather than incrementally over the year.[153] Both qualitative and quantitative studies show that taxpayers are, in effect, using the government as a commitment device, forcing themselves to save until the time of their tax rebate.[154] In general, we should expect that these kinds of opt-in regulations will be attractive to individuals in Box 4: those who recognize their need for, and who value, interventions to

---

for those individuals. *Id.* at 1622. In my view this is a problem created largely by the use of only a single instrument. If government is limited to only one instrument, the default, then there are tradeoffs implicit in the setting of the level of default savings. A more efficient approach would be to use a second instrument to further sort the inattentive savers between those with high and low savings needs. Allcott et al., *supra* note 138, at 77–78; *cf.* Ayres, *supra* note 36, at 2093 (noting possible use of multiple defaults for "subsets" of population).

148 *Id.*

149 *Id.* at 1225–26.

150 *See* Ayres, *supra* note 36, at 2089–91.

151 Galle, *supra* note 122, at 1754. Avishalom Tor, *The Next Generation of Behavioral Law & Economics*, *in* EUROPEAN PERSPECTIVES ON BEHAVIOURAL LAW AND ECONOMICS 17, 25–26 (Klaus Mathis ed. 2015), offers a similar story. Tor explains that legal rules can themselves cause "selection" effects, such as when risky rules tend to cause the affected population to be more risk-seeking.

152 Barr & Dokko, *supra* note 65, at 979; Richard H. Thaler, *Anomalies: Saving, Fungibility, and Mental Accounts*, 4 J. ECON. PERSP. 193, 193-95 (1990).

153 Damon Jones, *Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds*, 4 AM. ECON. J.: POL'Y 158, 158 (2012).

154 Barr & Dokko, *supra* note 65, at 979; Sara Sternberg Greene, *The Broken Safety Net: A Study of Earned Income Tax Credit Recipients and a Proposal for Repair*, 88 N.Y.U. L. Rev. 515, 561–62 (2013); Jones, *supra* note 153, at 159; Damon Jones, *Information, Preferences, and Public Benefit Participation: Experimental Evidence from the Advance EITC and 401(k) Savings*, 2 AM. EC. J.: APPLIED EC. 147, 149 (2010) (reporting reasons the advanced EITC program failed); Ruby Mendenhall et al., *The Role of Earned Income Tax Credit in the Budgets of Low-Income Earners*, 86 SOC. SERV. REV. 367, 377–78, 382, 398 (2012).

bolster willpower.[155]

We could probably tell a similar story about Box 1, the inattentive. Many of us know that we sometimes fail to pay as much attention as we should to some of life's important details.[156] Here, too, there are robust commercial markets for self-help devices, providing evidence that indeed some individuals will value interventions.[157] Our slenderer readers may be unfamiliar with Weight Watchers.[158] The Weight Watchers "points" system is basically just an easily-implemented tool for encouraging participants to pay attention to what they eat and drink, and to count calories and other nutritional data.[159]

A third set of tools, as Schwartz at points acknowledges, is experimentation and data crunching.[160] Consider, for instance, the possibility that some consumers make bad choices because of features of their choice environment, such as in studies finding that long menus of Medicare Part D drug coverage options caused seniors to pick plans that were clearly dominated by other available choices.[161] Schwartz seems to assume that these situations are hopeless, at least in the case where every choice would be rational for some consumers.[162]

As Goldin and Reck recently have shown, however, government can design experiments that at least would reveal what share of the population's choice has been changed by its framing.[163] "Consistent" consumers, whose choice is

---

155 Nava Ashraf et al., *Tying Odysseus to the Mast: Evidence from a Commitment Savings Device in the Phillipines*, 121 Q.J. ECON. 635 (2006); Esther Duflo et al., *Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya*, 101 AM. ECON. REV. 2350 (2011); *see also* Ashvin Gandhi & Michael Kuehlwein, *Reexamining Income Tax Overwithholding as a Response to Uncertainty*, 43 PUB FIN. REV. 220, 222 (2016) (reporting evidence that rules out most plausible rational-actor explanation for overwithholding).

156 See Raj Chetty et al., *Salience and Taxation: Theory and Evidence*, 99 AM. ECON. REV. 1145, 1170–74 (2009) (modeling behavior of households aware of their own inattention).

157 Michael S. Barr et al., *Behaviorally Informed Regulation*, *in* BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY 440 (Eldar Shafir ed., 2013). A notable data point here is the continuing popularity of software that allows us to plan and reminds us of those plans. A number of studies, summarized in Brigitte Madrian, *Applying Insights from Behavioral Economics to Policy Design*, 6 ANN. REV. ECON. 663, § 3 (2014), find that planning and reminders improve savings, vaccination rates, education performance, vehicle safety, and healthy eating.

158 WEIGHT WATCHERS (2018), https://www.weightwatchers.com/us/.

159 *Our Approach*, WEIGHT WATCHERS (2018), https://www.weightwatchers.com/us/our-approach ("Our new SmartPoints™ plan nudges you toward a healthier pattern of eating so that over time, smart choices become second nature").

160 Schwartz, *supra* note 31, at 1380, 1402–03.

161 Jason Abaluck & Jonathan Gruber, *Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program*, 101 AM. ECON. REV. 1180, 1195–96, 1198 (2011); Saurabh Bhargava et al., Chose to Lose? Employee Health-Plan Decisions from a Menu with Dominated Options, unpublished manuscript, Nov. 2014, at 21–22.

162 Schwartz, *supra* note 31, at 1390–93, 1403 n.54.

163 Jacob Goldin & Daniel Reck, Preference Identification Under Inconsistent Choice, unpublished manuscript, Mar. 25, 2015, at 7–29. A less technical version of the same argument is

unaffected by the frame, presumably are choosing based on some set of invariant underlying preferences, and therefore their expressed preferences are reliable evidence of their true preferences.[164] Government can examine which observable features of consistent choosers predict a given set of preferences, and then use the same observables to draw inferences about the preferences of inconsistent choosers whose preferences cannot be directly measured.[165] As I have pointed out elsewhere, these same types of tools can also be used to extend experiments of limited scope, especially field experiments, to conclusions about the population as a whole.[166] While it is well known that some cognitive failings are context-specific or overlapping,[167] this is not a reason to reject experimentalism, but rather a reason to design experiments so that they will have external validity.[168]

Ian Ayres and Quinn Curtis offer the kernel of a similar idea in their recent proposal to reform pension savings.[169] They suggest that government simply test directly for financial sophistication before it allows investors to choose from outside the limited default set of investment options.[170] It will not always be practical or cost-effective to administer individualized testing, especially for decisions that must be made quickly or frequently. But the direct-testing idea can be implemented for other infrequent, high-stakes choices, such as home mortgages, student loans, and health plans.

Saul Levmore also shows that in some instances the informational demands of internality regulation can be low if the causes of internalities are lumpy.[171] That is, suppose the main obstacle to road (or hockey) safety is overcoming the present self's short-term dislike for helmets, which is driven by peer pressure. If government requires a helmet, it can then leave to the individual the choice about how protective the helmet should be. In other words, the regulator does not have

---

Jacob Goldin, *Which Way to Nudge? Uncovering Preferences in the Behavioral Age*, 125 YALE L.J. 226, 260–69 (2015).

164 Goldin & Reck, *supra* note 163, at 4–5; *see also* Saul Levmore, *From Helmets to Savings and Inheritance Taxes: Regulatory Intensity, Information Revelation, and Internalities*, 81 U. CHI. L. REV. 229, 240–41 (2014) (considering possibility that data from older generations could be used to infer preferences of young savers).

165 *Id.* at 5–6; *see also* Allcott et al., *supra* note 138, at 78–79 (suggesting that government can at least infer bounds on extent of bias if consumers are fully responsive to prices).

166 Galle, *supra* note 26, at 862–64; *see also* Raj Chetty, *Behavioral Economics and Public Policy: A Pragmatic Perspective*, 105 AM. ECON. REV. (Papers & Proceedings) 1, 16–19 (2015) (explaining how behavioral data can be used to extrapolate policy improvements).

167 Schwartz, *supra* note 31, at 1392–95.

168 *Cf.* Sendhil Mullainathan et al., *A Reduced-Form Approach to Behavioral Public Finance*, 4 ANN. REV. ECON. 17.1, 17.3 (2012) (laying out a model of cognitive failures that "can be interpreted independent of a specific psychological mechanism").

169 Ian Ayres & Quinn Curtis, *Beyond Diversification: The Pervasive Problem of Excessive Fees and "Dominated Funds" in 401(k) Plans*, 124 YALE L.J. 1476, 1525–29 (2015).

170 *Id.* at 528.

171 Levmore, *supra* note 164, at 234–35.

to determine the optimal level of safety equipment, only to identify the fact that there is an initial barrier that is preventing many actors from choosing the safest level themselves.

B. Government or Markets?

Jonathan Klick & Gregory Mitchell and others further argue against internality regulation on the ground that it simply crowds out what would otherwise have been private responses.[172] They point out that government regulation of internalities may reduce individual incentives to invest in the power to self-regulate, which they call "cognitive hazard."[173] These claims echo the classic argument by Ronald Coase that private actors can potentially negotiate their way around externalities.[174]

As we have seen already, it is not clear the cognitive hazard story makes much sense for Box 3 internalities. Confirmation-biased actors by definition resist the notion that they are biased at all.[175] This may also hold for some Box 1 individuals: while some inattentive actors are aware and seek out help to overcome their own perceived failures, others may not notice their own inattention.[176] In both these cases individuals would not invest much effort in de-biasing themselves, whether the government helped or not.

In other instances there are already private markets for internality correction, but government intervention can be justified even when these private markets function efficiently.[177] One rationale for regulation is a public taste for distributive justice.[178] We may believe that individuals who have the bad luck to

---

172 Jonathan Klick & Gregory Mitchell, *Government Regulation of Irrationality: Moral and Cognitive Hazards*, 90 MINN. L. REV. 1620, 1626, 1633–41 (2006); CSERNE, *supra* note 111, at 53; Gary Becker & Casey Mulligan, *The Endogenous Determination of Time Preferences*, 112 Q.J. ECON. 729 (1997); Epstein, *supra* note 66, at 811–17.

173 Klick & Mitchell, *supra* note 172, at 1633–41; *see also* GARY S. BECKER, ACCOUNTING FOR TASTES 11 (1996) (arguing that individuals can invest in long-run thinking); Schwartz, *supra* note 31, at 1379 (arguing that government protection is unnecessary when consumers can learn from their own mistakes).

174 Coase, *supra* note 39, at 6–15.

175 Kahan, *supra* note 95, at 22–23.

176 *Cf.* O'Donoghue & Rabin, *supra* note 40, at 1841 (making this point about willpower failures).

177 To the contrary, in my view there are good reasons to believe markets for learning are frequently imperfect, whether due to limited opportunities for learnings, Bar-Gill & Warren, *Making Credit Safer*, *supra* note 4, at 7– 25, or other market failings, Brian Galle, *Carrots, Sticks, and Salience*, 67 TAX L. REV. 53, 102–03 (2013); *see* Jeremy A. Blumenthal, *Emotional Paternalism*, 35 FLA. ST. UNIV. L. REV. 1, 51–53 (2007). My argument here holds whether these suppositions are ultimately empirically supported or not.

178 Concerns for justice are consistent with my welfarist framework, since welfarism aims to maximize the satisfaction of all preferences, including society's preferences for fairness. MATTHEW D. ADLER & ERIC A. POSNER, NEW FOUNDATIONS OF COST-BENEFIT ANALYSIS 23 (2006); Louis

suffer from internalities should not be left worse off than others.[179] Many commentators offer similar theories for providing affordable health insurance regardless of economic need: women and individuals with disabilities, for instance, should not be worse off economically than men, simply because their health costs are on average higher.[180]

Obliging internality sufferers to turn to private-market solutions will often leave them worse off because private help is not free. Self-education is difficult and often impractical.[181] In many cases, private market solutions may allow the solution provider to extract all or nearly all of the "surplus," or benefit, from customers.[182] For example, employers who provide pensions help workers to overcome the workers' failure to save adequately for retirement. In addition to sacrificing some salary for this benefit, workers also take on the agency costs inherent in entrusting their bargaining adversary–the employer–with their own well-being.[183]

Even worse, private markets may extract consumer surplus without necessarily overcoming consumers' internalities. Another lesson from studies of tax withholding is that after households engage in costly commitment devices, they later backslide and, when tempted, spend more money trying to accelerate the very payments they voluntarily delayed.[184] These de-commitment services, known as "refund anticipation loans" in the tax refund context, are often subject to enormous fees.[185] Thus, as other commentators have observed, the possibility that private markets can both offer *and undo* self-commitments can be a reason

---

Kaplow & Steven Shavell, *Fairness vs. Welfare*, 114 HARV. L. REV. 961, 989–92 (2001).

179 SHLOMI SEGALL, HEALTH, LUCK, JUSTICE 100–04 (2010); Tom Baker & Peter Siegelman, *You Want Insurance With That? Using Behavioral Economics to Protect Consumers from Add-on Insurance Products*, 20 CONN. INS. L.J. 1, 43 (2013); *cf.* DANIELS, *supra* note 68, at 161–62 (suggesting that inequality of information about health risks can warrant regulation of those risks).

180 SEGALL, *supra* note 179, at 74–86, 101; Eric Rakowski, *Who Should Pay for Bad Genes?*, 90 CAL. L. REV. 945, 1352–67 (2002); *see also* Zamir, *supra* note 27, at 282–83 (offering this rationale in defense of paternalistic regulation more generally).

181 Bar-Gill & Warren, *supra* note 4, at 26–55, and Galle, *supra* note 177, at 104–05, survey the evidence.

182 *See* Galle & Utset, *supra* note 89, at 54–55 (describing this phenomenon in the consumer credit context).

183 Christine Jolls, *Employment Law and the Labor Market*, NBER Working Paper No. 13230, at 4 (2007); Brendan Maher, *Regulating Employment-Based Anything*, 100 MINN. L. REV. 1257, 1296–1303 (2015).

184 *See* Barr & Dokko, *supra* note 65, at 980; Jonathan Parker, *Why Don't Households Smooth Consumption? Evidence from a 25 Million Dollar Experiment*, NBER Working Paper No. 21369 (July 2015) (reporting that borrowing to accelerate refunds is most strongly correlated with measures of household impatience).

185 Barr & Dokko, *supra* note 65, at 979; Chi Chi Wu & Jean Ann Fox, Nat'l Consumer Law Ctr., Inc., Consumer Fed'n of Am., Coming Down: Fewer Refund Anticipation Loans, Lower Prices from Some Providers, but Quickie Tax Refund Loans Still Burden the Working Poor 4, 8–12 (2008).

for government action.[186]

Even if private providers do not extract most of the benefits of internality correction, government provision of commitment devices and the like can be an efficient source of redistribution because it resembles an improved version of the income tax. In standard public finance accounts of the tax system, government can improve over a simple tax on labor earnings if it can identify and tax things that are correlated with the ability to earn income.[187] Briefly, the reasoning is that in a progressive tax system, individuals with high earning potential may "mimic" low-income individuals in order to escape high tax rates; taxing correlates of income rather than income itself makes this mimicking more difficult.[188]

Free internality correction fits into this story. The same logic behind taxing correlates of income also justifies government provision of in-kind benefits, if those benefits are disproportionately useful to individuals with lower earning potential.[189] Many of the key building blocks of internalities, such as impatience, inattention, and addictive behaviors, have been shown to correlate with lower income.[190] Internality correction therefore closely resembles an efficient supplement to the income tax.

Finally, Klick & Mitchell do not consider the possibility that government support could encourage private self-help investments. Government support might, for example, increase the returns to investment: I might invest more effort in learning to rely on commitment devices if the devices available are cheap and effective.[191] Absent government assistance, some individuals might be too demoralized to make use of commitment or procrastinate learning how to do

---

186 David Laibson, *Golden Eggs and Hyperbolic Discounting*, 112 Q.J. ECON. 443 (1997); O'Donoghue & Rabin, *supra* note 40, at 1841.

187 This insight is usually traced to James A. Mirrlees, *Optimal Tax Theory: A Synthesis*, 6 J. PUB. ECON. 327 (1976), and the earlier work by Mirrlees discussed therein.

188 Emanuel Saez, *The Desirability of Commodity Taxation Under Non-Linear Income Taxation and Heterogeneous Tastes*, 83 J. PUB. ECON. 217, 226, 228 (2002).

189 LOUIS KAPLOW, THE THEORY OF TAXATION AND PUBLIC ECONOMICS 227 n.10 (2008); Helmuth Cremer & Phillipe Pestieau, *Redistributive Taxation and Social Insurance*, 3 INT'L TAX & PUB. FIN. 281, 282 (1996); Jean-Charles Rochet, *Incentives, Redistribution, and Social Insurance*, 16 GENEVA PAPERS ON RISK & INSURANCE THEORY 143, 160–64 (1991).

190 For evidence on impatience, see James J. Heckman et al., *The Rate of Return to the Highscope Perry Preschool Program*, 94 J. PUB. ECON. 114 (2010); Walter Mischel et al., *Delay of Gratification in Children*, 244 SCIENCE 933 (1989); Brian C. Cadena & Benjamin J. Keys, *Human Capital and the Lifetime Costs of Impatience*, 7 AM. ECON. J.: ECON. POL'Y 126 (2015). For salience, see Jacob Goldin & Tatiana Homonoff, *Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity*, 5 AM. ECON. J.: ECON. POL'Y 302 (2013). For addiction, see, e.g., Karen M. Jennison, *The Short-Term Effects and Unintended Long-Term Consequences of Binge Drinking in College: A 10-Year Follow-Up Study*, 30 AM. J. DRUG & ALCOHOL ABUSE 659 (2009), or watch a few episodes of *The Wire*.

191 *Cf.* Gruber & Koszegi, *supra* note 77, at 1284 (arguing that knowledge of a future price increase may help motivate addicts to quit addictive behavior before price increase).

so.[192] In essence, regulation serves as a commitment matching grant, multiplying the public's investments.

V. CHOICE OF INSTRUMENTS: CARROT, STICK, OR COMPROMISE?

Assuming that officials have committed to regulating internalities, how should they do it? Should we ban dangerous drugs? Tax them? Subsidize alternatives and treatments? Or oblige producers to print visceral and frightening images of the drugs' consequences on the side of each package? These choices among regulatory options or "instruments" is a central question for modern regulation.[193]

I have argued before that there are four basic categories of regulatory instrument.[194] Two of the categories involve explicit transfers of wealth, while the other two do not.[195] If government selects an instrument that transfers wealth, it must decide whether the regulated party will, relative to the pre-existing baseline, be made to pay for non-compliance (a "stick") or be rewarded for compliance (a "carrot").[196] The transferless categories include standard "command and control" regulation, such as bans or caps on production.[197] Many nudges resemble command and control regulation, in that they also do not involve any explicit wealth transfer.[198] But other features of the nudge are so different and, as Sunstein and Thaler say, "surprising," from a classical rational-actor perspective,[199] that I put them in a fourth category by themselves.

As we saw briefly in Part I.A., there now is a considerable literature examining the factors that officials should consider when they choose between these classes of instrument. Instruments may differ, among other ways, in their propensity to create moral hazard, in their effects on demand for the regulated product, in their impact on the public fisc, and in their distributive fairness.[200] While economists often favor wealth-transfer instruments over others,[201] the

192 *See* Galle & Utset, *supra* note 89, at 76–78 (introducing the concept of "meta-procrastination" and discussing its sources and implications).

193 Helfand et al., *supra* note 45, at 249–53.

194 Galle, *supra* note 26, at 848–54. I actually simplify my earlier categories a bit here for the sake of narrative economy. A more complete categorization would also include the division between *ex ante* and *ex post* and would sub-divide price instruments into priced (those denominated in dollars) and transfer (those that shift resources). Transfer instruments can be even further refined to distinguish between public and private transfers.

195 *Id.*

196 *Id.* at 851.

197 *Id.* at 848–49.

198 *Id.* at 846–47.

199 THALER & SUNSTEIN, *supra* note 12, at 85–86, 252–54.

200 Galle, *supra* note 26, at 848–53.

201 Helfand et al., *supra* note 45, at 287; Edward L. Glaeser, *Paternalism and Psychology*, 73 U. CHI. L. REV. 133, 150 (2006); Kaplow & Shavell, *supra* note 52, at *passim*.

choice between carrot and stick typically obliges us to trade off favorable results on some of these factors against unfavorable outcomes on others, implying that sometimes a reason to choose a transferless instrument is that it may represent a compromise position.[202]

Although prior literature has extensively considered these factors in the externality context, it has not to my knowledge examined the extent to which the received wisdom applies to internalities. This Part begins that task.[203]

A. Moral Hazard

Moral hazard is usually the strongest argument against carrots.[204] When government offers rewards to polluters to stop their misdeeds, it encourages new polluters to begin emitting so that they, too, can be paid to stop.[205] The underlying game theory logic is similar to the rationale of the United States government when it refuses to pay ransom to kidnapers.[206] In contrast, the promise of future liability encourages polluters to invest in pollution-reduction technologies, especially if liability may turn out to be retroactive;[207] rewards provide no incentive to remedy harm before the reward is paid, and may crowd out any voluntary compliance.[208] Even when distributional concerns might weigh in favor of carrots,[209] the possibility of these kinds of moral hazard strongly counsels in favor of at least partial stick liability for externality producers.[210]

---

202 Galle, *supra* note 26, at 872–79.

203 I emphasize "begin." A work of this length necessarily must omit many considerations, some potentially important. For example, the form of regulation may affect compliance and the public's relation to the law. Yuval Feldman & Orly Lobel, *Behavioral Tradeoffs: Beyond the Land of Nudges Spans the World of Law and Psychology*, *in* NUDGE AND THE LAW: A EUROPEAN PERSPECTIVE 336, 342–50 (Alberto Alemanno & Anne-Lise Sibony eds., 2015). Private actors who profit from consumers may respond to government interventions and understanding how these responses might vary across instruments is an important future direction for researchers. *Cf.* Ran Spiegler, *On the Equilibrium Effects of Nudging*, 44 J. LEGAL STUDIES 389 (2015) (offering formal models of firm responses to several nudges).

204 Coase, *supra* note 39, at 42; Wiener, *supra* note 35, at 726 & n.186.

205 *Id.* For qualifications to this story, see Galle, *supra* note 35, at 822–23.

206 *See* ROBERT MNOOKIN, BARGAINING WITH THE DEVIL: WHEN TO NEGOTIATE, WHEN TO FIGHT 7 (2010).

207 Saul Levmore, *Changes, Anticipations, and Reparations*, 99 COLUM. L. REV. 157, 1663 (1993).

208 WILLIAM J. BAUMOL & WALLACE E. OATES, THE THEORY OF ENVIRONMENTAL POLICY 212 (2d ed. 1988) (1975).

209 Gerrit De Geest & Giuseppe Dari-Mattiacci, *The Rise of Carrots and the Decline of Sticks*, 80 U. CHI. L. REV. 341, 363–65 (2013).

210 In essence, a partial stick is similar to a government co-pay requirement on its social insurance policy, and is efficient for similar reasons: it trades off the worst of moral hazard against only a small loss on redistribution. Raj Chetty & Amy Finkelstein, *Social Insurance: Connecting Theory to Data*, *in* V HANDBOOK OF PUBLIC ECONOMICS 112, 157–58 (Alan J. Auerbach et al. eds., 2012).

This version of the moral hazard argument seems tenuous when it comes to internalities. A rational actor cannot credibly kidnap themselves.[211] Nor would it make sense to delay self-help in order to encourage government payment, unless somehow the individual expects that the government will greatly over-pay.

Admittedly, carrots could contribute to moral hazard at the margins. Government interventions of all kinds can encourage risky behavior, as Klick & Mitchell emphasize.[212] If I know that a soda tax will help motivate me to eliminate any soft drink habit, I may be more willing to take the first sip. If I expect instead to be paid to switch to bottled water, my calculus perhaps tips a bit further towards downing that initial Coke. Rewards might also impact timing. A newly-introduced carrot might encourage "sophisticated" high-beta consumers to wait to kick their bad habit until the government assistance arrives, while a new tax would encourage quitting before the tax.[213]

For now, though, the possible absence of significant *incremental* moral hazard from carrots is enormously important to internality regulation. With moral hazard out of the way, the choice between carrot and stick is much closer.[214] Carrots for internalities deserve serious consideration.[215] Let's continue a bit further in that direction.

B. Income Effects

As other writers recognize, income effects can be important in the choice of regulatory instrument.[216] Income effects are the just the tendency of most consumers to demand more of a good when they have more money. A relatively unusual exception is the so-called "inferior" good, which is a good that people tend to want less when their wallet is fatter; ramen noodles may be a familiar

---

211 But see *Other People's Money* (1991).

212 *See supra* notes 172–176.

213 *Cf.* Gruber & Koszegi, *supra* note 77, at 1273–77 (finding that smokers are responsive to anticipated future tax increases). It also is possible that some interventions may be more likely to cause crowding out of private effort than others. For example, maybe resentment towards fines imposed on helmetless riders tends to strengthen confirmation bias among motorcycle riders, making them even more resistant to news about the dangers of riding. This is another area where better empirical evidence would be useful.

214 Galle, *supra* note 26, at 878.

215 My analysis of carrots here is limited to what I'll call traditional carrots, those that have no unexpected impacts on decisions. It may be possible to use rewards as the centerpiece of a "surprising" regulatory intervention, in which the behavioral impact is far larger than we would expect given the size of the reward. See Kevin Volpp et al., *Financial Incentive-Based Approaches for Weight Loss*, 300 JAMA 2631, 2631–36 (2008) (describing effectiveness of constant stream of very small but immediate rewards in changing the behavior of present-biased actors). Since the fiscal costs of these approaches is typically quite small, I will group them with nudges and other transfer-less instruments.

216 STERNER, *supra* note 50, at 167.

example to college-educated readers. Proponents of soda taxes note that because of income effects, subsidies for healthy alternatives could have a perverse impact.[217] By enriching a household, subsidies may also increase its demand for all consumables, including unhealthy food. This shift implies a higher optimal subsidy, as when consumers are more reluctant to give up their junk food, government suasion must be more powerful.[218]

The opposite is true if government uses sticks rather than carrots, or if government uses carrots but the internality-creating item is an inferior good. In those cases, income effects will depress consumption, allowing for a lower optimal tax or subsidy rate.

Regulation of internalities may also change a household's perception of its own wealth, but this is unlikely to affect the choice of instruments analysis. If we can achieve the same internality correction in any of three ways (carrot, stick, or transfer-less instrument), presumably any income effect that comes with internality correction will be the same under each.[219] For example, helping families save for retirement might make them feel richer; since consumption is a normal good, that feeling of greater wealth might increase current expenditures at the cost of savings.[220] But this effect would be equally true whether we penalize non-savers or reward savers. It is only the incremental income effects that result from the carrot/stick choice that factor into which of those options we would want.

Of course, that would not be true if our choice of instrument also affected how individuals perceive the benefit of internality-correction. For instance, we saw before that inattentive eaters might not be aware that a "nudge" has changed their eating habits. They might be somewhat more cognizant of monetary rewards, leading them to better understand that they now are making better use of their money (and therefore are effectively richer). I explore the relation between salience and income effects in more detail elsewhere.[221]

Ultimately, then, the importance of income effects is hard to predict without more empirical evidence. There is some suggestion already that "junk" food and

---

217 Gideon Yaniv et al., Junk-Food, *Home Cooking, Physical Activity and Obesity: The Effect of the Fat Tax and the Thin Subsidy*, 93 J. PUB. ECON. 823, 826–27 (2009).

218 *See* KAPLOW, *supra* note 128, at 496–97 (analyzing impact of income effects on optimal pigouvian tax). Of course, when government increases the optimal subsidy, it further worsens the income effect, and so on. But typically, the substitution effect of the subsidy is more powerful than the income effect, and so this iterative sequence converges to a solvable optimum.

219 *See* Louis Kaplow, *Myopia and the Effects of Social Security and Capital Taxation on Labor Supply*, NBER Working Paper No. 12452, at 7 (Aug. 2006) (discussing of the potential income effects of improving a household's inter-temporal budget allocation).

220 *See* GRUBER, *supra* note 45, at 650 (stating that income effects tend to reduce savings because present consumption is a normal good).

221 Galle, *supra* note 177, at 66–67, 86–89, 93.

cigarettes might be inferior goods, at least in the contemporary United States.[222] If other internality-creating choices are also inferior goods, then the argument for carrots becomes much stronger. In that case, the income effect of the carrot boosts, rather than undermines, the substitution effect of the price instrument. If not, alternatives such as sticks or transferless instruments look better by comparison, unless it were the case that carrots had particularly low income-salience.

C. Revenues

Another central issue in the choice of instruments is the instrument's effect on the public fisc. Carrots, of course, must be paid for.[223] Carrots aside, in most of the prior literature, the fact that some instruments provide funds that can be transferred to others in large measure explains the dominance of "price," or what I have called "transfer," instruments, over other choices.[224]

O'Donoghue & Rabin suggest that internalities present an especially strong example, going so far as to argue that Pigouvian-type taxes on internalities can be justified solely from a revenue perspective.[225] This was an argument that, in the pollution context, had been known in the 1990's as the "revenue recycling" or "double dividend" claim: the possibility that carbon taxes both could clean the environment and also raise money more efficiently than other sources.[226] Allcott & Sunstein assert, without reference to O'Donoghue & Rabin, that there is a double dividend from taxing internalities, but as we'll see the story is rather more complicated.[227]

The availability of a double dividend is, fittingly, important to my analysis for two reasons. For one, a double dividend would offer a reason to impose a tax on internalities, regardless of whether we otherwise believe internality regulation is a good idea: sin taxes would just be a particularly good way to raise money. The other significance of the double dividend is that it represents an opportunity cost if policy makers choose to use command and control regulation or nudges,

---

222 R.J. DeGrandpre et al., *Effects of Income on Drug Choice in Humans*, 59 J. EXPERIMENTAL ANAL. BEHAVIOR 483, 483 (1993) (generic cigarettes); Leonard H. Epstein et al., *Purchases of Food in Youth: Influences of Price and Income*, 17 PSYCH. SCIENCE 82, 88 (2006) (junk food); Matthew Harding & Michael Lovenheim, *The Effect of Prices on Nutrition: Comparing the Impact of Product- and Nutrient-Specific Taxes*, 53 J. HEALTH ECON. 53 2017) (junk food).

223 Madrian, *supra* note 157, at § 2.

224 *See* sources cited *supra* note 201.

225 O'Donoghue & Rabin, *supra* note 40, at 1829, 1832–33.

226 *E.g.*, Charles L. Ballard & Steven G. Medema, *The Marginal Efficiency Effects of Taxes and Subsidies in the Presence of Externalities: A Computational General Equilibrium Approach*, 62 J. PUB. ECON. 199, 200 (1993).

227 Allcott & Sunstein, Working Paper, *supra* note 14, at 8.

rather than taxes or other transfer instruments.

i. Is There a Double Dividend?

To understand O'Donoghue and Rabin, we first have to review briefly why it is that environmental economists came to mostly reject the "double dividend" theory. In brief, as Bovenberg & Goulder summarize in their handbook entry, the problem is that carbon taxes are differentiated consumption taxes.[228] As a result, carbon taxes are likely not only to distort consumers' choice of what to buy, but also to affect the after-tax return on labor income. That is, since we generally work in order to buy stuff, taxes on the stuff itself affects not only our choice about what to buy but also our labor/leisure decisions.[229] While carbon taxes may have some desirable effects, they also frustrate some consumers' preferences for high-carbon goods, and discourage labor supply.[230]

Boiling Bovenberg & Goulder's more complex calculations down into a simple equation, the effect of a carbon tax is:

$$E - C - L + R$$

where E is the gains from carbon reduction, C the loss from consumers' choice to switch to a second-best product, L the compensated labor-supply impact, and R any available gains from cutting other distorting taxes.[231]

This equation implies that pollution taxes are usually a less efficient revenue source than the income tax. Set aside E for now; these environmental gains can

---

228 A. Lans Bovenberg & Lawrence H. Goulder, *Environmental Taxation and Regulation*, *in* 3 HANDBOOK OF PUBLIC ECONOMICS 1471, 1486–1507 (Alan J. Auerbach & Martin Feldstein eds., 2002)

229 Anthony B. Atkinson & Joseph E. Stiglitz, *The Design of Tax Structure: Direct Versus Indirect Taxation*, 6 J. PUB. ECON. 55, 56 (1976)

230 Bovenberg & Goulder, *supra* note 228, at 1500.

231 We could also write a slightly more complicated version of the formula, with the added complexity perhaps justified by the fact that this version captures an important nuance. This more-nuanced version would look something like:

$E - C/p_C^2 - L/p_L^2 + R/p_R^2$

Here, C and L stand for something a bit different: instead of the total welfare loss, they are the total loss, in dollars, caused by the tax's distortive effects. These losses are spread out over a population, p, which differs for each factor depending on the incidence of that factor. To translate the dollar loss into a utility loss, we divide the dollars by the square of the population that experiences the dollar loss. This reflects, roughly, the basic proposition in tax economics that the deadweight loss of a tax rises in proportion to the square of its rate: small taxes do not inconvenience us much, but large taxes motivate ever-larger distortions in our behavior. By spreading a burden out over more people, we reduce its per-capita impact, effectively cutting the rate. This burden-spreading point is important to Bovenberg & Goulder. For them, it helps to explain why R can never offset C and L. By bringing in carbon-tax revenue, the government can cut the income tax by the same amount, producing a beneficial effect R/p_R^2. But the income tax affects many more people than will be impacted by at least some forms of carbon pricing. R and L are equal, by definition. If p_R is much greater than p_L, the R/p_R^2 term will always be smaller than L/p_L^2.

be achieved through non-tax policies. Bovenberg and Goulder argue that L and R will at best cancel each other out, because the carbon tax, like the income tax, discourages labor supply.[232] Since L and R each involve identical dollar amounts, the result is that the carbon tax is strictly worse than the income tax by the amount of the distortion, C.[233]

O'Donoghue and Rabin in essence point out that this result does not hold for internalities when taxes are too small to create any effects on labor supply.[234] By definition, correction of the internality on net improves consumer welfare, at least from the point of view of the government planner.[235] If there are negligible labor-supply effects, we could write the resulting equation as:

$$U = I - C + R$$

where I replaces E as gains from the government policy, and by assumption $I - C$ is always positive. That seems quite intuitive; O'Donoghue and Rabin's main contribution is to show that it sometimes is plausible one could achieve meaningful changes in behavior with taxes that are too small (in their total burden, not necessarily their rate) to affect labor/leisure decisions.[236]

But what about when internality-correcting taxes must be large in this sense? Cigarette taxes, for example, can consume a meaningful fraction of household income for low-income families.[237] In many cases, it is not convincing to ignore potential labor-supply (or other tax-avoidance) effects.

A 2006 article by Louis Kaplow provides some starting points for our analysis.[238] Kaplow analyzes the labor-supply effects of an actuarially fair social

---

232 Bovenberg & Goulder, *supra* note 228, at 1500.

233 *Id.* at 1501–02.

234 O'Donoghue & Rabin, *supra* note 40, at 1834–35.

235 *Id.* at 1829; *see also* Haavio & Kotakorpi, *supra* note 108, at 575.

236 O'Donoghue & Rabin, *supra* note 40, at 1836–38. Another contribution the pair offer is that internality-correcting taxes can be efficient even if the population is heterogeneous, such that taxes in some cases are falling on individuals with no internality at all. The intuition is that taxes on rational actors, if "small," are relatively non-distorting, because at the margin the rational actor is indifferent between her two options; if taxes change her choice, the welfare loss is correspondingly small. That is, the contribution of rational actors to C in my formula is minor. At the same time, because internality sufferers are far from their private optimum, small changes can produce large welfare gains (due to the exponential relationship between deadweight loss and the size of the error). Thus, I will generally be quite large relative to C. *See also* Mullainathan et al., *supra* note 168, at 17.16–17.17 for a slightly more elaborate model of the same idea.

237 *See* Jonathan Gruber & Botond Kőszegi, *Tax Incidence When Individuals Are Time-Inconsistent: The Case of Cigarette Excise Taxes*, 88 J. PUB. ECON. 1959, 1962 (2004) (estimating that smoking consumes about 3% of the budget of households in the bottom quartile of income); Gary Lucas, *Saving Smokers from Themselves: The Paternalistic Use of Cigarette Taxes*, 80 U. CIN. L. REV. 693, 693 (2012) (estimating that a New York pack-a-day smoker pays $2,500 per year).

238 Kaplow, *supra* note 219. Portions of the 2006 paper were published as Louis Kaplow, *Targeted Savings and Labor Supply*, 18 INT'L TAX & PUB. FIN. 507 (2011). As the working paper

security system.[239] He shows that the labor-supply effects of compelled savings for retirement may depend on taxpayer beliefs about whether their own savings decisions are optimal.[240] Let us consider that point in the context of the four boxes I set out in Part II.

Of the internality examples I have examined so far, Box 3 of Fig. One most closely resembles the standard externality case sketched by Bovenberg & Goulder. These are individuals, again, who perceive change as costly now and see little long-term benefit to compliance. Think of a tobacco tax. Individuals who change their tobacco consumption habits are likely to experience that shift as very costly. Further, individuals who believe that they do not face major long-term costs of smoking will perceive themselves as obtaining a lower utility for each dollar of earnings, potentially diminishing labor supply or shifting labor to less-productive uses. At the same time, there are revenue gains from the tax, and the policy planner recognizes that there is in fact a long-run internality health gain for the individual, yielding the simple equation:

$$U = I - C - L + R \qquad \text{(Eq. 1)}$$

This of course is the same as Bovenberg & Goulder's equation, with internalities replacing externalities.

Now consider Box 4, where individuals have high B and also high C. Imagine, for instance, the reaction to an alcohol tax by an actress we could call "Lindsay," who recognizes her alcohol dependence but lacks the willpower to end it.[241] Lindsay would experience the loss of her morning whiskey as a drop in her short-run wellbeing, even while she recognizes that there are long-run gains from abstaining. To the extent she does not climb fully onto the sobriety wagon, government would have revenue from taxing alcohol.

What is the effect on Lindsay's labor supply? As in Kaplow's story of the social security tax, arguably Lindsay should recognize that the government, by improving the way in which she has chosen to allocate her spending, has on net increased her returns to labor.[242] This effect should at least partly counteract the labor-discouraging impact of the tax, and might even on net increase her labor.[243]

---

version includes much more extensive discussion of several subjects of interest here, citations are to the original.

239 *Id.* at 4–13. Daniel Shaviro, *Multiple Myopias, Multiple Selves, and the Undersaving Problem*, NYU Law & Economics Working Paper No. 382 (Aug. 2014), http://lsr.nellco.org/cgi/viewcontent.cgi?article=1386&context=nyu_lewp, considers a similar case but relaxes the assumption of actuarial fairness.

240 Kaplow, *supra* note 238, at 12–13.

241 Any resemblance of our anecdote to any real individuals, living or dead, is strictly coincidental.

242 Kaplow, *supra* note 238, at 7.

243 Lindsay may also be able to work more as a result of her greater sobriety. Some portion of that improvement should be reflected in the I term in our equation. There may also be externality

At the same time, there are income effects that point in the opposite direction from the substitution effects. For example, since Lindsay is poorer as a result of the liquor excise, she will have to work harder in order to pay her rent. Conventionally, analyses of the labor impact of taxation omit this effect, because by assumption the consumer is "compensated" by the use of the tax to pay for public goods, which offset the loss of individual wealth.[244] But there is no such compensation for another kind of income effect Lindsay experiences. When the government helps her to set her priorities in order, she in effect is wealthier: by incentivizing her to spend less money on booze, government gives her more money to pay rent.[245] Since she now has more money available to pay rent, the government's assistance should tend to diminish her labor supply.[246]

By adding subscripts to the labor component to reflect the differing impact of the two competing substitution effects and the income effect, we could write the resulting social welfare calculation. Let L_{sb} represent labor supply responses from those who see the tax as benefitting their own well-being, and L_{st} as the traditional labor-supply impacts of a tax. L_i will be the labor impact of the income effect. Then we have:

$$U = I - C + L_{sb} - L_{st} - L_i + R \qquad \text{(Eq. 2)}$$

Equation 2 implies that there is some possibility of a double dividend from taxing internalities of this kind. Even setting aside I, it is possible that the remaining terms net out to a positive, implying that there would be social gains from using the Pigouvian tax to replace other sources of revenue.[247] Whether this is so in the real world would depend on the relative labor gains and losses from better allocating consumers' budgets.

Box 1 offers an even stronger case for double dividends. Recall that these individuals tend not to notice either changes in their behavior or to contemplate the long-run effects of those changes. What is the effect of a tax on inattentive behaviors?

Changes to the behavior of the inattentive may have minimal effects on

---

benefits for others from Lindsay's efforts. But for simplicity I will focus on the pure internality case.

244 GRUBER, *supra* note 24, at Ch. 21.

245 Kaplow, *supra* note 238, at 9; *cf.* Chetty et al., *supra* note 156, at 1173–74 (exploring possible income effects of improved allocation of resources for irrational consumers).

246 Some impulsive individuals may be particularly bad at planning their household finances. These actors—and actresses, in some cases—might not be as careful in matching labor supply to their needs or desires. If so, changes in their effective income may not produce expected labor-supply effects. *Cf.* Chetty et al., *supra* note 156, at 1173–74 (discussing how household's ability to allocate resource decisions affects social planner's welfare calculation).

247 Given the possibility of private "commitment devices" for individuals in Box 4, the social planner may prefer to calculate I as the incremental gains, if any, from public provision.

either consumer welfare or labor supply.[248] If I have little idea how much soda I drank, it is unlikely that I will perceive changes in that amount as affecting my short-run well-being.[249] And, since I am unaware that anything meaningful has changed, I have no reason to adjust my labor supply. Of course, it is also quite possible that incentives that depend on conscious responses, such as tax or tax incentives, also will not do much to change my behavior, producing little internality gain,[250] but that would make them potentially a very effective source of revenue.[251] In the best-case scenario, assuming minimal consumption losses or labor effects, but modest internality gains (for instance, because consumers are more attentive to price than they are to portion size) we could write the welfare effects of the tax on inattentive consumers as:

$$U = I + R \qquad \text{(Eq. 3)}$$

which would represent an unambiguous double dividend.[252]

On the other hand, taxes aimed at the highly inattentive could also be highly inefficient, depending on how they are designed and how taxpayers respond. For example, repealing existing tax incentives for retirement savings might be perceived by inattentive savers as an increase in the tax on labor.[253] A tax on non-savers could be similarly ineffective, if inattentive savers are attentive to labor taxes. The penalty would distort labor income as much as any other income tax, but if the non-savers do not know why they're being punished it may not change

248 Galle, *supra* note 26, at 867–68.

249 *See* ELSTER, *supra* note 10, at 8 (noting that inattentive agents cannot respond to incentives); Kahneman, *supra* note 58, at 1451 ("The operations of System 1 . . . are difficult to control or modify").

250 O'Donoghue & Rabin, *supra* note 40, at 1835; Shaviro, *supra* note 239, at 34–43. For a survey of the evidence on taxpayer responsiveness to "low-salience" taxes, see Galle, *supra* note 177, at 63–67; David Gamage & Darien Shanske, *Three Essays on Tax Salience: Market Salience and Political Salience*, 65 TAX L. REV. 19, 26–53 (2011).

251 Usually, the welfare loss from a "hidden" tax of this kind is that the consumer obtains a different basket of goods than she would have purchased had she been aware of the tax. Chetty et al., *supra* note 156, at 1173. By assumption here, the consumer does not have clearly defined preferences for the amount of the internality good. But the tax would reduce her consumption of other goods, to an extent she likely would not have chosen if she were aware of the tax. Consumers may be able to minimize these distortions if they notice their shrinking budget and can re-optimize purchases accordingly. *Id.* at 1174. So, while labor distortions are lessened, there may still be welfare losses from this misallocation of the taxpayer's budget. If so, we should add a $-C$ term, which would make the case for a double dividend ambiguous.

252 Kaplow, *supra* note 238, at 8 & n.10, appears to tell a similar possible story about willpower failures. He suggests that, if workers wrongly believe at the time they set labor effort that the government will not impose any commitments on them, there will be no labor supply effects. But unlike the inattention story I describe, there is no known empirical evidence of this behavior. In a more realistic setting in which labor supply is constantly being decided, it is unclear how long workers could persist in their mistaken beliefs about the government's response. At some point, Kaplow's willpower story becomes an inattention story.

253 Kaplow, *supra* note 238, at 1–2.

behavior in the desired direction. More research on these issues would be welcome.

In sum, whether an internality-correcting tax provides a double dividend depends on the nature of the mental processes that affect consumer choice. Further complicating the analysis, there may be a mix of processes that produce similar outcomes. Some smokers may fall into Box 3, others in Box 4; that is, some may want help to quit, while others believe they do not need to "right now."

Optimal policy choice then may depend on government's ability to correctly identify the mix of each. This task may not be as daunting as it sounds. For example, we expect that individuals in Box 4 will accept commitment devices where available (barring, say, ideological objections to accepting any help from the government), while individuals in the other boxes would not. This difference allows government to establish policies that induce individuals in Box 4 to reveal their type.

ii. Double dividends and choice of instruments

Assuming that in some cases there may be a double dividend, let's look at how that possibility would affect the government's choice of instruments. Here again, it will make a big difference which "box" our regulated party falls into. Once more, prior commenters have argued that the transfer of resources from the regulated party to other private individuals or the public makes price instruments unambiguously superior, on welfare terms, to transferless regulation.[254] That is, since the welfare effect of a carbon tax is $E - C - L + R$, while the effect of a similar command & control regime is presumably $E - C - L$, the tax is always superior to the extent of any revenue gains.[255]

The same could be true of internality regulation. To the extent that consumers are aware of the effect of the transferless regulation, we might expect the regulation to provide the same labor-supply effects that a tax would. That is, if the regulation is experienced as an unwanted psychic cost (the hassle of going back for a second tiny cup of soda, say), it will diminish the pleasure of the drinking experience, lowering the returns to labor.[256] If government policies are perceived as helpful because they lead to better allocation of the consumer's spending choices, transferless regulation should match the impact of a tax or other "stick," but lack the corresponding revenue.

In earlier work, I attempted to show that this syllogism is not true to the extent that government has available transferless instruments that have lesser

254 *See* sources cited *supra* note 201.

255 *Id.*

256 Glaeser, *supra* note 201, at 150.

effects on consumer welfare and labor supply than the carbon tax.[257] For example, nudges and other "surprising" instruments may sometimes change consumer behavior without consumers necessarily noticing that much important has changed.[258] If so, then it becomes ambiguous which instrument is better on basic utility terms, as $E - C - L + R <> E - C$.

To be more concrete, consider the choice between raising tobacco taxes and the global efforts to label cigarette packaging with disturbing images of adverse health outcomes.[259] Let's posit that some fraction of smokers fall in Box 4 of Figure 1; that is, they would prefer to quit but lack the willpower to do so, and appreciate government efforts to motivate their cessation efforts.

It is unclear if the images are inferior on revenue terms. Both options, if effective, would encourage greater labor effort by improving the smoker's perceived returns to working (because she smokes less, improving her health), while reducing incentives to work because the smoker will perceive herself to be richer. The higher tax, in addition, brings in revenue, and as we just saw may even on net produce a double dividend. The disturbing images, of course, do not bring in money (and may reduce revenue if the government maintains a tobacco excise at a lower rate).

Whether nudges are the better choice than taxes turns on the relative labor effects of the nudge and the tax. Do disturbing images make smokers feel less inclined to work, because the discomfort they feel when they smoke reduces the total reward they can buy with their labor effort? If so, then taxes are superior: both choices have similar labor-supply effects, while taxes bring in money. But if not, then even if there is a double dividend from cigarette taxes the nudge is likely a better choice. Taxes have lower labor and more revenue, while the nudge has less revenue but more labor supplied.[260] That is, the nudge is superior if

$$(I - C + L_{sb} - L_{st} - L_i + R < I - C + L_{sb} - L_i) = (R < L_{st}) \qquad \text{(Eq. 4)}$$

I showed in my earlier work that under standard assumptions $R < L_{st}$: the nudge is preferable to the tax.[261]

That is also the case, *a fortiori*, for Box 3 taxpayers. As we saw, there is no double-dividend scenario for taxing individuals in Box 3.[262] If nudges are better even when using them means giving up a double dividend, they must be better when it does not.[263]

---

257 Galle, *supra* note 26, at 867–71.

258 *Id.* at 867–68.

259 *See* sources cited *supra* note 17.

260 Note that since labor is taxed, the extra labor supply under a nudge also has a revenue benefit. For simplicity, I simply assume that this benefit is already reflected in the L terms.

261 Galle, *supra* note 26, at 869–71.

262 *See supra* notes 240–241.

263 In terms of my simple equations, a nudge is better for Box 3 taxpayers when $I - C - L + R$

Perhaps surprisingly, nudges may lose out for Box 1 individuals, the inattentive. In the best-case scenario for taxation I just sketched, inattentive actors do not adjust labor supply in response to changes in their consumption. Switching to a nudge then sacrifices the double dividend, without gaining any off-setting labor supply benefits. In math terms, $I + R > I$.

In sum, it looks preliminarily as though "nudges" and other surprising instruments are the best choice for consumers who experience willpower or bolstering problems, while sin taxes make more sense for the inattentive. Again, we do not yet fully understand the labor-supply effects of many internality-correction options. If the labor supply effects of sin taxes for the inattentive fall short of the best-case scenario, that might tip the balance back towards nudging.

iii. A Note on Non-Labor Distortions

Recent work in tax economics suggests that the impact of taxes on labor supply may be only a small portion of the total impact of most taxes.[264] Instead, the deadweight loss of taxation is mostly caused by other behavioral shifts individuals undertake in order to avoid tax— for instance, individuals may choose to go into business for themselves so that their income cannot be reported to the government.[265] Depending on other factors, the diminished importance of labor distortions can strengthen the argument for a pigouvian tax.[266]

Applying this framework in the internality context is a complex problem. A key question, certainly, would be to what extent transferless policies inspire the same kinds of avoidance behaviors that a tax would create. On that front we have even less empirical evidence than we do on the labor-supply question. My view is thus that it's too soon to try to build a complete analysis, but this will remain an important open area for future work.

D. Information and Targeting

Another standard argument in favor of price instruments over regulation is that they provide better information about private costs.[267] Typically government cannot directly observe the private marginal cost of compliance.[268] However, a

---

$< I - C - L$, which is to say never.

264 David Gamage, *The Case for Taxing (All of) Labor Income, Consumption, Capital Income, and Wealth*, 68 TAX L. REV. 355, 376–400 (2015).

265 *Id.*

266 *Cf.* John Brooks, Brian Galle, & Brendan Maher, *Cross-Subsidies: Government's Hidden Pocketbook*, 106 GEO. L.J. 1229 (2018) (arguing that Gamage's framework supports use of consumption taxes in some instances).

267 Don Fullerton et al., *Environmental Taxes*, *in* DIMENSIONS OF TAX DESIGN: THE MIRRLEES REVIEW 423, 430 (James Mirrlees et al. eds., 2010). Kaplow & Shavell, *supra* note 52, at 4.

268 *Id.*

price instrument induces rational externality producers to comply if their private costs are less than the price set.[269] By iterating this process, government can gather enough data about private cost structures to better approximate the optimal price.[270] This information is also potentially critical to effective targeting of a policy: government does not want to distort the behavior of those who are already performing in their own self-interest.[271]

Of course, for our purposes the critical assumption in the standard account is that the observed response to price is a rational one. Yet we already know that in many cases it is not.[272] Eighty-five percent of Danish workers ignored a large new tax incentive for retirement savings.[273] Does that tell us that the cost of retirement savings was very high, or just that Danes prefer not to think about Act Five of their lives? It seems that in many cases price instruments are no better than others at revealing private information.[274] We have seen that Schwartz and others rely on this fact as a basis for objecting to any form of internality regulation.

I argued earlier that experiments and self-targeted instruments answer Schwartz's critique, and they also serve to level the playing field between price and other instruments. Because the Danish experiment had a control group of individuals whose costs of savings were indistinguishable from the treatment group—and that control group responded strongly to the tax incentive—we can infer that the unresponsiveness of the bulk of the population was due to behavioral factors, not cost.[275] Well-designed experiments like this allow government not only to identify individuals who need a little help, but also to

---

269 *Id.*

270 Strnad, *supra* note 2, at 1254–55.

271 *Id.*; *cf.* Kaplow, *supra* note 238, at 22 (discussing significance of policy choice when population is heterogeneous in their propensity for error); Tor, *supra* note 151, at 26–27 (same). Allcott & Sunstein rely on a version of the targeting argument to favor energy subsidies over clean-fuel mandates, but their analysis may be a bit off. They suggest that a subsidy will be superior to a "mandate that all consumers take action" because the mandate will cause compliance among some consumers for whom compliance is inefficient. Allcott & Sunstein, Working Paper, *supra* note 14, at 7. This confuses the form of an instrument with its price. Allcott & Sunstein appear to assume that the mandate would apply to every consumer–in effect, that its price would be infinite. But transferless instruments, including many command & control approaches, can have an effective or "shadow" price of any amount. Galle, *supra* note 26, at 862. To take one example, the mandate could exempt any consumer with private compliance costs above what would have been the subsidy amount. This would effectively set the price of compliance at *tau* for either instrument. The duo acknowledges this point later in their discussion, Allcott & Sunstein, *supra* note 14, at 9, so perhaps we should understand their claim simply to be that a flat ban on inefficient energy use by all consumers is bad policy, rather than a general claim about the merits of taxes over mandates.

272 *See* Weiss, *supra* note 3, at 1312.

273 Chetty et al., *supra* note 84, at 1141.

274 *See* Mullainathan et al., *supra* note 168, at 17.8–17.9, 17.22.

275 Chetty et al., *supra* note 84, at 1169–72.

draw inferences about the private cost structures of the targeted groups.[276] In a regulatory environment in which government is already conducting experiments before it regulates, the need to rely on price instruments to reveal information is considerably lessened.

If anything, price instruments might be less appealing in an internality context because they may be more difficult to design as self-targeting. "Linear" taxes, or taxes that apply a uniform per-unit rate, are difficult to make asymmetrical.[277] It's true, of course, that a teetotaler will not pay much alcohol tax.[278] But the large man who *rationally* consumes alcohol at a slow, steady pace will pay far more than the slender woman who irrationally binges. That is poor targeting.

Nonetheless, with some creativity policy makers can likely reduce the inflexibility of tax-like instruments. Consider a system of opt-in taxation.[279] Individuals could agree to be subject to a higher tax rate on some goods. Ian Ayres's StickK, a company that allows users to pledge to pay a penalty if they fail to meet personal goals, has already adopted this method.[280]

Similarly, government could allow households to opt out of government subsidies for overly tempting products. Some states are currently considering a prohibition on junk food for families who receive SNAP benefits,[281] but a more empowering option that would also reveal better information would be an opt-in system in which households have the power to move selected categories of food and beverage on or off a "banned" list.[282] Perhaps modifying the list could not be done in-store, which would help to reduce the likelihood that the family would buy the items it does not want to be tempted by. Manuel Utset and I have also proposed a kindred idea in the consumer credit context, in which recipients of government rebates are defaulted into saving a portion, but have the power to access the funds in emergencies and to change the default savings level.[283]

---

276 Galle, *supra* note 26, at 861–63.

277 Fleischer, *supra* note , at 1686, 1697–1701; Haavio & Kotakorpi, *supra* note 108, at 576.

278 O'Donoghue & Rabin, *supra* note 40, at 1831 (claiming that tax distortions on rational actors are "second order").

279 Pratt, *supra* note 74, at 131–32. See Ted O'Donoghue & Matthew Rabin, *Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes*, 93 AM. ECON. REV. (Papers & Proceedings) 186, 190 (2003), for an early version of a voluntary sin tax proposal.

280 IAN AYRES, CARROTS AND STICKS: UNLOCKING THE POWER OF INCENTIVES TO GET THINGS DONE Ch. 8 (2010).

281 Anemona Hartocollis, *New York Asks to Bar Use of Food Stamps to Buy Sodas*, NEW YORK TIMES, Oct. 6, 2010, at A1; Steve Mistler, *Maine DHHS Renews Push for Ban on Buying Soda and Candy with Food Stamps*, PORTLAND PRESS HERALD, Nov. 23, 2015.

282 *Cf.* Janet Schwartz et al., *Healthier by Precommitment*, 25 PSYCHOLOGICAL SCI. 538, 538–46 (2013) (reporting experiment in which consumers could trigger loss of an existing subsidy if they failed to improve their healthy shopping).

283 Galle & Utset, *supra* note 89, at 84–87.

What about the inattentive? Two Finnish economists, Haavio and Kotakorpi, argue that taxes cannot be made to vary with an individual's propensity to make mistakes, both because that is unobservable (they say) and because mistake-free buyers would purchase tax-free and then resell at a discount to the error-prone.[284] Transaction costs would often remove the second concern. Even assuming that both concerns were in full force, regulators could design around them by taxing observable behaviors that are correlated with internalities, not the internality itself.[285]

An example here could be a tax that scales up with portion size. The bigger the bottle of soda, or the more cigarettes per package, the greater the tax rate per ounce or per cigarette. Analyses of inattentive behaviors show that portion size is a major driver of consumption—or, to put it another way, the minor nuisance of having to acquire another serving slows consumption considerably among those who tend to be overconsumers.[286] Of course, retailers would respond to the scaled-up tax rate by selling smaller portions, but that is exactly the goal of the policy; by reducing portion size, we also reduce internalities. For rational actors, the bother of buying a second cup is trivial. We therefore have an instrument that, despite being a tax, is nonetheless asymmetric: rational actors will not pay it.

E. Distribution

Distributive considerations can also be important to choosing an instrument, and typically will favor transferless instruments.[287] Some are relatively straightforward. Critics of sin taxes often complain that they fall more heavily on the poor.[288] Gruber and Koszegi respond that, by assumption, an internality-correcting tax on net improves well-being for each individual, so that if the tax falls more on the poor this simply means that it provides even greater benefits for

284 Haavio & Kotakorpi, *supra* note 108, at 587.

285 *Cf.* Mullainathan et al., *supra* note 168, at 17.18 & n.17 (suggesting that effective prices can be varied with degree of internality by concentrating enforcement based on observable correlates of internality).

286 *See* Chandon, *supra* note 83, at 16 (connecting portion size and inattention to overeating); Andrew B. Geier et al., *Unit Bias: A New Heuristic That Helps Explain the Effect of Portion Size on Food Intake*, 17 PSYCHOL. SCI. 521, 5224 (2006) (suggesting that "immediate presence" of tempting goods drive the effects of portion size on consumption).

287 While it is possible, and perhaps even preferable, for policy makers to analyze each rule as though it were enacted simultaneously with a perfectly offsetting adjustment to the income tax, Kaplow, *supra* note 128, at 488, 494, in practice this step may not always be feasible, *id.* at 499. Among other reasons, it may be difficult to observe the distributive effects of some policies, such as the self-targeting instruments I have described here. In that situation, Kaplow suggests analyzing the policy as though it were transferless, but immediately followed by an income tax adjustment with the appropriate redistributive characteristics. *Id.* at 498.

288 Lucas, *supra* note 237, at 738–39; Pratt, *supra* note 74, at 121–24.

the poor than its burdens.[289] But this latter analysis overlooks opportunity costs. Assuming that the government has available some other, transferless, instrument that potentially provides individuals with the same benefits as the tax, the tax will necessarily be more regressive than the tax-free alternative, unless it is offset by refunds or other support.[290]

Free or subsidized support for internality-sufferers can also be an efficient way of redistributing to the poor. In general, the distributive features of externality correction policies do not offer reasons to enact them, as the same redistributive benefits could be obtained through a simpler, less-distortive income tax.[291] Recall, however, that in-kind transfers can be more efficient than cash transfers via the tax system in the special case where the in-kind benefit disproportionately benefits individuals with low earning potential.[292] Then the government can support indigent households without creating incentives for individuals with high earning potential to stay home. Internality correction, we saw earlier, can have this feature. Since internality-correcting taxes or other "sticks" would undercut the benefit of efficient redistribution by diminishing the size of the net benefit the poor receive, these options are less appealing than others in some cases.

F. Summary

Let's try to pull together some of these threads. First, in many cases carrots are likely to remain the least favored policy. Carrots for internalities lack the ruinous moral hazard problems they pose for externalities, but they remain socially costly because they must be paid for through some kind of tax revenue. As a form of price instrument, carrots for externalities can generate better information than transferless policies, but we have seen that these information benefits will often be minor or redundant in the internality context. Thus, the extra cost of carrots is only likely to be worth paying if their income or redistributive benefits are large enough to justify the markup. Both of these factors can arise sometimes but are not at all universal.

Next, nudges and other "surprising" transferless instruments, such as

289 Jonathan Gruber & Botond Koszegi, *Tax Incidence When Individuals are Time-Inconsistent: The Case of Cigarette Excise Taxes*, 88 J. PUB. ECON. 1959, 1960 (2004); *see also* Avigail Kifer, The Incidence of a Soda Tax, in Pennies and Pounds, unpublished manuscript. Nov. 25, 2014, at 5 (making same claim about soda taxes).

290 *See* Emmanuel Farhi & Xavier Gabaix, Optimal Taxation with Behavioral Agents, NBER Working Paper No. 21524, Aug. 26, 2015, at 26; *cf.* Haavio & Kotakorpi, *supra* note 108, at 576 (noting that sin taxes transfer wealth from the irrational to the rational). On the refund possibility, see Hines, *supra* note 127, at 65.

291 KAPLOW, *supra* note 189, at 32.

292 *See supra* text accompanying notes 187–190.

defaults and choice architecture, look like potentially strong options for overcoming willpower or confirmation-bias failures, especially when willpower failures are correlated with low earning capacity. Again, there will often be little information lost by abandoning transfer instruments for nudges. The nudge is less regressive than a tax would be and may even serve as an efficient mode of supporting the poor. Depending on the labor-supply effects of a given nudge, it may be superior on revenue terms to a tax alternative. And even the income effects of the nudge are preferable in the case of tempting inferior goods.

The case is less clear cut for the inattentive. For these individuals, a tax potentially could be a very efficient revenue-raiser, although there would be serious distributive fairness concerns if it ended up falling mostly on lower-income households. The income effects of the tax are also more likely to be useful in most cases, since most consumption goods are ordinary.

It's worth emphasizing, too, that government need not rely on a single instrument for any given social problem. As the externality literature recognizes, there are some good arguments for relying on multiple instruments, including my recent argument that it allows for a way of imposing multiple *ex ante* price points.[293] In addition, we have seen that different cognitive failings may produce the same mistaken behavior. If some over-eaters are inattentive while others suffer willpower failures, it could make sense to use a different instrument to help each of the sub-populations.[294] In some cases, though, instruments may be in tension with one another. A nudge aimed at Box 4 consumers might reduce demand among Box 1 individuals, diminishing tax revenues earned from taxing the inattentive.[295] If these revenues were the main reason for adopting the tax, that combination might not make sense.

Another argument for multiple instruments arises if actors vary in their sensitivity to dollar instruments. In that case, I have shown, the optimal approach is to increase the price of the instrument, but not by so much as to fully correct the behavior of the most insensitive. [296] The intuition for this result is similar to the argument for using several *ex ante* prices: because the social cost of errors

---

293 Galle, *supra* note 122, at 1730–33; David M. Driesen, *Emissions Trading Versus Pollution Taxes: Playing "Nice" with Other Instruments*, 48 ENVTL. L. REV. __, manuscript at 32 (forthcoming 2018); Michael P. Vandenbergh et al., *Regulation in the Behavioral Era*, 95 MINN. L. REV. 715, 719 (2011); *see* Emanuel Saez, *The Optimal Tax Treatment of Tax Expenditures*, 88 J. PUB. ECON. 2657, 2659–60 (2004) (explaining use of tax subsidies and adjustments to level of direct government provision as complementary tools); *see generally* Vidar Christiansen & Stephen Smith, *Externality-Correcting Taxes and Regulation*, 114 SCANDINAVIAN J. ECON. 358 (2012).

294 Allcott et al., *supra* note 138, at 77–78. Similarly, if there are both consumption externalities and internalities, different instruments may be necessary to target both effectively. Madrian, *supra* note 157, at § 2; Goldin & Lawson, *supra* note 14, at 441.

295 *Cf.* Kaplow, *supra* note 238, at 22 (noting that mandatory savings and savings subsidies produce inefficient results when combined).

296 Galle, *supra* note 177, at 77–81.

grows exponentially, it is better to make a few small mistakes than one big one.[297] Even better, though, would be to eliminate one of the small mistakes. A second, behavioral instrument aimed at the group that is most insensitive to a traditional tax or subsidy would improve over the dollar instrument alone.[298]

Enforcement costs might offer a third reason for multiple instruments. Suppose, for example, that our instrument is a carrot, which must be paid for with tax revenues. The benefits of offering the carrot are counter-balanced by the economic distortions caused by raising taxes to pay for the carrot. A standard result in the externality literature is that, in situations where government faces tradeoffs of this kind, it may not be optimal to set subsidy levels at the full internality-correcting level (*tau*, or "τ" in Figure One).[299] At prices very close to *tau*, there is little marginal benefit from further policy change: government has already helped those who are the worst off, and the gains from helping the next worst-off grow steadily smaller. At the same time, the costs of implementing that policy are growing—in our example, the tax burden of paying for more and more carrots. Balancing these two factors against each other, it often will not be cost effective to help everyone.

Multiple instruments can help to solve the costly tradeoff problem. As with any standard tax, the economic distortion of each instrument should grow exponentially with its effective rate.[300] This implies that, by using two small "taxes" or instruments instead of one big one, the government will often face less of a tradeoff when it implements its policy.[301] It can, cost-effectively, get closer to the full internality-correcting price.

Obviously, there remains an enormous amount of uncertainty with all of these prescriptions. My goal is not to be able to provide definitive answers to, say, how to regulate vaccinations. The point instead is to identify for further study the factors that we need to know in order to make the best policy.

VI. APPLICATION: TOBACCO REGULATION

To repeat, at this point there remain important unknowns in evaluating the best policy for any given internality. To make my analysis concrete, though, I offer some preliminary thoughts, given the available evidence, on how my

297 *Id.* at 78–79.

298 *See* Farhi & Gabaix, *supra* note 290, at 25, 28 (modeling combination of tax and nudge when some actors are insensitive to taxes).

299 Bovenberg & Goulder, *supra* note 228, at 1486.

300 *See id.* (explaining equivalence of taxes and other costly regulations).

301 This argument assumes the distortive behavior produced by the two instruments does not overlap. For a more general discussion of this assumption and its importance, see David Gamage, *How Should Governments Promote Distributive Justice? A Framework for Analyzing the Optimal Choice of Tax Instruments*, 68 TAX L. REV. 1, 21–44 (2014).

argument would apply to a significant real-world source of internalities: smoking.

Recent U.S. efforts to follow other countries around the world in requiring that packages of cigarettes prominently display graphic images of smoking's health consequences were stymied by a panel of the U.S. Court of Appeals for the D.C. Circuit.[302] The panel ruled that the images infringed on the First-Amendment right of manufacturers to control their brand message, the administration declined to seek *certiorari*, and the government went back to the drawing board.[303] But two years later, the full D.C. Circuit, sitting *en banc* in a different dispute, ruled that the relatively searching scrutiny it had used in the earlier case was not justified in the commercial speech context.[304] Therefore it appears there is again an opportunity to revive the graphic-images rule.[305] What can the government say in defense of graphic images?

Smoking is a cognitively complex behavior, with different smokers seeming to exhibit different kinds of at least arguably irrational behavior. Evidence suggests some smokers are classic examples of Box 4 willpower failure–no surprise, since nicotine is addictive.[306] Another contributing cause for some smokers is a form of inattention bias, in which the immediate visceral temptation of habitual cues that trigger the urge to smoke bypass rational thought processes.[307] These we might place in Box 1. A small group of smokers seem to reject evidence that their personal risks of smoking are serious, even though these individuals tend to be among the heaviest smokers.[308] That, of course, accords well with our Box 3.

To meet the challenge of this very heterogeneous group of consumers, tobacco control policy likely must be equally complex. For each sub-population of smokers, a different regulatory instrument may be optimal. Further, it may be optimal to use multiple instruments even within each sub-group. A number of studies show that graphic images have helped to motivate and encourage quitting

---

302 *R.J. Reynolds Tobacco Co. v. Food & Drug Admin.*, 696 F.3d 1205, 1208 (D.C. Cir. 2012).

303 Letter from Eric Holder, Att'y Gen., to John Boehner, Speaker, U.S. House of Representatives (Mar. 15, 2013), http://www.justice.gov/sites/default/files/oip/legacy/2014/07/23/03-15-2013.pdf.

304 *Am. Meat Inst. v. U.S. Dep't of Agric.*, 760 F.3d 18, 22–23 (D.C. Cir. 2014) (en banc).

305 For more detailed discussion of the First-Amendment issues, see Rebecca Tushnet, *More Than a Feeling: Emotion and the First Amendment*, 127 HARV. L. REV. 2392, 2404–15, 2442–43 (2014).

306 Gruber & Koszegi, *supra* note 77, at 1278; Joni Hersch, *Smoking Restrictions as a Self-Control Mechanism*, 31 J. RISK & UNCERTAINTY 5, 6 (2005).

307 Bernheim & Rangel, *supra* note 55, at 44–45; George Loewenstein, *A Visceral Account of Addiction*, *in* GETTING HOOKED: RATIONALITY AND ADDICTION 235, 237–45 (Jon Elster & Ole-Jórgen Skog eds., 1999).

308 AUSTRALIA GOV'T DEP'T OF HEALTH AND AGING, *supra* note 17, at 57.

and its follow-through, and to discourage adolescent smoking.[309] Which instrument or instruments, then, should we choose?

Let's take Box 4 willpower-failure sufferers to start. For them, my earlier analysis implies that the graphic images policy is the best choice, given one key factual assumption. If it is the case that the images do not have significant impacts on labor/leisure decisions or related distortions typically associated with a sales tax, then the images are on net preferable strictly on revenue terms. That is, although they bring in less money than a tax would, they also produce less deadweight loss, so that on net the government comes out ahead with the images. It's worth emphasizing we do not currently have evidence on that question. It seems likely, though, that those who have a long-term preference for quitting would view the images as on net improving, not diminishing, the value of their take-home pay. The images also have helpful distributive effects. Since smokers tend to be poorer, and the policy on net helps smokers, it is actually progressive overall. It is particularly progressive compared to a tax alternative.[310]

There is, though, a fair argument in favor of subsidies or other "carrots" for smoking cessation, and likely the optimal policy is a mix of carrots and graphic images. The correlation between low willpower, smoking, and low earning potential makes smoking-cessation subsidies a highly efficient tool for redistribution. But carrots also have downsides, including the tax cost of paying for them, which is in turn compounded by the carrots' potential income effects. Both of these problems likely grow exponentially with the size of the subsidy.[311] Further, the redistributive rationale would be turned on its head to the extent that subsidies are claimed by smokers of higher income.

As I mentioned earlier, tradeoffs of this kind offer a strong case for multiple instruments.[312] In this instance, it would probably not be optimal to pay the price of a carrot for every last smoker, so the optimal carrot is less than the full internality-correcting price. Graphic images could be used to make up the

309 For surveys, *see id.*; Hammond, *supra* note 17, at 329–34. The D.C. Circuit, in finding that the U.S. FDA had failed to show evidence that vivid images would reduce smoking, cherry-picked a single sentence from Hammond's review in which he expressed reservations about the empirical methodology of one paper. *R.J. Reynolds Tobacco Co. v. Food & Drug Admin.*, 696 F.3d 1205, 1220 (D.C. Cir. 2012). Overall, he reports, "[T]he research literature unequivocally demonstrates the impact of comprehensive health warnings." Hammond, *supra* note 17, at 334.

310 Lucas argues that "psychic" taxes are regressive, Gary Lucas, *Paternalism and Psychic Taxes: The Government's Use of Negative Emotions to Save Us from Ourselves*, 22 S. CAL. INTERDISCIPLINARY L.J. 227, 297–98 (2013), but fails to distinguish between psychic and real taxes. A tax that is collected in dollars, because of the diminishing marginal utility of money, has much greater impact on poor households. *See* Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1121 (1972).

311 *See supra* text accompanying notes 218–219.

312 *See supra* text accompanying notes 299–301.

resulting gap between *tau* and the subsidy price.

Let's move on now to Box 1. The optimal policy for the cue-triggered smokers in Box 1 is probably a mirror image of the ideal Box 4 strategy: instead of carrots and images, the best choice is taxes and images. Our analysis earlier suggests that taxes on inattentive smokers are probably a highly efficient revenue source, though that would be less true to the extent that they trigger labor-supply or related distortions. The revenue benefit, and its accompanying helpful income effects, have to be traded off against the sharply regressive impact of a substantial tobacco tax.[313] If the resulting optimal tax were less than *tau*,[314] the graphic images could be added to the policy mix in order to obtain full deterrence of the internality. Images are also helpful to the extent that some inattentive smokers are inattentive to the cigarette tax, but are still sensitive to graphic images (for example, because the images disrupt the tempting cues that otherwise trigger smoking).[315]

Combining our strategies for Box 1 and Box 4 is tricky. While it is not necessarily absurd to try to enact both a tax and a subsidy at the same time, it may be more sensible simply to rely on graphic images and other "transferless" policies instead. If the tax and subsidy exactly offset—say, if all the new cigarette tax revenues are used to support smoking cessation—then on net what we have done is enact a mandate to purchase smoking cessation. [316] Smokers would also retain a marginal incentive to cut back, since they could still reduce their personal contribution to the cessation program with each puff they snuff.[317] A mandate to buy cessation could be defensible, as perhaps for some smokers the cost of cigarettes is crowding out cessation spending.[318] But it would require large administrative costs to administer both the tax and the subsidy.[319] We might

313 *See* Farhi & Gabaix, *supra* note 290, at 26 (modeling efficiency tradeoff between distortions and redistribution for low-salience taxes). While one could imagine strategies for avoiding the regressive impact of the tax—for example, implementing the "tax" as a smoking license fee, and granting free licenses to poorer households—these approaches would mostly eliminate the efficiency advantage of the tax by turning it into a *de facto* tax on income.

314 Note that, because of the revenue benefits, the optimal Box 1 tax could conceivably exceed *tau* if there were no redistributive concerns.

315 *See* Bernheim & Rangel, *supra* note 55, at 44–45 (suggesting that cue-triggered smokers should be unresponsive to future tax cost of smoking).

316 *See* KAPLOW, *supra* note 189, at 13–34 (analyzing externality-correcting policy enacted together with exacting offsetting tax adjustment).

317 *Id.*

318 *Cf.* Susan H. Busch et al., *Burning a Hole in the Budget: Tobacco Spending and its Crowd-Out of Other Goods*, 3 APPLIED HEALTH ECON. & HEALTH POL'Y 263, 266–71 (2004) (reporting that smoking expenditures displace rent and food).

319 For an overview of enforcement issues in tobacco excise collection, see Department of the Treasury Report to Congress on Federal Tobacco Receipts Lost Due to Illicit Trade and Recommendations for Increased Enforcement, Feb. 4, 2010, http://www.ttb.gov/pdf/tobacco-receipts.pdf.

justify the costs by arguing that cessation and taxing policies reached different individuals than the images alone could. If not, though, it seems better to simply rely on the images, which provide marginal incentives to quit with something like one-third the administrative apparatus.

On the other hand, if the policies do not perfectly offset, we are left with a stub version of either one, potentially with helpful economic results. For instance, if the rich pay cigarette taxes but do not take up free government cessation programs (perhaps because they can afford better programs on their own), then we have achieved a small, but quite efficient, redistributive tax.

The presence of Box 3 smokers might push us towards the three-instrument option. Although evidence is still very preliminary, some research does suggest that graphic images can prompt stronger bolstering and denial responses among the group of heavy smokers, diminishing the efficacy of the image.[320] It might be worth paying the extra administrative costs of the tax/subsidy/graphic image combination to reach these individuals, although admittedly they are a relatively small share of smokers.

If there is such a thing as a "rational smoker,"[321] their presence also would suggest that multiple instruments could be optimal. Externalities or other market failures aside, the government should not try to change the behavior of rational actors.[322] Where the government faces a mix of rational and irrational individuals, we have seen, the *tau* or price it presents should usually be a weighted average of the zero price that should face the fully rational and whatever other prices should face various internality sufferers—essentially, balancing the harm to the rational against the help for the irrational. If government has some evidence that allows it to make educated guesses about who is in which pool, it can use multiple instruments to strike a better balance for each group.[323]

Extensive heterogeneity in smokers' need for government help, then, would offer another rationale for using multiple instruments to combat smoking. Gary Lucas argues that government should limit itself to offering opt-in commitment devices, such as voluntary smoking licenses, in order to avoid burdening possible rational smokers.[324] The trouble with this suggestion is that it leaves those who do not seek out commitment devices, our Box 3 and naïve Box 1 smokers, without any help at all. Better would be to implement a general policy for all smokers, with a relatively lower tau, that accounts for the possibility that some in

320 AUSTRALIA GOV'T DEP'T OF HEALTH AND AGING, *supra* note 17, at 41.

321 Becker & Murphy, *supra* note 108, at 694–95.

322 GRUBER, *supra* note 24, at 3.

323 Galle, *supra* note 122, at 1730–34.

324 Lucas, *supra* note 237, at 743–44. *See* Lee Anne Fennell, *Revealing Options*, 118 HARV. L. REV. 1399, 1483–85 (2005), for a more detailed proposal.

that group may have no or smaller internalities. On top of that, individuals could opt into more costly policies, reflecting the fact that those who have opted in have, on average, more need for help. In other words, the voluntary license could, and likely should, exist side-by-side with graphic images.

In short, there are good reasons for the U.S. Food & Drug Administration to go forward with its stalled regulatory project on graphic images for tobacco control. And, if confronted by a skeptical court wondering why taxes and subsidies are not preferable alternatives to infringing on the commercial speech of cigarette makers, the FDA now has some good answers.

VII. CONCLUSION

The analogy between externality regulation and internalities is powerful. The lessons of the externality literature not only help us to see why we should regulate internalities—answering, for example, the heterogeneity and information constraint objections to paternalistic regulation—but also how. But, as I have tried to explore here, internalities are also different. They present unique informational challenges we are still learning to overcome. And some standard verities of externality control, such as the clear advantages of price instruments, and the clear inefficiency of carrots, are not at all obvious when translated to internalities.

I do not mean to claim that the answers I offer here are the best or the final word. In general, my goal instead has been to sketch policy possibilities and reveal places where answers depend on unknown facts. My hope is that this work helps establish an empirical research agenda for myself and others, and to stimulate discussion about what we think we do know.

At a minimum, though, I hope that I have shown it is more than possible to make an efficiency case for nudges and other kinds of cognitively-informed regulation, particularly in the internality context. While refinements and counter-arguments certainly are likely to come, the economic case for nudging is too good to dismiss with a wave and the cry, "paternalism!"

Righting Research Wrongs: An Empirical Study of How U.S. Institutions Resolve Grievances Involving Human Subjects

Kristen Underhill*

Abstract:

Tens of millions of people enroll in research studies in the United States every year, making human subjects research a multi-billion-dollar industry in the U.S. alone. Research carries risks: although many harms are inevitable, some also arise from errors or mistreatment by researchers, and the history of research ethics is in many ways a history of scandal. Despite regulatory efforts to remedy these abuses, injured subjects nonetheless have little recourse to U.S. courts. In the absence of tort remedies for research-related injuries, the only venue for resolving such disputes is through alternative dispute resolution (ADR)—or more commonly, *internal* dispute resolution (IDR) through a process offered by the research institution. The federal regulations on human subjects are silent on resolving subject grievances, and to date, little is known about how institutions handle these disputes. This Article is the first empirical study of how U.S. universities and hospitals resolve subjects' claims of physical injury, dignitary harm, non-compensation, deviations from research protocols, and maltreatment by research staff. I have conducted in-depth interviews with personnel from 30 hospitals and universities to understand how institutions respond to grievances involving research subjects. These interviews reveal highly flexible dispute resolution processes managed by institutional review boards (IRBs), the institutional authorities mandated by federal law to protect human subjects. Although many interviewees spoke intuitively of procedural justice—including elements such as voice, neutrality, and courtesy—these interviews also indicated problems with

* Associate Professor of Law, Columbia Law School. Associate Professor of Population & Family Health, Mailman School of Public Health. J.D. (Yale 2011); D.Phil. (University of Oxford, 2007). I am grateful to Jennifer Gerarda Brown, Scott Burris, Carl Coleman, Sue Fish, Celia Fisher, Stephen Latham, Robert Klitzman, Susan Sturm, and William Sage for helpful conversations in planning the research and feedback on earlier versions of this manuscript. I am grateful to Stephanie Boegeman for assistance coding transcripts, and I am grateful to the IRB chairs, directors, administrators, and managers who spoke with me as part of this research. Data collection for this paper was supported in part by the Fordham HIV Prevention Research Ethics Training Institute via a training grant sponsored by the National Institute on Drug Abuse (R25-DA031608, PI: Fisher), and in part by the National Institute of Mental Health, (K01-MH093273, PI: Underhill). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. All errors herein are my own.

neutrality, expertise, representation of participants, one-sided appeals, and access to the dispute resolution process itself. This Article takes a close look at current practices, and then suggests strategies for improvement, addressing both the federal regulations and options for institution-led reforms.

I. INTRODUCTION

Research is an enormous enterprise; more than 19 million individuals participate in research studies per year,[1] and the annual costs of research in the U.S. include an estimated $32 billion in NIH funds[2] and over $50 billion in pharmaceutical funding alone.[3] Although federal regulations, state laws, and professional organizations apply countless mandates to institutions that conduct human subjects research, the processes for resolving research participants' concerns are a curiously unregulated space. Where grievances arise, U.S. courts have recognized claims relating to physical injuries, negligent study design and oversight, and insufficiency of informed consent.[4] But courts cannot and do not respond to most research-related injuries. Litigation is procedurally unavailable for large classes of research participants, such as international subjects or subjects in intramural federal projects.[5] Moreover, many research-related disputes are not amenable to courtroom remedies. Recent work suggests that there is a high frequency of non-justiciable complaints in healthcare settings,[6] and a few such concerns in research may include study staff rudeness, offensive recruitment efforts, or post-trial access to study drugs. Prior findings suggest widespread confusion among subjects about study protocols,[7] and this confusion may engender other subject complaints. Where litigation is not feasible, or where complaints are not cognizable in courts, institutions may seek to provide alternative fora for resolving research-related disputes. These ADR practices, however, have gone entirely unnoticed by scholarship.

Responsiveness to research subjects' injuries and complaints is a legal, ethical, and practical imperative for research institutions. At institutions that receive federal funds for research, federal regulations governing research with

---

1 Adil E. Shamoo, *Adverse Events Reporting—The Tip of the Iceberg*, 8 ACCOUNTABILITY RES. 197, 197 (2001).

2 *Budget*, NAT'L INSTS. OF HEALTH, https://www.nih.gov/about-nih/what-we-do/budget [https://perma.cc/2LNT-VL3A].

3 Pharmaceutical Research and Manufacturers of America, *2015 Biopharmaceutical Research Industry Profile* 36, http://phrma-docs.phrma.org/sites/default/files/pdf/2015_phrma_profile.pdf.

4 Roger L. Jansson, *Researcher Liability for Negligence in Human Subject Research: Informed Consent and Researcher Malpractice Actions*, 78 WASH. L. REV. 229 (2003).

5 Elizabeth R. Pike, *Recovering from Research: A No-Fault Proposal to Compensate Injured Research Participants*, 38 AM. J.L. & MED. 7, 29-30 (2012).

6 *See* Orna Rabinovich-Einy, *Deconstructing Dispute Classifications: Avoiding the Shadow of the Law in Dispute System Design in Healthcare*, 12 CARDOZO J. CONFLICT RESOL. 55 (2010).

7 Matthew E. Falagas et al., *Informed Consent: How Much and What Do Patients Understand?*, 198 AM. J. SURGERY 420 (2009); James Flory & Ezekiel Emanuel, *Interventions to Improve Research Participants' Understanding in Informed Consent for Research: A Systematic Review*, 292 JAMA 1593 (2004); Adam Nishimura et al., *Improving Understanding in the Research Informed Consent Process: A Systematic Review of 54 Interventions Tested in Randomized Control Trials*, 14 BMC MED. ETHICS 28 (2013).

human subjects (the "Common Rule") delegate oversight over research protocols to institutional review boards (IRBs). IRBs are tasked with *a priori* review and approval of research protocols, after determining an appropriate balance of risks and benefits, equitable selection of subjects, and reviewing procedures for securing informed consent from participants or their legally authorized representatives.[8] In approving and monitoring protocols, U.S. IRBs often take as their guiding principles those set forth in the Belmont Report, a 1979 set of guidelines issued by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Although the Common Rule does not specify the procedures that institutions must use when grievances arise during research, federal regulations do require that participants receive "an explanation of whom to contact . . . in the event of a research-related injury."[9] This implies that responding to such contacts is indeed a legal imperative for research institutions, and most institutions house that responsibility within the IRB.

As an ethical matter, the duty to respond to participants' concerns over the course of research follows from the Belmont Report's emphasis on respect for subjects' autonomy, justice and the equitable selection of study subjects, and minimization of research burdens (beneficence and non-maleficence).[10] Because unforeseen problems may arise during research studies, each of these ethical goals requires that when participants allege injuries or grievances, institutions responsible for conducting research must remain responsive to these ongoing problems. Several scholars have noted that "researcher ethnocentrism" can limit researchers' ability to identify ethical problems in their own protocols, and researchers sitting on IRBs may be no different;[11] providing a feedback loop for subject complaints is an essential means of augmenting IRB review and oversight. As a practical matter, providing a forum for the resolution of research-related complaints may avert litigation, identify unforeseen problems in research protocols, promote stable relationships between research institutions and communities who may participate in research, and encourage participation among subjects who may be concerned about accountability in the event of injury.

Prior literature suggests that research institutions do, in fact, maintain internal processes for the resolution of research-related disputes, and IRBs provide these procedures as part of their research oversight role. But almost nothing is known to date about how these processes work. Scholarship on internal dispute resolution (IDR) systems—dispute resolution procedures maintained internally by corporations or other institutions—reflects concerns about procedural fairness.

---

8 45 C.F.R. § 46.111 (2018).

9 45 C.F.R. § 46.116(b)(7) (2018)

10 NAT'L COMM'N FOR PROTECTION OF HUMAN SUBJECTS OF BIOMEDICAL AND BEHAVIORAL RESEARCH, DEP'T OF HEALTH, EDUCATION & WELFARE, THE BELMONT REPORT (1979).

11 REBECCA DRESSER, SILENT PARTNERS: HUMAN SUBJECTS AND RESEARCH ETHICS 2 (2017).

When one party to a dispute has structured the process by which that dispute is resolved, there are many opportunities to build institutional advantage into these procedures.[12] The need for procedural fairness is keen when parties waive claims and other venues, such as an agreement not to sue, or when other venues are unavailable from the outset (e.g., because litigation is unavailable, or because the complaint is not legally cognizable). IRBs themselves are in a curious role. They have a federal mandate to protect subject well-being, independent of the institution, and the institution may not authorize research that lacks IRB approval. IRBs do not do research themselves, and their practices and decisions are rarely the subject of subject complaints. They are thus infrequent "parties" to the disputes. But IRBs are nevertheless institutional bodies and composed largely of institutional employees and staff, and they are not blind to institutional liabilities. This Article will therefore consider IRB-managed processes as "internal" to the institution, despite IRBs' independent grant of authority to approve and oversee human subjects research.

This Article proceeds on the premise that providing *procedurally* just grievance procedures in human subjects research is an entailment of the ethical duty to provide resolution to research-related complaints and injuries. Importantly, IRBs enact and implement these systems amid long-standing power imbalances between researchers, research institutions, and participant communities. The history of research abuses worldwide is long, and biomedical research in the U.S. has provided some of the most acute examples of studies that violated subjects' rights, autonomy, dignity, and humanity.[13] The current regulatory system is intended to curb these abuses, bolstered by ethical guidance such as the Belmont Report, the Declaration of Helsinki, and the Nuremberg Code. But despite these regulatory frameworks, power disparities between participants and research institutions persist. This is in part a result of epidemiology. The burden of ill health, and the risk of ill health, is unevenly distributed along lines of socioeconomic status, race, ethnicity, education, disability, and other axes of social marginalization.[14] Research protocols for the study of disease prevention, etiology, progression, and treatment, are therefore likely to recruit and enroll participants

12 *See, e.g.*, Lauren B. Edelman, Christopher Uggen & Howard S. Erlanger, *The Endogeneity of Legal Regulation: Grievance Procedures as Rational Myth*, 105 AM. J. SOCIOLOGY 406 (1999); Mark Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC'Y REV. 95 (1974).

13 *See, e.g.*, HARRIET WASHINGTON, MEDICAL APARTHEID (2014).

14 *See* Paula Braverman & Laura Gottlieb, *The Social Determinants of Health: It's Time to Consider the Causes of the Causes*, 129 PUB. HEALTH REPS. 19 (2014); Sandro Galea et al., *Estimated Deaths Attributable to Social Factors in the United States*, 101 AM. J. PUB. HEALTH 1456 (2011); Steven H. Woolf & Paula Braverman, *Where Health Disparities Begin: The Role of Social and Economic Determinants – and Why Current Policies May Make Matters Worse*, 30 HEALTH AFF. 1852 (2011).

with relatively less socioeconomic power—perhaps due to convenience and cost, but also as a function of the distribution of disease, as well as the separately impoverishing effects of disease. Biomedical research with healthy, compensated volunteers may also draw poorer subjects willing to trade off their time, convenience, and (sometimes) safety for pay. Multiple studies have shown how participants may approach research as a form of employment,[15] but compensation for research participation is held down to avoid problems of unduly influencing poor individuals to take research risks.[16] Some have also argued that current practices of payment for research participation exploit an "underclass" of healthy volunteers compensated to test experimental medications in Phase I trials—the first (and riskiest) human tests of new drugs.[17]

Given these dynamics, when a dispute arises due to perceived injury or misconduct experienced by research participants, research institutions often hold a comparative advantage in sophistication, access to human and financial resources, and access to the legal system—compared to both participants and investigators. Institutions are also repeat players, compared to participants who may only take part in one or a few studies, and they may experience a comparative advantage due to expertise or relationships strengthened by multiple experiences. These comparative disadvantages for research participants present intertwining ethical and procedural questions when designing a dispute resolution system.

Researchers' interests are also at stake. When resolving disputes between participants and investigators, IRBs also have the task of balancing investigators' interests, which may at times diverge from the interests of the institution. For example, complaints alleging researcher misconduct, protocol deviations, or harassment may expose institutions to liability, but the stakes are high for the investigators themselves, who could face termination of their research protocol, their entire research program, or their employment. These situations can be precarious for subjects, investigators, and research institutions alike, and IRBs faced with the management (or even merely the initial intake) of these disputes must navigate these conflicts. Although researchers do not have the historic structural disadvantage of research participants—they are well-educated and at times legally sophisticated parties—researchers' experience of fair process is essential for the long-term function of these dispute resolution programs.

The goal of this Article is to provide the first description of IDR processes

---

15 *See, e.g.*, ROBERTO ABADIE, THE PROFESSIONAL GUINEA PIG: BIG PHARMA AND THE RISKY WORLD OF HUMAN SUBJECTS (2010); Carl Elliott & Roberto Abadie, *Exploiting a Research Underclass in Phase 1 Clinical Trials*, 1 N. ENG. J. MED. 2316 (2008).

16 *See* Emily A. Largent and Holly Fernandez Lynch, *Paying Research Participants: Regulatory Uncertainty, Conceptual Confusion, and a Path Forward*, 17 YALE J. HEALTH POL'Y L. & ETHICS 61 (2017); William M. Sage, *Paying Research Subjects: The U.S. Example*, *in* ESSAIS CLINIQUES, QUELLS RISQUES? 137 (A. Laude & D. Tabuteau, eds., 2007).

17 ABADIE, *supra* note 15; Elliott & Abadie, *supra* note 15.

used by research institutions to address injuries and other grievances brought by participants in human subjects research. My in-depth interviews with informants at federally funded U.S. hospitals and universities have revealed that institutions maintain permanent, highly flexible IDR processes in which the IRB manages initial complaint intake, complaint investigation, involvement of institutional and sometimes external stakeholders, identification of potential remedies, decisions that are binding for research protocols, and enforcement of those decisions. These processes accommodate not only physical injuries, but also non-justiciable claims and concerns brought by people who are not (or not yet) enrolled in research protocols. The highly flexible and sometimes unwritten nature of these processes allows IRBs substantial discretion in the dispute resolution process, and IRBs use this discretion to maximize participation and voice for subjects, investigators, and other community stakeholders. Informants often described the goals of their IDR systems with reference to Belmont Report principles, including respect for autonomy and justice. IRB informants also noted their federal mandate to protect research subjects, and discussed research subjects with attention to potential vulnerability or disparities in resources and sophistication. This case study provides a useful demonstration more generally of how procedural flexibility in ADR can serve participation and legitimacy interests for complainants.

Despite the wide breadth of these IDR systems, however, this study identified recurring shortcomings of IDR processes for research-related disputes. This Article will consider three shortcomings in particular. First, as a procedural matter, the design of these systems uniformly omitted consultation with participants, or with non-institutional personnel who could represent participant interests. IRBs typically began with informal office practices for handling subject complaints, and then codified these practices into more formal systems when pursuing institutional accreditation under the Association for the Accreditation of Human Research Protection Programs (AAHRPP), which requires a written policy on complaint resolution. Design features mitigate the problem of non-consultation: for example, built-in procedural flexibility allowed individual subjects some control over the process at the time of their complaint, IRB personnel who designed the systems might be said to represent participant interests already, and some IRBs involved trusted local authorities at the moment when disputes arose. But the lack of participant consultation at the time of system design was a missed opportunity to establish systems that would be accessible and trusted by participants.

Second, informants consistently believed that uptake by subjects was low, compared to hypothesized rates of injuries and other complaints. There are numerous explanations for a lower rate of uptake, including a low frequency of grievances, low salience or importance of grievances to research subjects, and high effectiveness of initial researcher responses (i.e., before participants decide to contact the IRB). But a low rate of uptake may also indicate deficiencies in the

process. Some informants suggested that participants may be suspicious of IRB-maintained systems as non-neutral processes, while many suggested that participants are unaware that the institution is willing to remedy research-related complaints. Based on my study of the available processes, as well as typical disclosures and informed consent forms, another possibility is that procedural flexibility *itself* can complicate efforts to make such processes predictable, and to make procedural information available in advance. The flexibility for IRBs to determine procedures on a case-by-case basis may undermine the predictability and legitimacy of the process for *prospective* claimants (those considering complaining), even though the IRB may seek to use that flexibility in ways that benefit *actual* claimants who are using the process.

Third, these interviews also indicated significant ambiguity regarding the capacity of IRBs to undertake dispute resolution, with respect to both neutrality and skills. Although these bodies must protect research participants, research-related disputes systems ask the same personnel to act neutrally toward investigators and the institution, which may be concerned about legal exposure, public image, and sustainability of relationships with participant communities. IRB personnel are also colleagues with ongoing relationships with investigators, and informants acknowledged that the stakes of some complaints for investigators are high. Sensitivity to investigator interests may account for the common practice of allowing investigators to appeal IRB decisions, while participants are not generally given notice of an appeal opportunity. In many ways, the use of IRBs to resolve research-related disputes is efficient: it takes advantage of existing scientific expertise; the federal regulations already give these bodies enforceable control over research protocols, which is often needed for durable remedies; and IRBs' central mandate to safeguard participant well-being may provide a much-needed thumb on the scale in favor of participant interests. But some informants in this study noted difficulties in maintaining impartiality in the face of institutional pressure and investigator pushback, and IRB personnel often noted their lack of training in dispute resolution, mediation, or investigations. Managing research-related disputes can also tax IRBs' human resources, especially given the range of potential procedures that may be necessary to fully address a complex dispute.

Based on these findings, this Article offers several recommendations to improve the design of IDR systems for resolving research-related complaints. Because participants are party to all such disputes, and particularly in light of the power disparities between research institutions and participants, institutions should involve participants themselves in the initial design or improvement of a dispute resolution system. Baseline data on the frequency of participant grievances is largely unavailable, particularly for complaints alleging intangible harm, and it is difficult to be certain that the low uptake of IDR systems is problematic. But in light of preliminary evidence that systems are underutilized, I suggest a greater

emphasis on dispute resolution systems in the informed consent process, perhaps including procedural information and requiring a verbal discussion in addition to written informed consent, where practicable. Finally, although it may not be necessary to take these procedures out of the IRB, I suggest that institutions may consider providing IRB personnel with training in dispute resolution, conflict management, or mediation, as well as additional personnel for highly complex complaints. Furthermore, it may improve neutrality to provide for independent external review of IRBs' dispute decisions, which may be invoked by the participant, investigator, and IRB itself. It may be unwise to establish these as federal regulatory requirements, given the advantages of procedural flexibility in this context. But research institutions may in fact adopt these practices voluntarily, given the ethical and practical advantages of a functioning IDR program.

This Article proceeds in the following Parts. Part II will situate research-related disputes in the context of other ADR uses in healthcare settings, and then identify the sources of authority, ethical guidance, and regulatory flexibility for research institutions to design processes that address participant injuries and concerns. Part III presents the empirical study and a process-specific appraisal of institutional systems for research-related disputes. This section will note the multiple roles of the IRB throughout the IDR process, as well as IRBs' uses of procedural flexibility to serve what they perceive to be participants' interests. Part IV discusses informants' appraisal of these systems, followed by a more critical evaluation of strengths and weaknesses. Part V concludes by considering strategies for improving IDR in this context.

II. IDR FOR RESEARCH-RELATED DISPUTES

Despite a wide-ranging set of federal regulations, federal laws, state laws, and professional requirements governing research with human subjects, there is a persistent gap in formal guidance for resolving disputes that arise in human subjects research. The federal regulations that govern most research in the United States are silent on this issue, as are federal and state laws and aspirational ethical guidance governing domestic and global research. This gap in regulation corresponds to a near-total absence of knowledge about the processes by which research-related injuries and disputes are resolved.[18] Most scholarship in this area focuses on the problem of financial compensation for physical injuries that arise in the course of research.[19] Although many such injuries are unavoidable risks of

18 Kristen Underhill, *Legal and Ethical Values in the Resolution of Research-Related Disputes: How Can IRBs Respond to Participant Complaints?*, 9 J. EMP. RES. HUM. RES. ETHICS 71 (2014).

19 Carl Elliott, *Justice for Injured Research Subjects*, 367 N. ENG. J. MED. 6 (2012); Michelle M. Mello, David M. Studdert & Troyen A. Brennan, *The Rise of Litigation in Human Subjects Research*, 139 ANNALS INTERNAL MED. 40 (2003); E. Haavi Morreim, *Clinical Trials Litigation: Practical Realities as Seen from the Trenches*, 12 ACCOUNTABILITY RES. 47 (2005); Pike, *supra* note

clinical research, some have provided a valid basis for litigation, including justiciable claims against institutions and individuals who conduct research, IRBs and institutional officials who oversee research, research sponsors, and manufacturers of products tested in clinical research protocols.[20] Institutions may seek to settle such claims quickly,[21] but the IDR processes that may facilitate settlement are entirely unknown. Moreover, even when physical injuries are alleged, litigation is unavailable for several categories of claimant and injury, making alternative dispute resolution processes the only option for dispute resolution.[22]

The scholarly focus on physical injuries has also obscured a much wider universe of potential grievances by research participants, including claims with more precarious footing in U.S. courts. These may include claims of dignitary or intangible harm, participant abandonment, inadequate informed consent, negligent protocol design, post-research access to drugs or devices, access to incidental research findings, or concerns about compensation, or complaints about the lack of privacy or confidentiality.[23] Where such claims have been unsuccessful in litigation, ADR processes are once again the only available forum for dispute resolution. The remainder of this Section will consider other uses of ADR in healthcare settings, available guidance for IRBs responding to research-related complaints, and predictable categories of disputes.

A. Uses of IDR in Healthcare

IDR programs are on the rise in healthcare settings, largely inspired by changes in the resolution of medical malpractice claims. These systems include communication-and-resolution programs for medical errors,[24] disclosure and apology programs for the proactive disclosure of errors,[25] and the use of ombudsmen or other internal complaint-handling processes for both justiciable and

---

5.

20 David B. Resnick, *Liability for Institutional Review Boards: From Regulation to Litigation*, 25 J. LEGAL MED. 131 (2004); Mello et al., *supra* note 19.

21 Mello et al., *supra* note 19, at 43.

22 Here, I consider ADR to include institutional processes for compensating injuries through insurance, if the institution is one of the few that insures against research-related injuries. *See* Pike, *supra* note 5.

23 *See* infra. Richard S. Saver, *Medical Research and Intangible Harm*, 74 U. CIN. L. REV. 941 (2005) [hereinafter Saver, *Medical Research*]; Richard S. Saver, *At the End of the Clinical Trial: Does Access to Investigational Technology End as Well?*, 31 W. NEW ENG. L. REV. 411 (2009) [hereinafter Saver, *At the End*]; Morreim, *supra* note 19; Underhill, *supra* note 18.

24 William M. Sage et al., *How Policy Makers Can Smooth the Way for Communication-and-Resolution Programs*, 33 HEALTH AFF. 11 (2014).

25 Maria Pearlmutter, *Physician Apologies and General Admissions of Fault: Amending the Federal Rules of Evidence*, 72 OHIO ST. L.J. 687 (2011); Joanna C. Schwartz, *A Dose of Reality for Medical Malpractice Reform*, 88 N.Y.U. L. Rev. 1224 (2013).

non-justiciable complaints in hospital settings.[26] Institutions are also experimenting with private or court-annexed medical malpractice arbitration.[27] Mandatory arbitration has been particularly controversial in the nursing home setting, and as of this writing, the Centers for Medicare and Medicaid Services has proposed a rule that would loosen requirements needed for nursing homes to impose binding arbitration agreements.[28] Licensing boards for physicians and nurses offer another forum for the resolution of complaints against individual providers, including complaints from patients and referrals from other authorities.[29] Because these are external, rather than IDR programs, however, they are less applicable to research-related disputes.)

IDR is also used outside the context of medical errors and patient complaints. Healthcare ethics committees have emerged as a method for managing disputes about courses of treatment for patients, reconciling the interests of patients, families, and caregivers.[30] Bioethics mediation processes, including particularly the approach suggested by Nancy Dubler and Carol Liebman, integrates mediation skills into clinical ethics consultation, promoting shared decision-making and consensus in clinical conflicts.[31] Outside clinical care, health insurers offer internal procedures for managing coverage disputes, with external review mandated by state law (in most states)[32] and the Affordable Care Act.[33] Some disputes that arise

---

26 Orna Rabinovich-Einy, *Escaping the Shadow of Malpractice Law*, 74 L. & Contemp. Probs. 241 (2011); Orna Rabinovich-Einy, *Beyond IDR: Resolving Hospital Disputes and Healing Ailing Organizations through ITR*, 81 ST. JOHN'S L. REV. 173 (2007); Rabinovich-Einy, *supra* note 6.

27 Matthew Parrott, *Is Compulsory Court-Annexed Medical Malpractice Arbitration Constitutional?*, 75 FORDHAM L. REV. 2685 (2007); Stephanie Smith & Janet Martinez, *Analytic Framework for Dispute Systems Design*, 14 HARV. NEGOT. L. REV. 123, 128 (2009) (reporting Kaiser case study).

28 82 Fed. Reg. 26649 (June 8, 2017) *CMS Issues Proposed Revision Requirements for Long-Term Care Facilities' Arbitration Agreements*, CTRS. FOR MEDICARE & MEDICAID SERVICES (June 5, 2017), https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2017-Fact-Sheet-items/2017-06-05.html.

29 Timothy S. Jost et al., *Consumers, Complaints, and Professional Discipline: A Look at Medical Licensure Boards*, 3 HEALTH MATRIX 309 (1993).

30 Thaddeus Mason Pope, *Multi-Institutional Healthcare Ethics Committees: The Procedurally Fair Internal Dispute Resolution Mechanism?*, 31 CAMPBELL L. REV. 257 (2009).

31 NANCY NEVELOFF DUBLER & CAROL B. LIEBMAN, BIOETHICS MEDIATION: A GUIDE TO SHAPING SHARED SOLUTIONS (2011).

32 Nan Hunter, *Managed Process, Due Care: Structures of Accountability in Health Care*, 6 YALE J. HEALTH POL'Y, L. & ETHICS 93 (2006).

33 Patient Protection and Affordable Care Act, Pub. L. 111-148 (2010); Interim Final Rules for Group Health Plans and Health Insurance Issuers Relating to Internal Claims and Appeals and External Review Processes under the Patient Protection and Affordable Care Act, 75 FR 43329, July 23, 2010 (codified at 26 C.F.R. § 54, 26 C.F.R. § 602, 29 C.F.R. § 2590, 45 § C.F.R. 147 (2017)); Group Health Plans and Health Insurance Issuers: Rules Relating to Internal Claims and Appeals and External Review Processes 76 F.R. 37207, June 22, 2011, codified at 26 C.F.R. § 54; 29 C.F.R. § 2590, and 45 C.F.R. § 147 (2017). The ACA mandates external review mechanisms for coverage determinations in group health plans and individual plans in the federal and state marketplaces.

in healthcare and health research settings are also those of large organizations more generally, including concerns about employment and discrimination, interpersonal conflicts, shared credit and workload, and organizational concerns. Susan Sturm and Howard Gadlin have discussed the National Institutes of Health's ombudsman program for handling these types of disputes, noting the interplay between individual-level and systemic analyses and solutions for organizational problems.[34]

Aggregating these IDR processes raises questions about healthcare exceptionalism:[35] whether process values or goals should be differently weighted in healthcare settings because there are distinctive interests at stake. Procedural scholars have long enumerated the underlying purposes and values of procedural due process adjudication, and claims about the values served by process have extended from litigation and administrative adjudication to ADR and IDR.[36] The design of dispute resolution procedures are now widely acknowledged to serve not only accuracy, [37] but also other values, particularly given the impossibility of perfect accuracy in any system.[38] One such value is participation by claimants, either because participation is an intrinsic good,[39] or because it is instrumental[40] in producing a psychological experience of fairness,[41] promoting dignified treatment,[42] or conferring legitimacy on decisions.[43] Other values may include system legitimacy (including "the appearance of fairness"[44]), predictability,[45]

---

34 Susan Sturm & Howard Gadlin, *Conflict Resolution and Systemic Change*, 2007 J. DISP. RESOLUTION 1, 2 (2007).

35 Hunter, *supra* note 32.

36 *See, e.g.*, Rebecca Hollander-Blumoff, *The Psychology of Procedural Justice in the Federal Courts*, 63 HASTINGS L.J. 127 (2011); Rebecca Hollander-Blumoff & Tom R. Tyler, *Procedural Justice and the Rule of Law: Fostering Legitimacy in Alternative Dispute Resolution*, 2011 J. DISP. RESOL. 1 (2011).

37 Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L.J. 455 (1986); Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181 (2004).

38 Laurens Walker, *Avoiding Surprise from Federal Civil Rule Making: The Role of Economic Analysis*, 23 J. LEGAL STUD. 569 (1994); Solum, *supra* note 37, at 185.

39 *See, e.g.*, JERRY MASHAW, DUE PROCESS IN THE ADMINISTRATIVE STATE 177 (1985).

40 *See* Matthew J.B. Lawrence, *Procedural Triage*, 84 FORDHAM L. REV. 79 (2015) (describing the importance of participation with respect to psychological, dignitary, and legitimacy theories).

41 *See* LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 275-80 (2002) (noting that "a taste for fairness" may explain individual preferences for some procedures in adjudication, but also expressing skepticism that strong preferences exist); *see also* Lawrence, *supra* note 40, at 92 (examining psychological theories that consider the inherent value of participation in dispute resolution, including satisfying a preference for fair treatment).

42 Jerry L. Mashaw, *Administrative Due Process: The Quest for a Dignitary Theory*, 61 BOS. U.L. REV. 885, 886 (1981); MASHAW, *supra* note 39.

43 Solum, *supra* note 37; *accord* Lawrence, *supra* note 40.

44 Redish & Marshall, *supra* note 37.

45 Mashaw, *supra* note 39 at 175-76, also quoted by Redish & Marshall, *supra* note 37.

equality of parties,[46] accountability of parties,[47] "revelation" and explanation of the events that led to the claim,[48] and respect for dignity and privacy.[49]

Many (although not all) grievances arising in healthcare settings present a unique combination of physical or mental vulnerability, information asymmetry, emotional weight, socioeconomic disparity, cultural difference, urgency, and visceral need, particularly conflicts involving individual patients and healthcare providers.[50] In this context, process values such as revelation, equality, accountability, participation, and dignity take on greater salience; IDR innovations such as disclosure-and-apology, communication-and-resolution, and bioethics mediation express these values clearly. Because research with human subjects presents many of the same contextual features, we may expect similar process values to have a high priority in IDR for research-related complaints.

The legitimacy of not only the IDR process, but also the larger system of healthcare services is also an important priority for inherent and instrumental reasons. Medical mistrust is a formidable barrier to accessing care[51] and promoting quality in care delivery[52] and perceived mistreatment in medical contexts can foster litigation and violence.[53] Both undermine the core goals of healthcare institutions, many of which are nonprofit corporations principally engaged in patient care. Given the goals of institutional legitimacy, such institutions may be more receptive to addressing non-justiciable disputes based on interests rather than legal rights.[54] IDR is the only process option for these types of disputes. Many IDR processes in healthcare settings were established as alternatives to public adjudication of justiciable claims, such as medical malpractice claims sounding in tort or coverage disputes sounding in contract. But IDR innovations in health law also extend to

---

46 Redish & Marshall, *supra* note 37, at 484-85; MASHAW, *supra* note 39 at 171.

47 *See* Galanter, *supra* note 12. This is particularly problematic for some forms of ADR, such as internal dispute resolution, whereby one party designs the procedural rules and provides the forum. *See* Lauren B. Edelman & Mark C. Suchman, *When the Haves Hold Court: Speculations on the Organizational Internalization of Law*, 33 L. & SOC'Y REV. 941 (1999).

48 Redish & Marshall, *supra* note 37 (quoting Frank I. Michelman, *Formal and Associational Aims in Procedural Due Process*, 18 DUE PROCESS: NOMOS 126, 127 (1977)); Hunter, *supra* note 32.

49 Mashaw's theory considers dignity the overarching underlying value served by equality, predictability, participation, and privacy. MASHAW, *supra* note 39, at 172-82.

50 Hunter, *supra* note 32.

51 Thomas A. LaVeist, Lydia A. Isaac, & Karen Patricia Williams, *Mistrust of Health Care Organizations is Associated with Underutilization of Health Services*, 44 HEALTH SERVS. REV. 2093 (2009); Kristen Underhill et al., *A Qualitative Study of Medical Mistrust, Perceived Discrimination, and Risk Behavior Disclosure to Clinicians by U.S. Male Sex Workers and Other Men Who Have Sex with Men*, 92 J. URBAN HEALTH 667 (2014).

52 Rabinovich-Einy, *supra* note 6, at 69; *see also* Mark A. Hall et al., *Trust in Physicians and Medical Institutions: What Is It, Can It Be Measured, and Does It Matter?*, 79 MILBANK Q. 613 (2001).

53 *Id.* at 68, 78.

54 *Id.*

disputes that would not support litigation in public courts. Bioethics mediation, healthcare ethics committees, internal complaint-handling mechanisms and hotlines at hospitals, and fora such as ombudsman programs in large health-related organizations all address both justiciable and non-justiciable claims. The availability of fora for these disputes promotes not only participation values, but also legitimacy of the care system more generally. These themes are all present in the context of research-related disputes, to which we now turn.

B. Authority and Guidance for IDR in Research Settings

Although the institutions that conduct human subjects research are subject to complex and overlapping federal and state laws, as well as informal ethics guidance and the requirements of professional self-governance and accreditation, the resolution of research-related disputes is an almost entirely unregulated space. This Section will describe the authority and existing guidance for research institutions addressing participant complaints.

The regulatory provisions governing research with human subjects include 45 C.F.R. § 46 (for research at institutions receiving federal funding through most agencies and departments) and 21 C.F.R. § 50 and 21 C.F.R. § 56 (for research that will be submitted as part of an application for FDA approval of a new drug or device). These regulations grant IRBs (which may be internal or external to research institutions) the authority to approve and monitor research protocols on an ongoing basis. As part of this authority, IRBs are empowered to withdraw approval, suspend, or terminate studies.[55] This authority entails stoppage or modification of a protocol in response to a complaint or injury. Although IRBs have authority over research protocols, however, the federal regulations are silent on the processes by which participant grievances should be resolved. The Common Rule refers to these processes only directly: as part of informed consent, participants must receive contact information for a party who can provide "answers to pertinent questions about the research and research subjects' rights, and . . . in the event of a research-related injury to the subject."[56] Institutions almost universally satisfy this requirement by providing participants with the contact information of the IRB, although the regulations do not specify that the IRB is the correct or only resource for questions about rights and injuries.[57] Dispute resolution receives no further attention in the recent revisions to the Common Rule.[58]

The Office of Human Research Protections (OHRP) within the Department of

55 45 C.F.R. § 46.109 (2018).
56 45 C.F.R. § 46.116(b)(7) (2018); 21 C.F.R. § 50.25(a)(7) (2018).
57 Underhill, *supra* note 18.
58 Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149 (Jan. 19, 2017).

Health and Human Services, which is tasked with enforcing the Common Rule,[59] has issued formal guidance to assist institutions in their oversight of human subjects research. These guidance documents, however, address only subject concerns that fall into the categories of "adverse events" or "unanticipated problem involving risks to subjects or others."[60] Adverse events are narrowly defined as "untoward or unfavorable medical occurrence[s] . . . temporarily associated with the subject's participation in the research," while unexpected problems are incidents that are "unexpected . . . related or possibly related to participation in the research . . . [and] suggest that the research places subjects or others at a greater risk of harm . . . than was previously known or recognized."[61] Even within these categories, the focus of OHRP guidance is on how institutions should report the events and correct the research protocol—rather than providing mechanisms for addressing the harm experienced by the individual subjects. OHRP does not direct IRBs to enact a complaint resolution policy separate from these procedures. [62]

Many states also govern human research by statute or regulation, but like the federal regulations, these are typically silent on the mechanisms by which institutions resolve disputes with individual participants. State statutes governing research in California, for example, require that participants in medical research receive "the name, address, and phone number of an impartial third party, not associated with the experiment, to whom the subject may address complaints about the experiment."[63] "Impartial third party," however, is not defined, nor is the procedure by which this third party should resolve the dispute.[64] New York state law requires that research protocols falling outside federal regulatory requirements be reviewed by a "human research review committee"[65] and that researchers secure informed consent from subjects,[66] but does not address the resolution of research-

59 Sharona Hoffman, *Continued Concern: Human Subject Protection, the Institutional Review Board, and Continuing Review*, 68 TENN. L. REV. 725 (2001).

60 45 C.F.R. § 46.108(a)((4)(i) (2018).

61 DEP'T HEALTH & HUMAN SERVS., OFFICE FOR HUMAN RESEARCH PROTECTIONS, Unanticipated Problems Involving Risks & Adverse Events Guidance (2007).

62 The self-assessment tool for OHRP's Quality Assessment Program asks whether the IRB operates a "hot line or 800 number for potential or enrolled participants to file complaints or direct questions regarding human subjects protection issues," as well as whether the IRB provides an advocacy program or ombudsman for participants, but no additional guidance appears to be available in this area. Office for Human Research Protections, QA Self Assessment Tool, https://www.hhs.gov/ohrp/education-and-outreach/human-research-protection-program-fundamentals/ohrp-self-assessment-tool/index.html (Retrieved March 5, 2013).

63 Cal. Health & Safety Code § 24173(c)(10).

64 Research with prisoners may be an exceptional case. California also requires that "provisions have been made for compensating research related injury" occurring to prisoners enrolled in research, and that the Department of Corrections provide a process for hearing grievances occurring in research. Cal. Penal Code § 3515(d), 3518.

65 N.Y. Pub. Health § 2444.

66 N.Y. Pub. Health § 2442.

related complaints.

Apart from federal and state law, a quasi-binding requirement for institutions to address research-related complaints arises from professional accreditation. Modern IRBs are often part of broader "human research protection programs" in research institutions, which encompass functions such as protocol review and approval, research ethics instruction for investigators and research staff, development of institutional policies, ensuring compliance of research protocols with state law, monitoring conflicts of interest in research, and managing unanticipated problems and adverse events. Human research protection programs can apply for accreditation by the Association for the Accreditation of Human Research Protection Programs (AAHRPP), which has two requirements relevant to the management of research-related disputes. First, researchers and staff must "have a process to address participants' concerns, complaints, and requests for information."[67] Second, organizations as a whole must "ha[ve] and follo[w] written policies and procedures that establish a safe, confidential, and reliable channel for current, perspective, or past research participants . . . that permits them to discuss problems, concerns, and questions; obtain information; or offer input with an informed individual who is unaffiliated with the specific research protocol or plan."[68] This duty is not located with the IRB; for example, organizations could fulfill the requirements using an ombudsman or research subject advocate. AAHRPP has set no requirements for structure of these processes, but simply requires that they exist, and implies that they should handle all types of concerns—including those that are not justiciable in public courts.[69]

Aspirational ethics documents provide "soft law" principles that plausibly imply that researchers and research institutions have a duty to address the full range of participant complaints.[70] As noted above, there is a vast array of ethical guidance documents in medical research, including the Belmont Report, the Nuremberg Code, the CIOMS guidelines, the WMA Declaration of Helsinki, and others. These offer additional values that might be relevant to dispute systems design here, but no specific procedural guidance. On the basis of the Belmont report, for example, the design of a dispute resolution system in this field might seek to promote participant autonomy, beneficence, non-maleficence, and justice defined as

67 Assoc. for Accreditation of Human Research Protection Programs, AAHRPP Accreditation Standards, Oct. 1, 2009.

68 *Id.*

69 IRB professionals can pursue individual certification through the Certified IRB Professional program (CIP) run by the Public Responsible in Medicine & Research (PRIM&R) organization. This program, however, does not provide specific training on the management of research-related disputes. PRIM&R, CIP Body of Knowledge/Content Outline, https://www.primr.org/certification/cip/bodyofknowledge/.

70 Underhill, *supra* note 18.

equitable access to the benefits and burdens of research.[71] But these broad norms leave wide latitude for procedures that attempt to address grievances arising in the course of research.

In some ways, this flexibility is typical of research oversight more generally, in which the regulation of research is broadly decentralized and delegated to IRBs as what Laura Stark has called "declarative groups—their act of deeming a practice acceptable would make it so."[72] The federal regulations do not dictate the outcome of any particular protocol, but rather leave these decisions up to IRBs themselves, even permitting IRBs to waive informed consent requirements entirely under certain conditions.[73] IRBs also retain procedural flexibility in the format of their deliberations, and institutional practices on IRB membership and deliberation vary; variation across IRBs is reinforced by consulting prior decisions within the institution as precedent.[74]

D. Gaps in Understanding Research-Related Disputes

Despite near-total freedom for the design of IDR processes in this field, the actual dispute resolution practices of research institutions operating in this regulatory gap have gone entirely unexamined. Drawing on the literatures above, many similar process values will be important for the resolution of disputes in this field. These include the values of participation, legitimacy (including legitimacy of the process and broader legitimacy of scientific research), predictability for potential and actual disputants, equality and accountability in a context where research subjects are less powerful than research institutions, revelation for subjects interacting with a highly specialized field of knowledge, and dignity and privacy interests for all disputants. Moreover, dispute systems for resolving research-related disputes likely have similar proximate goals to other ADR processes, such as efficiency, durability, and party satisfaction.

I have previously noted the range of grievances that may arise in human subjects research.[75] Most previous scholarship in this area has focused on physical injuries that are inherent risks of research,[76] or that arise from negligence in protocol design, approval, or implementation.[77] Litigants bringing tort claims

---

71 NAT'L COMM'N FOR PROTECTION OF HUMAN SUBJECTS OF BIOMEDICAL AND BEHAVIORAL RESEARCH, DEP'T OF HEALTH, EDUCATION & WELFARE, THE BELMONT REPORT (1979).

72 Laura Stark, *Victims in Our Own Minds? IRBs in Myth and Practice*, 41 L. & SOC'Y REV. 777 (2007) [hereinafter Stark, *Victims*]; LAURA STARK, BEHIND CLOSED DOORS: IRBS AND THE MAKING OF ETHICAL RESEARCH 164 (2012); Laura Stark & Jeremy A. Greene, *Clinical Trials, Healthy Controls, and the Birth of the IRB*, 375 N. ENG. J. MED. 1013 (2016).

73 45 C.F.R. § 46.116(f) (2018).

74 Stark, *Victims*, *supra* note 72; STARK, *supra* note 72, at 165.

75 Underhill, *supra* note 18.

76 Pike, *supra* note 5.

77 Mello et al., *supra* note 19.

against research institutions have alleged wrongs including negligent protocol design or implementation, lack of informed consent, emotional distress, fraud, misrepresentation, battery, medical malpractice, products liability claims, privacy violations, breach of contract, wrongful death, state law violations, conspiracy, participant abandonment, unjust enrichment, and IRB misconduct including negligent study approval and oversight.[78] Additional claims may include failure to disclose individual study results, premature study termination, and withholding or denying access to treatments after the study has concluded.[79] Complaints made outside litigation have included allegations of noncompliance with protocols, delayed payments, unwanted requests for study participation, perceived HIPAA violations,[80] and lack of confidentiality.[81] Research on the therapeutic and preventive misconceptions suggests that many participants do not fully understand protocols at the time of informed consent,[82] which can generate complaints later. Many of the concerns visible in healthcare settings more generally—such as perceived rudeness, long wait times, miscommunications, and other "small-scale disputes"[83]—are almost certainly present in the research context as well. Other complaints may have more in common with workplace grievances; many participants in non-therapeutic research see their participation as paid work, and view study terms as conditions of employment.[84] There has been no systematic study, however, of how institutions may seek to resolve the universe of participant concerns.

For many if not most of these claims, IDR processes are the only available venue for dispute resolution. Litigation is a poor fit for many of these disputes. Some of the claims noted above have been rejected by courts (e.g., claims to post-trial access[85]) or do not allege legal violations (e.g., unwanted requests for study participation). Litigation is also legally or practically unavailable for some categories of research subjects. As Elizabeth Pike has pointed out, international participants may be barred from recovery due to the Federal Tort Claims Act and

78 Underhill, *supra* note 18 (citing additional sources), Mello et al., *supra* note 19; Saver, *At the End*, *supra* note 23; Saver, *Medical Research*, *supra* note 23; Morreim, *supra* note 19.

79 Gordon 2009, Saver, *Medical Research*, *supra* note 23

80 HIPAA does not provide for a private right of action, but state courts may rely on HIPAA for setting the standard of care in tort actions for privacy violations. Byrne v. Avery Ctr. for Obstetrics & Gynecology, P.C., 314 CONN. 433 (2014) https://apps.americanbar.org/litigation/litigationnews/top_stories/030215-hipaa-disclosure.html.

81 Underhill, *supra* note 18 (citing sources).

82 Flory & Emanuel, supra note 8.

83 Rabinovich-Einy, *supra* note 6.

84 Peter Davidson & Kimberly Page, *Research Participation as Work: Comparing the Perspectives of Researchers and Economically Marginalized Populations*, 102 AM. J. PUB. HEALTH 1254 (2012). The journal GUINEA PIG ZERO—an "occupational jobzine" for study participants—is emblematic of this view. http://www.guineapigzero.com/. *See also* sources cited *infra* n. 88.

85 Saver, *At the End*, *supra* note 23.

the Alien Tort Statute,[86] and US participants in federally conducted research may find their claims precluded due to sovereign immunity and the discretionary function exception to the Federal Tort Claims Act. Litigation has a number of drawbacks in the research context as well, including high costs that may raise the costs of research and lead IRBs to make excessively conservative decisions about study approval.[87]

The literature on research-related disputes sheds little light on IDR options. Although some institutions provide no-fault compensation programs for research-related injuries, such programs are rare,[88] and we know little about the processes or process values they employ. Several protocols have set up study-specific ADR (not necessarily IDR) processes; interestingly, the two published papers on these processes are in HIV/AIDS research, reflecting the history of participant advocacy and community-based research.[89] Both programs resembled arbitration. In one program, a series of HIV vaccine trials in India created a three-person arbitration board to handle all grievances.[90] The other program was an informal arbitration system established for a consortium of AIDS treatment trials and was established to promote participants' "right to be treated with dignity"; participants could have their complaints represented by a social worker before a study panel, with the option to appeal the panel decision to the IRB.[91]

Only one published paper has described institutional practices for complaint resolution, published by IRB professionals at the Baylor College of Medicine.[92] The Baylor system provides for an "iterative process that seeks to identify the truth about research-related complaints through fact-finding efforts."[93] As understood by this IRB, due process requires objectivity and the opportunity for all parties "to speak to the 'truth' as they perceive it."[94] Procedural elements include the requirement of a written complaint, IRB classification of the complaint as noncompliance or scientific misconduct, notification of a compliance assessment team and the principal investigator, a formal audit of study materials and "fact-

86 Pike, *supra* note 5; see also Sarah Gantz, *Judge Dismisses $1 Billion Guatemalan Syphilis Experiment Case against Hopkins, Others*, BALTIMORE SUN, Sept. 9, 2016; Estate of Alvarez v. Johns Hopkins Univ., 205 F. Supp. 3d 681 (D. Md. 2016).

87 Mello et al., *supra* note 19; Underhill, *supra* note 18.

88 Pike, *supra* note 5; Elliott, *supra* note 19.

89 Underhill, *supra* note 18.

90 J.L. Excler et al., *Preparedness for AIDS Vaccine Trials in India*, 127 INDIAN J. MED. RES. 531 (2008).

91 Lisa E. Cox & Thomas M. Kerkering, *Grievance Procedures as Assurance for the HIV-Infected Clinical Trial Participant*, 1993 AIDS PATIENT CARE 20 (1993).

92 Kathleen J. Motil, Janet Allen, and Allison Taylor, *When a Research Subject Calls with a Complaint, What Will the Institutional Review Board Do?*, 26 IRB: ETHICS & HUMAN RESEARCH 9 (2004)

93 *Id.* at 13.

94 *Id.* at 9.

finding" through interviewing relevant parties, review of factual findings by an IRB subcommittee, a face-to-face "hearing" involving the investigator and IRB subcommittee (but not the subject), full IRB deliberation and a preliminary decision imposing corrective actions or sanctions on the investigator, an option for the investigator to appeal the decision, and a final decision letter by the full IRB setting forth factual determinations and a binding corrective action plan. This arbitration-like process appears to prioritize accuracy and investigator voice, but says little about voice or remedy for the individual participant.

IDR has structural limitations in this context, particularly when the ADR process are maintained by institutions themselves. IRBs who maintain ADR processes have divided loyalties to their institutions, their colleagues, and the participants they are tasked with protecting, and IRB administrators may be concerned about their own liability in the event of litigation.[95] Financial incentives for researchers and institutions may encourage unethical behavior in both the oversight and implementation of research protocols.[96] And like all IDR programs, this context raises concerns about privatizing legal norms, transmuting rights-based claims into organizational issues, providing a highly unequal forum, and deterring publicly useful litigation.[97] But where IDR may be the only practicable option for resolving many of these disputes, it is important to interrogate the process choices that institutions have already made.

III. AN EMPIRICAL STUDY OF IDR FOR RESEARCH-RELATED DISPUTES

No previous research has examined the role of IDR in the resolution of research-related disputes. This Part will introduce the study methods, followed by results describing the frequency and nature of complaints, process options, uses of procedural flexibility, and informants' appraisal of their processes. Throughout, I will use "informants" to refer to individuals who participated in my interviews, and "subjects" or "participants" to refer to individuals who lodge (or may lodge) complaints with their IRBs. Where I have quoted informants directly, I have selected quotes that are most striking or most typical of responses across the full set of informants.

A. Methods

The goal of this empirical study was to understand the structure and animating procedural values of ADR processes that research institutions use to manage

95 Mello et al., *supra* note 19.

96 Carl Elliott, *The University of Minnesota's Medical Research Mess*, *The N.Y. Times*, May 26, 2015. http://www.nytimes.com/2015/05/26/opinion/the-university-of-minnesotas-medical-research-mess.html.

97 *See, e.g.*, Edelman & Suchman, *supra* note 47.

disputes involving human subjects. This Part presents the result of in-depth, semi-structured qualitative interviews with human research protections program officers at 30 hospitals and academic institutions throughout the US. All procedures were approved by the Yale Human Subjects Committee and advised by an expert panel of 6 scholars and IRB professionals. Data were protected by a Certificate of Confidentiality (COC) from NIH, which aims to facilitate research on sensitive topics by shielding individual participant data from subpoena.[98]

The population of interest for this study was IRB chairs, directors, and other designated IRB personnel who have discretion in responding to complaints; all individuals in the study had at least 1 year of experience reviewing human subjects protocols and had discretion in managing institutional responses to participant complaints. I interviewed one person per institution, with the exception of one institution, where I ran a joint interview with two IRB officers. Twenty-six of the 31 informants were chairs or directors of their IRBs; the others were managers or administrative chairs.

The unit of analysis for this study was the institution; I included IRBs that reviewed protocols for a hospital or academic institution, were located in the US and subject to US federal regulation, and had an OHRP-approved federal-wide assurance number.[99] A majority of eligible institutions[100] were academic institutions that encompassed both medical and nonmedical schools; I oversampled hospitals and universities lacking medical schools to ensure adequate data from these types of institution. The final sample included 20 universities with

98 Leslie E. Wolf et al., *Certificates of Confidentiality: Protecting Human Subject Research Data in Law and Practice*, 43 J.L. MED. & ETHICS 594 (2015).

99 Although institutions have procedural freedom to have complaints resolved outside the IRB, common practice is for this to be a core IRB function, and I thought IRBs would be aware of dispute resolution practices even if they occurred outside the IRB office.

100 Because this study was funded through the Fordham HIV Prevention Research Ethics Training Institute, a second part of the interviews specifically considered management of disputes arising in biomedical HIV prevention research. To fulfill this part of the study, I further limited the sample to institutions that had received funding from any source within the previous 5 years to conduct social science research or clinical research on biomedical HIV prevention. I identified eligible institutions by searching all active protocols in the NIH RePORTER database as of May 2013, clinicaltrials.gov, and consulting trials networks for biomedical HIV prevention research. One hundred and sixteen unique institutions were eligible for inclusion. I used a computer-generated random number sequence to select simple random samples of ten institutions at a time for recruitment. I approached 73 institutions to secure the sample of 30 included interviews; the other 43 institutions either did not respond (35) or declined (8) for reasons including busy schedules or lack of expertise handling participant complaints. This design introduces some inevitable weaknesses; for example, there may be some social desirability in responses, and participating institutions may have been more comfortable discussing their procedures—perhaps because they had better-defined procedures or fewer negative experiences with research-related disputes. These limitations are inherent in most qualitative research designs, but they are balanced in this study by the strength of a simple random sampling procedure among eligible studies, full data saturation on all themes of interest, and stratification of recruitment across three different types of institution.

medical schools, 4 universities without medical schools (where almost all protocols were for social and behavioral research), and 6 hospitals.[101] Institutions were located in all four US Census Bureau regions, and the sizes of their research portfolios ranged from 20 to more than 5,000 active protocols enrolling human subjects.

I did not include external or centralized (independent) IRBs; although centralized IRBs approve research protocols (and may experience liability for negligent approvals), they have different liabilities from institutions who receive funds from research sponsors, employ investigators and research staff, and provide material support and physical space for study activities. I also did not focus on private industries; although drug and device manufacturers conduct human subjects protocols, they also may have a somewhat different set of liabilities as manufacturers. They also often rely on centralized IRBs, or may subcontract trials to hospitals or clinics. Limiting the scope of this project to hospitals and universities provides a first cut at the question of how research institutions resolve complaints and injuries involving human subjects, and subsequent work should focus on other research settings.

Each interview lasted 60-90 minutes and focused on the types and frequency of complaints, experiences with litigation involving human subjects, the development and application of ADR procedures for resolving research-related disputes, the need for guidance or training to handle research-related disputes, and the principal values or priorities of the institutional ADR processes.[102] I conducted and audio-recorded all interviews, then analyzed verbatim transcripts thematically using NVivo 11, which allows the application of a formal coding structure to qualitative data. I used an initial set of planned codes for data analysis, but added new themes as they emerged from the data.

---

101 This sample size is appropriate for the collection of nuanced, in-depth data that explores variation and meaning in experiences, and it allows for data saturation. *See* Janice M. Morse, *Determining Sample Size*. 10 QUALITATIVE HEALTH RES. 3 (2000); Janice M. Morse, *The Significance of Saturation*, 5 QUALITATIVE HEALTH RES. 147 (1995). Data saturation refers to having collected sufficiently rich data to understand the key relationships at stake in the study—that is, collecting data until no new themes emerge with additional interviews—and although no formal metrics of saturation exist, qualitative researchers monitor their findings throughout studies to ensure that they research saturation before concluding data collection. *See* Morse, *The Significance of Saturation*; *see also* CONSTRUCTING GROUNDED THEORY: A PRACTICAL GUIDE THROUGH QUALITATIVE ANALYSIS (Kathy Charmaz, ed. 2006). I monitored for data saturation throughout this work by completing and transcribing a debriefing after each interview, then rereading debriefing reports to identify new and recurring themes. The final sample enabled a thorough exploration of the themes of this paper.

102 Each individual participant provided informed consent to interviews, completed an interview by phone, and received $100 for their time. To protect institutions that may be experiencing research-related litigation, informed consent used an anonymous verbal process, and all data were deidentified before analysis.

Table 1 reports information on the individual informants, while Table 2 reports information about the institutions they represented.

Table 1. Characteristics of individual respondents

Characteristic	Percentage (n = 31)
Position	
Director	77%
Manager	13%
Chair or Administrative Chair	6%
Administrator	3%
Gender	
Female	74%
Male	26%
Median Time in Current Position	4.5 years (range 0.25-25 years)
Median Time in Research Protections	12 years (range 2-25 years)
Median Time Managing Complaints	9 years (range 1-25 years)
Highest Degree	
B.A./B.S.	16%
M.A./M.S.	39%
J.D.	6%
M.B.A.	10%
Ph.D./M.D.	29%
Certified IRB Professional (C.I.P.) Qualification	
Currently Certified	42%
Previously Certified	10%
Lapsed	32%
Not Known	16%

B. Frequency and Types of Disputes

Despite a wide and colorful variety of complaints that encompassed injuries, noncompliance, human resources issues, unwanted recruitment efforts, and cultural concerns, the overall frequency of complaints was far lower than might be expected. This low frequency was surprising to many informants, who sought to explain low system uptake as not only a result of good research practices, but also a result of subjects' lack of understanding of their rights, interests, and dispute resolution options.

Table 2. Characteristics of institutions.

Characteristic	Percentage (n = 30)
Type of Institution	
University with Affiliated Hospital	67%
University without Hospital	13%
Hospital	20%
U.S. Census Region	
Northeast	40%
West	23%
South	20%
Midwest	13%
Other	3%
AAHRPP Accreditation	
Current	50%
Pending	10%
Not Accredited	37%
Not Known	3%
Median Active Protocols	2,000 (range 20-5,000)
Median Annual Complaints per 1,000 protocols	2.2 (range 0-43.5)
Written Policy or Procedure for Complaint Resolution	73%
Previously Experienced Litigation Involving Human Subjects	30%
Policy for Compensating Subjects for Physical Injury	
Compensated Some or All Injuries	47%
Compensated Depending on Sponsor Agreements	30%
Never Compensated	17%
Uncertain	7%

1. Uptake of the Process

At all but one institution, the IRBs were listed as the resources for participant complaints on patients' informed consent forms; the remaining institution provided participants with information for a patient relations office, which reported any "non-trivial" complaints back to the IRB. Institutions handled between 20 and 5000 protocols; the largest research programs were at universities that included medical schools, while hospitals and universities without medical schools had smaller research programs. Given the variety and commonplace nature of potential participant complaints noted above, the numbers of complaints received by IRBs

were surprisingly low, at an average of approximately 5 complaints per year per thousand active protocols. This figure reflects several outliers with larger numbers of complaints; the median complaint frequency was 2 complaints per year, per thousand active protocols. Complaints were somewhat more frequent at universities with medical schools (median 2.4 per year per thousand protocols), compared to universities without medical schools (median 0.5) and hospitals (median 1.8). Several institutions noted temporary spikes in complaints linked to identifiable events (e.g., media coverage of a protocol using an emergency exception to informed consent), but the stable frequency of complaints was around 2-5 complaints per thousand protocols per year.[103] Complaints were not spilling over into litigation instead; approximately one third of the institutions had been involved in litigation involving research subjects or staff, but these incidents were far less frequent than the number of complaints received.

This Part will explore potential causes for the low frequency of complaints below. The low figures observed here align, however, with fields such as medical malpractice and complaints about healthcare professionals, in which "most people choose to 'lump' their grievance (i.e., put up with it or ignore it) or to avoid expressing it by 'exiting' (abandoning or limiting) the troublesome relationship. In the medical context . . . the vast majority of patients do not sue for negligently caused injuries . . . . Studies of complaining and claiming behavior are, therefore, studies of atypical behavior."[104]

2. Subject Matter of Complaints: Rights and Interests

Despite this low uptake, when subjects *do* use the process, the subject matter of their complaints varies widely and encompasses both legally cognizable and non-justiciable claims. Complaints are typically brought by subjects themselves, or family members of participants who are minors, participants who have diminished capacity to consent, and participants who are ill or deceased. Study staff may also bring complaints as whistleblowers, particularly when complaints concern the conduct of principal investigators. I did not include here complaints from principal investigators about IRB actions; many institutions reported these, but because my focus is on research subject disputes, they were outside the focus of this study.

Most institutions reported that complaints about the speed and adequacy of

103 This is lower than a 2011 AAHRPP study that reported an average of 7.9 complaints per year per thousand protocols, among 193 AAHRPP-accredited institutions. My study included institutions with and without AAHRPP accreditation; the average for AAHRPP-accredited institutions was 4.0 complaints per year per thousand protocols. AAHRPP, *Metrics on Human Research Protection Program Performance* (2011).

104 Jost, *supra* note 29, at 314

participant compensation are most prevalent, particularly for participants enrolling in non-therapeutic protocols (who may be more interested in compensation, rather than receiving an experimental intervention). Other complaints include concerns about the availability and adequacy of the informed consent process (especially for non-English speakers, minors, or elderly participants); waivers of informed consent or HIPAA authorization; data privacy and confidentiality; the release of research reports or publications that did not protect participant confidentiality; disrespectful, nonresponsive, harassing, discriminatory, or dismissive treatment by research staff; staff noncompliance with study protocols; sexual harassment by research staff; dissatisfaction with emergency procedures for managing psychological events during research studies (e.g., threats of suicide); study requests for personal identifiers, especially Social Security numbers; unexpected, painful, or offensive study activities; requests for the return of biological samples; requests to discontinue participation; student concerns about the use of educational records; complaints about physical accessibility of study premises for individuals with disabilities; anger about premature study termination, where studies had been stopped by researchers, sponsors or the IRB; requests for access to individual study results or other records; concerns about future use of study samples or data; adverse social or legal consequences of participating in study procedures;[105] malfunctioning equipment or technology provided by a study; and cleanliness of study facilities. Complaints also include physical injuries, particularly where participants believe they had not received a timely and thorough response from the investigative team.

Almost all institutions reported additional complaints from individuals *not* enrolled in research protocols. These complaints include community objections to study advertising (e.g., concerns about how study posters depict LGBT individuals); concerns about study recruitment and consent processes where sensitive protocols have received media attention; frustration with being found ineligible for participation in a particular study (particularly for patients who want access to a therapeutic protocol), or being excluded from a study midway through due to noncompliance or changes in eligibility; complaints that studies are wasting money on answering trivial or obvious research questions; concerns about repeated requests for study participation after refusal; student concerns about pressure to participate in professors' research; complaints from community organizers about a mismatch between expected and actual research activities in the populations they represent; complaints about researchers' misuse of access to medical records

105 For example, one protocol enrolled a sex worker in an HIV vaccine trial, which causes the body to produce HIV antibodies despite the absence of infection. These antibodies caused her to test positive in an HIV antibody test when she was later arrested for prostitution, which triggered mandatory name-based reporting to the state and possibly enhanced penalties for the prostitution offense.

databases; and concerns from patients who were angry that the IRB had not yet approved a research protocol that they perceived to be beneficial. Some particularly sensitive complaints from non-participants also included concerns about culture, reputation, or identity; for example, complaints alleged that research results would harm the reputation of a community or organization, or that researchers were making inappropriate use of biological samples to study a Native American community.

Numerous institutions had received complaints from individuals who had been identified and contacted as potential subjects on the basis of their medical records or a state registry (e.g., asked to be in a prostate cancer study because their medical records included a prostate cancer diagnosis), which did not fit their expectations of medical record privacy. Institutions who reported these complaints had almost universally enacted institutional policy changes barring investigators from cold-calling participants on the basis of their medical records.

C. Process Goals and Values

Despite the heterogeneity across institutions in location, type of institution, and size, there were remarkable similarities in how institutions viewed their proximate goals and underlying process values. This Section will discuss each in turn, noting similarities in how informants described their systems.

1. Proximate Goals

Informants reflected on a number of institutional goals for their dispute resolution systems. These goals included system outputs that are separable (and often measurable) results of the process (e.g., participant satisfaction), as well as a common set of desirable procedural features (e.g., neutrality of the decision-maker, transparency).

Most informants noted that the IRB's institutional role is to protect subjects enrolled in research protocols; the quality of decisions depended on how well they fulfilled this substantive goal, in addition to complying with federal regulations and ethical guidance. For these institutions, complaints are a source of feedback for modifying risky protocols or practices, and the resolution process sometimes led to system-level changes to policies applying across the institution.[106] Subject and investigator satisfaction with the process—if not the outcome—is also a primary goal at all institutions, and informants often described "customer satisfaction" or a "consumer service" approach for subjects as an overriding emphasis. As expected, another salient goal of this process is to protect the institution itself from litigation and adverse media exposure, in part by satisfying

106 *See* Sturm & Gadlin, *supra* note 34.

individual subjects' concerns, but also by maintaining an active feedback loop that identifies systemic risks. IRBs noted that individual or repeated complaints can identify defects in institutional policies, providing opportunities for revision and reform. As one informant noted, "The most important thing is to ensure the patient is, feels comfortable in the resolution . . . I guess secondly would be to ensure that we've implemented whatever processes need to happen to ensure it doesn't happen again." Or as another said, "[We have] more policy-type resolutions so that I can go back to [an] individual and say . . . the institution has now changed its policy in a way this will not happen again . . . . A quality resolution . . . is not just a quick band-aid fix, but more of a long-term, proactive [step]."

Like many ADR processes, these systems aim for efficiency, speed, and accuracy, in part assured by the procedural flexibility inherent in the system design.[107] Speed was often cited as a goal of complaint resolution, with multiple informants noting that lengthy complaint processes may foster escalation of the dispute, particularly if parties are not kept abreast of progress. Consistency was another procedural goal, often fostered by written or standardized procedures. Conserving financial and administrative costs, however, was not typically a priority, and the costs of the ADR process were viewed as small in comparison to the threat of litigation and reputational exposures for the institution. Informants reported willingness to devote considerable time and resources to complaints in the interests of accuracy and fairness, and the low number of complaints enabled IRBs to prioritize thoroughness over administrative costs. ("As far as time, manpower, and all of that is concerned, I think you have to spend what you have to spend in order to make it a fair process.") Moreover, very few complainants sought financial compensation for their grievances, with the exception of participants with uncompensated injuries or complaints related to expected payment for study activities.

In order to fulfill these proximate goals, institutions sought to create processes with a number of ideal safeguards. These included an easily accessible forum; having a written procedure or having the same personnel respond to all complaints; a full opportunity for subjects and investigators to provide their version of the facts, including in-person or phone meetings with the IRB; options for the subject to elect anonymity or choose not to pursue corrective action; an opportunity for parties to choose facilitated negotiation or mediation; an initial triage point that allows for emergency actions such as study suspension; transparency about the process and communication of the outcome to investigators and subjects; provision of a third-party neutral with the authority to issue decisions that bind the institution; consultation of all complaint stakeholders and institutional actors, including trusted members of the subject's community where relevant; privacy of

107 *See* infra, Section III.E.

deliberations and decisions by the third-party neutral; a thorough fact-finding process that consults all relevant parties; a written, reasoned decision; opportunities for the investigator to weigh in on the corrective action plan; and an option to appeal.

2. Values of the Process

Informants' beliefs about the underlying value of a complaint resolution process reflected many, if not all, of the process values described in Part II above. The value of participation resonated most deeply throughout the interviews, both as an instrumental value (necessary to reach a resolution, promote legitimacy, or defuse conflict) and as an inherent value (an independent good for subjects who exert their autonomy by complaining). Informants intuitively described some themes arising in the procedural justice literature, such as an "opportunity to be heard" (voice); the need to treat participants empathetically and respectfully (courtesy); the need to provide a forum that approximates a neutral third party (neutrality); and the need for the IRB to be trustworthy or receive buy-in from trusted community authorities (trust). Procedures that involve all possible stakeholders to a complaint also advance participation values, and may also increase the legitimacy of both the forum and the substantive decisions made by the IRBs.

Equality between the subject and investigator is a second value, given IRB's efforts to provide neutral decisions and full participation opportunities for both sides. Accountability of the investigator for wrongs was an important corollary to this principle; importantly, however, this accountability is one-sided. Although the IRBs can compel investigators to take corrective actions, they have neither the authority nor the desire to sanction subjects. Of course, a final IRB decision that is adverse to the subject forecloses other options, particularly for non-justiciable complaints. But subjects cannot be made worse off ex post. The focus of accountability was also on individual investigators rather than the institution more generally, save for physical injuries (which may be compensated by institutional funds) and complaints that specifically alleged misconduct elsewhere in the institution (e.g., negligent approval of protocols by the IRB itself).

Informants' focus on consistency and the need for procedural transparency with complainants and investigators suggested that predictability was an important goal. Informants did not, however, identify the need to provide procedural information to subjects *before* the act of bringing a complaint. Indeed, very little about the process was disclosed ex ante, in part because IRBs maintained so much procedural discretion that precise procedural details were not known in advance. As overseers of the informed consent process, IRBs are well acquainted with the problems of how best to disclose information to research subjects. The difficulties

of obtaining informed consent are notorious.[108] Limited time is available for obtaining informed consent; participants may already be overwhelmed with information about the details of the research protocol; and the informed consent process often fails to present information in an accessible way. Prior studies have consistently found deficiencies in informed consent. One review found that participants lacked adequate comprehension of the study in 29% of research protocols, and lacked comprehension of the risks of surgery in 36% of surgical research protocols.[109] Participants in only 44% of protocols knew that they could withdraw from the study.[110] Studies worldwide have found similar results, showing that comprehension varies widely, and that randomization and placebo-controlled trials present particular stumbling blocks for comprehension.[111] When it is already difficult to present significant facts about the research protocol in an accessible way, researchers may be limited in their ability to disclose detailed procedural information about participants' dispute resolution options. Against this backdrop, informants in the present study generally had not questioned the current practice of disclosing IRBs' contact information without further details about the dispute resolution process.

Informants were less likely to describe privacy as an independent value, with the exception of privacy for investigators who experience disciplinary sanctions. Instead, procedures safeguarded privacy interests in an effort to promote participation values, particularly in the use of procedures to receive and manage complaints made by subjects who wished to stay anonymous or confidential. No informant described using a formal confidentiality or nondisclosure agreement during the process, but internal deliberations of the IRB were wholly confidential as an institutional practice.

Finally, a number of legitimacy interests were served by a well-functioning complaint process. These included the legitimacy of the IRBs' substantive decisions about complaints, but also legitimacy of the institution, particularly in its relationships with research communities, as well as the legitimacy of science more generally, as some later quotes will demonstrate. As one informant noted, "We protect human subjects and we facilitate research at the institution, because research with human subjects does improve healthcare at the end of the day . . . . it's important that our institution be trusted to have the best interest of our patients and you know, um, society in the research that we're doing."

---

108 Christine Grady, *Enduring and Emerging Challenges of Informed Consent*, 372 N. ENGL. J. MED. 855 (2015); Tom L. Beauchamp, *Informed Consent: Its History, Meaning, and Present Challenges*, 20 CAMBRIDGE Q. HEALTHCARE ETHICS 515 (2011).

109 Falagas, *supra* note 7.

110 *Id.*

111 Amulya Mandava, et al., *The Quality of informed Consent: Mapping the Landscape. A Review of Empirical Data from Developing and Developed Countries*, 38 J. MED. ETHICS 356 (2012).

In addition to these intuitive process values, some informants also sought to advance the values of the Belmont Report on ethical research with human subjects. These particularly included respect for subjects' autonomy and the need for beneficence and non-maleficence toward research subjects. These values might be reclassified into the interests above, such as dignity and equality—but the Belmont Report is unique to the lopsided power structures in the research setting, and may be less instructive for other types of IDR.

D. Elements of Process

Despite the lack of regulatory guidance on how institutions should handle research-related disputes—which might be expected to generate some heterogeneity in dispute system design—almost all institutions have developed similar and procedurally flexible dispute resolution systems. Even where some institutions differ slightly (e.g., a few hold institutional insurance for subject injuries; a few request local community leader involvement for disputes arising in foreign or culturally distinct groups; a few have a patient representative), the contours of the basic process remain the same. Although this may result in part from the process of AAHRPP accreditation, AAHRPP does not mandate particular dispute system design features, and even the institutions that were not accredited handle disputes similarly. This Section will therefore group all types of institutions together for the analysis.

Across institutions, complaint resolution processes most commonly resembles binding arbitration for disputes that are not the result of a factual misunderstanding. For minor disputes arising solely from a misunderstanding or miscommunication, the process may be more similar to facilitated negotiation or even simple education. All processes are developed and managed internally by the IRB, and they rely on the IRB to issue binding decisions as a third-party neutral vis a vis the participant and investigator. Processes follow a rough timeline of complaint receipt, internal discussion of procedural options, "fact-finding" carried out directly by the IRB or a research compliance team, deliberation by the IRB, and issuance of a binding, written resolution enforced by the IRB's authority to approve or disapprove the research protocol. The remainder of this Section will consider the origins of these processes, procedural similarities across institutions, and the chronological series of steps in the process.

1. IRB as Dispute System Designer: Process Origins and Design

Almost all the institutional processes in this study arose informally within the IRB and solidified over time, as IRBs received specific, but rare complaint calls from research subjects. Where institutions had an ex ante process, it was typically created as part of a broader reorganization of the IRB, or it was imported by a new

director or chair familiar with a process from a previous institution. A minority of institutions had no written process for managing complaints; they considered this an "office practice," or believed that they experienced too few complaints per year to require a written procedure ("I mean it happens maybe five times a year so, uh, knock on wood"). The likelihood of having a written policy differed little depending on the type of institution (hospitals, universities with or without medical schools). These written procedures were internal, and although several institutions post them internally, none described making them available to research subjects at the time of enrollment. No institution mentioned consulting subjects or subject representatives systematically at the time of process design.[112]

Where institutions had written procedures, most had developed them to fulfill the requirements set by AAHRPP.[113] Many, but not all, had consulted other institutions' policies at the time of accreditation. Informants at other institutions, including non-accredited programs, noted that they had developed written policies unprompted to increase efficiency ("[Before our written procedure,] not everybody was consistent, things were getting missed."), and to increase consistency across protocols and over time ("I think that's your biggest, you know, benefit is making sure that everything is handled in a fair, unbiased, consistent manner.").[114] The central goals of process standardization were to ensure similar treatment across all participants and investigators, and to reduce biased procedural decisions that may arise from prior knowledge of the investigator.

> Whenever an issue came up that we needed to resolve, we realized that we shouldn't do it ad hoc, you know depending on the PI [principal investigator]. If we knew the PI was a good guy to do one thing versus, um, doing something else. So we, we realized back then you're much better off to have upfront processes put into place -- to treat everyone the same -- and go down the same algorithm of decisions -- versus a hit or miss, which is you know what we were doing before we had the SOP [statement of procedure] in place.
>
> It's really important to us as an institution and as an office

112 One institution did, however, involve a patient representative throughout the process and involved that person in the process design.

113 Eighteen of the thirty institutions were accredited or pending accreditation, including all six hospitals, 12 of the universities with medical schools, and none of the universities without medical schools.

114 One institution had also interpreted the federal regulations and OHRP guidance on mandatory reporting of unanticipated problems to require a written process for resolving complaints, in the event that complaints alleged such problems or noncompliance. Other institutions, however, had not interpreted the regulations this way.

> specifically that we want to set precedent. Like we want to treat each case as very similar, we want to have a very similar outcome and so if we determine that we have a different outcome, we want to look at why . . . . There are investigators that have, uh, have kind of proven themselves to be very quality investigators, and then certainly I think every institution has investigators that are known to be a little bit less by the book . . . . But if the same complaint came in, the equivalent complaint came in under the same two, you know, under these two investigators, they should be handled exactly the same with the same neutral approach.

Both those with written and unwritten procedures, however, believed that a written procedure would be important in the event that a subject complaint resulted in litigation. For example, one institution without a written procedure suggested that they may be "at risk for not having it more codified . . . . But, you know, usually something bad has to happen and then you become codified." An institution with a standard written process noted that a key motivation was the belief that own compliance with internal procedures would have value in litigation.

> Some complaints . . . were bypassing [the director's] office and going right to [the IRB] committee. And they were meritless. And then there were other complaints that would come to me but there was no formal process -- there was no standard operating procedure . . . . And so we just codified the, um, process flow . . . . You know, if it did get to litigation we -- we could say that we were or were not following our own internal policies. So [we shifted] from no policy to policy. Based on experience, we knew what worked and what wasn't working. We knew where exposures were . . . legal exposures, regulatory exposures.

Most institutions noted that they continued to revise and update their processes over time, to respond to changes in complaints or the institutional environment ("We learn what works and what doesn't work and what's more efficient for the participant and the study team . . . . It's a continual learning basis."). Whether procedures were written or unwritten, however, basic procedural features and proximate process goals were similar across institutions, and all relied on the IRB as a third-party neutral, as the next sections will note.

2. IRB as Complaint Line: Initial Contact

All IRBs provide their contact information to subjects via the informed consent form, or if a verbal informed consent process is used, subjects receive

independent notice of the IRB contact information. Most subjects communicate complaints by phone, although some IRBs noted receiving isolated complaints by email or (sometimes-anonymous) written letter or email, and these IRBs responded by phone if possible. Phone calls may direct to a general office, but they are then redirected to a single person such as the director, administrator, chair, or manager of the IRB.[115] Institutions that received complaints through other channels, such as those going to the president or provost's office in a university, typically referred these back to the IRB. Most institutions do not require a written complaint; instead, the IRB personnel prepares a written description on behalf of the subject at the time of the call, and some fill out standard forms during the call to ensure that they are obtaining all the relevant information.

Almost without exception, the informants emphasized the importance of the initial conversation with an aggrieved subject. The immediate goals of this conversation are to obtain a detailed description of the complaint, to identify the relevant protocol and investigator, to identify any previous efforts to resolve the complaint with the investigator, to identify threats of violence or psychological needs, and to understand the remedy that the subject was requesting, if any. But at the time of first contact, IRBs also seek to provide the subject with a full opportunity to voice their complaint without interruption, to ensure that the subject feels heard and respected, to express respect and empathy, and to convey that the subject has been heard by someone who has the institutional authority to resolve the dispute.

> The number one thing we're trying to do is to listen, even if we don't get a complete understanding of the complaint, I mean, that's another goal, but the most important thing is that the person on the other end hangs up the phone feeling that they were heard. They want to get to somebody right away, without having to go through lots of different people, who has the authority and responsibility to listen to them and to, who can help them. So that's number one. And then number two, our perception is that, uh, they want somebody who's going to listen, um, in an empathetic way.
>
> The primary goal actually is to ensure that the subject feels heard. To make sure that whoever is calling, whatever the concern is, that they have some hope that in fact, uh, someone is going to take their call seriously. And while we obviously cannot, uh, promise to the caller that whatever resolution happens will be done, you

115 Several forms also provide numbers for multiple contacts at the IRB, in cases where the IRB chairs also conduct research and may have complaints arising in their own research studies.

> know, to his or her satisfaction -- we can at least reassure the caller that, um, they've, they've reached someone who is going to help them.
>
> I want to, um, allow them to tell their story . . . being, you know, caring and, um, respectful . . . . I would confirm back to them that, you know, we understand that it's upsetting to them . . . . Once I've heard from them I like to clarify back to them what I heard and what my understanding is of their concern . . . [I'm] making it clear that their concerns have been heard and understood. People really need to be heard.

Informants noted that hearing the subject could serve instrumental reasons—it can help defuse emotions and ensure that the process is responsive, and sometimes having a voice fulfills the subject's entire goal in complaining ("Some people will call and say, you know, here's my grief, but at the end of the day they just want to vent and don't really want me to follow up with that, and don't want to leave their name and number."). Informants also noted, however, that this also serves inherent values that might be described as dignity interests, at least in our taxonomy of process values—here, these interests include the desire to be "taken seriously" and to have someone in power acknowledge the emotional impact of the perceived wrong. These expressions of empathy can also promote legitimacy of the process and institution, as one informant noted:

> Usually if they know that you're concerned about them . . . this reflects on us as much as anybody else. We want research to be done ethically. We want all research participants to feel like they can come to us with any um concern or complaint and so I usually reassure them to let them know that we take every complaint seriously, that we'll investigate it, and we'll work with them until the problem is resolved.

3. IRB as Communicator: Ongoing Communications with Participants and Investigators

At the time of initial contact, most institutions also offer participants some input on process and offer procedural safeguards. All institutions offer subjects the opportunity to make their complaint anonymously, without disclosing their identity at all, or confidentially, without disclosing their identity to the investigator.[116] (They note, however, that anonymity or confidentiality may limit

116 One institution even maintains a fully anonymous, non-staffed phone line that anonymizes calls, for people who wish to leave a message without any link to their identity.

the options for resolutions in complaints regarding compensation, investigator misconduct, or harassment.) IRBs typically give participants the option to continue the process toward resolution or corrective action, or to stop the process after the initial call. One IRB member noted that giving the participant this flexibility was an important part of respect for autonomy, which is a core principle of the Belmont Report ethical guidance for research. In the dispute resolution context, this aligns with the broader dignitary and participatory values of process.

> If they, if they want it to just be a venting session, I'm here to listen. But if they, if they do need some additional follow-up I wanna make sure that they have the control as much as is appropriate . . . . Research again is not . . . your standard clinical treatment . . . . Our participants are volunteering to be in this research, that they're not compelled, and I think it's important that we respect and honor their contributions to the research. They can withdraw at any time and I, I guess it's just part of the respect of persons, kind of getting back to that ethical principle, um, in the Belmont Re[port that I think is, is important.

Another recurring theme throughout these interviews was the need to maintain continuous contact with participants and investigators, including informing them of the steps of the process as they occur. Informants viewed this communication channel as in part an extension of voice and the value of participation, as well as serving broader dignity goals; as one informant noted, "It's important to be transparent . . . . it usually turns out to be much worse if you don't keep the, the complainant in the loop so that they feel like they're actually being listened to . . . I think transparency and neutrality are more important because I'm not really sure there is such a thing as the right resolution." Some IRBs set frequencies for re-contacting subjects and investigators during a complaint, such as making contact on a weekly basis.

Informants also noted the need for transparency of process to improve satisfaction among both investigators and subjects. One informant, for example, described a change in practice to discontinue an informal process that was "never really clear on the policy" and "would cherry-pick what they wanted to do." In their new process, "if somebody had a complaint we would send an email and explain what our steps are going to be [to the subject] and a researcher if we were going to audit them . . . we tr[y] to be user-friendly and have clear understanding of what the role is and what's going to happen . . . and it's made the situation better." Another agreed that transparency directly affects perceived legitimacy: "Communication in really key . . . in order to be transparent . . . . I think even in the tough situations most people are respectful of how you undertake the process, knowing that it is a difficult process."

IRBs are aware that the stakes of complaints are high for investigators, particularly when subjects express concern about the investigators' own conduct or noncompliance. As one informant noted, researchers are "typically in a defensive stance" during complaints. Transparency of process was viewed as an important safeguard for investigators, who may also have more notice of IRB practices through investigator training, repeated interactions with the system, or access to internal institutional policy documents. Information about process can also alleviate investigators' feelings of being wrongly accused or the target of bias, as one informant noted: "We have to let them know that we have to investigate every single call regardless of feelings, regardless of anything . . . and a lot of times they know it's a process that we have to go through."

4. IRB as Mediator: Process Selection and Resolution of Minor Complaints

After the initial contact, the IRB director or manager makes a preliminary determination about the severity and likely veracity of the complaint.[117] Where there are urgent or emergency issues involving risks to subjects, the most senior IRB official (the chair) or a subcommittee of the IRB will immediately assemble and recommend emergency measures, such as suspending study activities. But for most types of minor complaints, the IRB personnel will begin by contacting the principal investigator of the research study by phone or email, to identify whether the complaint can be easily resolved. Many complaints are easily classified as minor issues that can be resolved via communication between the subject and principal investigator (e.g., missing compensation), or via a direct, second conversation with the individual (e.g., explaining why the person was not eligible for a particular protocol). The IRB director, manager, or chair typically takes these actions directly,[118] notifying the principal investigator or re-contacting the complainant to explain features of the study or informed consent form. A number of IRB chairs noted a practice of directly facilitating conversations between subjects and investigators, with the chair personally serving as a third-party mediator to ensure that the communication went smoothly.

> [I] try to set up a, you know, a meeting between them and the investigator so they can address these issues . . . . Most conflicts I think it's best when everybody is sitting down and talking to each other . . . . That's one of our first outreaches with any sort of problem, whether it's just an investigator or a study problem, is to

117 Several informants noted that concerns about veracity can be particularly important for complaints arising in psychiatric studies. "A lot of the complaints may also be from psychiatric patients . . . . So I sort of probe how closely their complaint is grounded in reality."

118 Where subjects report not having spoken with the investigator yet, many IRBs will suggest that the subject do so directly before proceeding.

> try to get everybody in the same room and talk about it. If it looks like it's a problem that could be solved by just people talking to each other or looking at what the different options are, that's always, that's always our first approach.

The time to resolution for these minor complaints is typically hours or days, and multiple informants described the procedure in these cases as a "customer service" approach, centered on listening and the subject's desire to be heard.

5. IRB as Fact-Finder: Iterative Investigation and Consultation for Serious Complaints

Where complaints do not arise from miscommunication or misunderstanding, however, the process escalates to resemble arbitration, in which the IRB takes on both a fact-finding and adjudicatory role and imposes a resolution that is enforced by institutional authority over the research protocol. The IRB chair, along with any other IRB personnel who initially received the complaint, makes a preliminary classification of the issues, rights, and individuals at stake, and determines whether other institutional actors should be involved in the resolution process. Where the IRB reports to an additional institutional authority, such as the vice president or chancellor for research, the IRB personnel will likely include this person in the decision about involving other departments.

Depending on the nature of the complaint, the IRB may choose to involve a wide array of offices or personnel within the institution. The role of these personnel is typically to provide guidance or to assist in fact-finding. These may include the institution's general counsel (for complaints that include legal claims, injuries, or potential legal violations, such as failures of informed consent or HIPAA violations), any insurance program for research-related injury, the human resources office for complaints involving whistleblowers or investigator misconduct, the risk management office, the regulatory affairs department, FERPA officials, the office for privacy and HIPAA, media affairs (for disputes receiving media attention), institutional officials serving as research subject advocates or patient advocates, university ombudsmen, campus police or security for disputes where subjects or investigators may threaten violence, institutional officials for sponsored projects, and departmental heads or chairs of the investigator's department. All dispute resolution processes for complaints alleging noncompliance with protocols will also involve a compliance team, which may be a subcommittee of the IRB, a single IRB officer such as a quality improvement officer, or a separate arm of the broader human research protection program.

The IRB chair and other institutional officials may also gauge whether the complaint requires contacts with people outside the university—for example, research sponsors who may need to approve protocol changes, local police who

were arresting participants leaving a study for sex workers, a state agency that had made a name-based registry of cancer patients available to researchers, a local school board for a dispute about informed consent for school-based research, or a tribal council for a dispute over the return of biological samples to tribe members. IRBs also work with foreign IRBs, for international research protocols that require review by institutions in multiple countries.

After identifying the relevant stakeholders, rights, and interests, the IRB typically begins a flexible and sometimes iterative process of fact-finding, consideration of facts and interests, and communication with the investigator, research staff, participant, and other institutional or outside actors. The fact-finding process may include a formal audit of study materials or less formal interviews with the investigator and study staff. The IRB may conduct this process itself through a subcommittee or individual staff members;[119] it may also use a compliance office or risk management team.[120] The process can last up to six months or even a year for complicated or contentious disputes, but more typically lasts about one month.

6. IRB as Client: Outsourcing Disputes

During consultation with other institutional stakeholders, senior members of the institution may decide to reallocate control of the dispute resolution process to legal counsel or human resources departments. Where this occurs, the IRB loses jurisdiction over the dispute. "[If] the institution wants to move forward with it or take it to a different level or address that we kind of bow out from a jurisdiction perspective."

Even when the IRB retains management of the dispute, however, they may rely on institutions' legal counsel for guidance, interpretation of applicable institutional policies or external regulations, or communication with research participants' counsel. Some informants believed that legal counsel were reliable supporters and valuable resources for most complaints. But others noted that legal counsel could actually complicate complaint resolution; their concern for institutional liability encourages defensive communication with subjects, rather than the empathy and concern that most IRBs thought was the necessary tone to achieve a resolution.

> We don't necessarily have to bring the attorneys in right from the beginning, and they don't drive the process . . . . They're focused,

119 One hospital also reported having institutional legal counsel attend fact-finding interviews, "to give [research staff] comfort and reassurance" that they will not be penalized for honest responses.

120 In one dispute concerning the behavior of the IRB itself, the IRB asked another institution's IRB to assist in the factfinding and dispute resolution process.

> of course, on protecting the university . . . and that's great. But that often is at odds with trying to resolve the participant complaint. In an ideal world, everybody would agree that resolving the complaint is not only the right thing to do but will prevent the litigation. But sometimes those are a little bit at odds and so we get into sort of a—if the attorneys are prominently involved—sort of a protective mode where um, we're not necessarily free to be as compassionate. Even if we're not agreeing with the participant necessarily, we want to be able to still interact with them in a way that displays empathy and compassion, and sometimes that can be a little bit of a challenge when the attorneys are involved.

A few research institutions had instituted a procedural innovation to address protocols that take place in international or culturally distinctive settings, where subjects may be uncomfortable with approaching the institution directly. These institutions sometimes required investigators to appoint a local community leader to assist in resolving disputes arising in any protocols; this person could liaise between the subject community and the institution where needed. The community representative was listed on informed consent forms and became a point of contact for receiving complaints, and also an active part of the resolution process for any complaints that rose to the level of the IRB.

> We look for an alternate, uh, position in the community, a trustworthy person in the community to accept those and refer them to us for handling . . . . It's all a part of being sensitive to the population that are being recruited . . . . It includes having a person in that community who would be perceived as being impartial and would listen and refer the, the problems and concerns to us . . . . It can be used in remote, anything that is remote from our site or which is culturally inaccessible, like an Indian tribe . . . . [And] for our sake they would be um, um at a leadership community leadership level that they would in an informed way communicate with the IRB here.

This institution raises interesting questions about the relationship between the research institution and the participant population.

7. IRB as Adjudicator: Deliberation, Decisions, and Appeals

When fact-finding is complete, IRBs proceed to deliberation, which remains internal to the IRB for most types of complaints. Factual findings and the results

of conversations with various stakeholders are recorded and assembled by the IRB, along with guidance from other relevant institutional actors. The IRB may designate a subcommittee or ask the full board to examine the factual findings, guidance, and interests at stake. This decision body recommends a preliminary solution that may be acceptable to the parties, including any proposed remedies or corrective action plans. Many IRBs at this stage will communicate directly with the principal investigator in advance of the final decision, attempting to find a voluntary set of protocol corrections or a remediation plan that the investigator would find feasible and acceptable. Several informants described this process as prioritizing transparency and participation throughout the crafting of a resolution, while others noted that unrealistic corrective action plans may undermine the durability of the resolution:

> We do try to be transparent, um, listen to both parties, and then come to collaborative solutions that would really involve all parties trying to create the solution . . . . My preference is not to impose solutions as much as to say, "What would be your solution given your particular environment that you conduct the study in?" . . . . Of course, if it's a regulatory piece then we have no flexibility, then we tend to impose, but even within that imposition it would be my style to say, "Well, how is that going to work for you?"
>
> We work together on a solution that's more of a learning experience. We don't want it to be punitive for either party . . . especially our PIs because sometimes . . . they didn't realize they were doing anything wrong . . . . So depending on the solution, a lot of times we may involve the PI into the solution.
>
> We don't want to impose . . . a bunch of strict regulations on a study team that will in essence make them be noncompliant in the future if they're unable to fulfill that corrective or preventive action plan.

The process concludes with a full IRB decision to approve a corrective action plan and to formally issue a written letter to the principal investigator, setting forth the facts and corrective action requirements. IRBs often notify the subject of the final resolution as well, although the subject does not typically receive a copy of the same letter. The IRB determines what will be disclosed to the participant at this time, which may be in writing or by phone, and may contain less detailed information. As one informant noted, "We may say [an investigator was] disciplined but we won't say . . . what the specific disciplinary action was

because . . . we have to keep in mind the faculty member and the investigator, their rights."

According to many informants, the IRB's authority to make binding decisions on research-related complaints arises from the federal regulations, which task IRBs with the approval or disapproval of research protocols. As one informant noted, "Because our IRB, you know . . . [we] have that federal regulatory mandate to be the final decision makers . . . even when people appeal [an adverse decision], it typically doesn't result in a significant change."

After the final decision, almost all institutions give the principal investigator a right to appeal for reconsideration by the chair, the full IRB, a vice president for research, or the chief medical officer. No institution described making this option available to the subject, because subjects cannot experience sanctions as a result of a complaint. But when prompted, many IRBs said that a subject who is dissatisfied with the resolution of their complaint could likely obtain reconsideration as well.

Subjects who invoke the dispute resolution process do not give up other legal remedies; nothing forecloses a public lawsuit after the process ends.[121] Informed consent forms do not require subjects to use the dispute resolution process at the institution—mandatory arbitration is curiously absent in this context. But because so many disputes are based on non-justiciable interests rather than legal rights, the IRB's decision is typically the only available remedy. Investigators can (and sometimes do) sue institutions in connection with research-related disputes, but individual subjects typically are not involved in public investigator-institution disputes.

8. IRB as Enforcer: Remedies

IRBs noted many options for remedying research-related injuries, all enforceable by the sanction of closing research protocols that do not comply with remediation plans. Financial settlements were possible but rare, and the negotiation of these settlements typically involved legal counsel. Only a small handful of institutions had a public policy of compensating research-related injuries, either by insurance or institutional funds; a majority, however, noted that they either paid for treating injuries at their own facilities, or they eventually provided funds for treating any research-related injuries that are not covered by subjects' own health insurance. This is an important informal policy, given that most consent forms specifically state that research sponsors and the institution are not obligated to pay for treating research-related injuries. Informants did not describe apologies as an available remedy, but noted that subjects did receive explanations of events where

121 It is possible, however, that subjects who receive compensation for injury do need to waive the ability to sue as a condition of settlement. *See* Pike article. But these are a minority of cases.

relevant.

Other remedies include changes to individual protocols, such as mandatory changes to training and supervision procedures for research staff, changes in recruitment strategies, changes to the informed consent process, or changes in criteria for initial or continued eligibility. Some of these protocol changes are reportable to study sponsors, as are complaints that are determined to arise from serious adverse events or unanticipated problems involving risks to subjects. IRBs can also require training or directed education of investigators or staff on issues like informed consent or record-keeping. For more severe or irremediable violations, IRBs can terminate studies or entire lines of research, mandate the destruction or nonuse of data, or require the return of biological samples to subjects. Where investigations reveal serious or recurring noncompliance, scientific misconduct, or HR violations, researchers may also experience professional sanctions or discipline through the HR department.

Some complaints led to thoroughgoing changes in institutional policies, such as the discontinuation of recruitment practices that involve cold-calling, changed policies for the supervision of students, a discontinuation of studies that consented participants under the influence of alcohol, new policies for training researchers and staff, and changes to institution-wide informed consent practices.

9. IRB as Record-Keeper: Missed Opportunities

Most institutions kept written records of complaints, but these were typically filed under individual protocols; only a few institutions systematically recorded complaints using a method that would allow for analysis over time or across protocols. Feedback from dispute resolution programs could assist IRBs in identifying research risks and burdens, but IRBs are neglecting this opportunity to use disputes as information. Ideally, IRBs should record complaints in a manner that would allow personnel to aggregate or compare issues across protocols. A periodic analysis of these complaint data could help IRBs anticipate risks and burdens at the protocol approval stage, rather than waiting for complaints to arise. IRBs could also use these data to identify recurring complaints arising from particular departments or protocol types, which could be remedied by improvements in investigator training or institutional research procedures.

E. The Centrality and Limitations of Procedural Flexibility in IDR

Throughout the interviews, the IRB informants consistently noted the advantages of a highly flexible complaint resolution process.[122] Even where

122 The informality of IRBs' own protocols, records, and procedures may amuse many researchers who prepare highly detailed and inflexible protocols to comply with IRB requirements.

procedures were written, informants described leaving broad latitude to select among process options, or supplementing the written process to include additional elements.

> We have to have a written policy that we handle complaints, but we leave it as open as we possibly can, um, we provide a range of possible responses depending on what the, you know, the level of severity, etc . . . . You don't want to lock yourself into having to, you know, you don't want to say in your policy we will respond in writing to all complaints if that's not appropriate . . . . So you then leave yourself open to being able to, um, respond in a, you know, um, issue specific manner that's appropriate for what's going on.

> [The process is] just based on the situation at hand . . . . We have on paper a policy and process . . . . But if we, you know, run into an obstacle or a snag or, you know, if we needed additional information, we might make a decision that's not written somewhere. But again, only with the same intent, which is . . . [that] all parties are being, you know, properly addressed, you know, properly, um, given the proper opportunity to kind of speak.

Informants believed that the principal benefits of procedural flexibility were the opportunity to tailor the process for complaints with a range of rights and interests; to involve all relevant institutional and outside personnel in the response; and to provide full voice to any unforeseeable parties that may have a stake in the events or their resolution. Some of these benefits serve efficiency—that is, standardized procedural features may waste time and resources. For example, it is costly in time and manpower to conduct full audits for complaints that might be easily resolved through facilitated negotiation. These efficiency benefits may indirectly serve the value of participation, by freeing up time and attention for more resource-intensive complaints. As informants described it, however, procedural flexibility also directly serves participation and legitimacy by promoting voice and inclusion of all parties. Informants believed it would be costly to legitimacy, destructive to community relations, or corrosive to the durability of a resolution if processes exclude stakeholders beyond the subject and investigators—as in the examples involving tribal leadership, community-based organization, school boards, or local trusted officials in overseas protocols. The procedural flexibility embedded in these ADR systems allow for the involvement of all relevant stakeholders on a case-by-case basis, which informants commonly described as a

---

Flexibility, however, serves several key values in the IDR process, as this Section suggests.

procedural goal. As one informant noted:

> Each situation almost is unique . . . . And the biggest principle that we try to follow that's sort of a general principle . . . is to spend a lot of time being absolutely certain that we have consulted with all the appropriate parties . . . . And that means at information gathering, identification of an appropriate, uh, resolution and action plan, and then conducting and carrying out that action plan, and then closing the loop when the whole thing is done. So that's kind of the general principle that we do that is common to all the complaints . . . because we've had some real problems when that didn't happen.

Many informants stressed that a standardized process would be inadequate to handle complaints, and some believed that the interpersonal skills of the dispute *processors* are likely more important for a thorough resolution, compared to the process elements itself.

> These complaints are as variable as there are people . . . . I'm wondering if you could or whoever develops this could come up with enough of a cookbook or a recipe, um, that it's going to be applicable to the next five cases that came in the door . . . . Some of it depends on who you have handling [complaints], just how adept they are at dealing with people, um, more than processes . . . . I don't know that this is going to be an area that just immediately lends itself to here's, here's, here's the one template or recipe you can all follow and apply this to every complaint you get

But despite the virtues of procedural flexibility, informants also noted that a flexible process introduced complexity, unpredictability, bias, and difficulty in passing on institutional knowledge. Informants noted that flexibility may lead to a lack of transparency and inconsistency.

> I think a strength is our flexibility or the nuances. I mean, I enjoy the autonomy to handle these things in the, uh, in a professional expeditious manner, as I see fit given the nature of them. But I also see, particularly as like a noncompliance gets tied in with this, the fact that we don't have, if you will, very transparent, codified, step by step procedures that we follow every single time can bite us.
>
> The flexibility is the upside and it's the downside . . . . It means

> that I am making decisions . . . . And that's my job . . . . But I have to decide, you know, pretty quickly what the correct response is and who to contact and where to go with it . . . . So having that, um, in the hands of a, a single individual . . . almost always it's, it's a single individual who's handling it and that, I think, might be, that could be a problem.

Informants also noted that the embedded discretion for IRBs to select among processes can also make it difficult to train successors in the process more generally, which could lead to inconsistency over time.

> One doesn't really know if there's a right or wrong way of dealing with this. You just do whatever makes sense for the participant . . . . [There's] a lot of flexibility. And a lot of discretion. Uh, it's up to the discretion of, um, me for the most part. That's -- that's the problem . . . . [It's] not impossible [to train someone else]. The challenge is that, um, it's a subjective process that depends on my view of what's going on initially . . . . it would be difficult to document, if necessary the triage process, because that is based on, largely on subjectivity and intuition and a lot of intangible characteristics.

Informants therefore viewed the deliberate exercise of procedural flexibility as a means of serving participation and legitimacy values, as well as the more proximate goal of system efficiency. But flexibility was not an unabated good, and it complicated values such as procedural predictability and equality, as well as the proximate goals of consistency and system transparency. The following section will consider informants' appraisals of process goals and values more generally.

IV. APPRAISING IRB-MANAGED IDR SYSTEMS

Apart from asking informants to describe their procedures, the interviews also asked informants to appraise the strengths and weaknesses of their complaint resolution systems. Informants identified a number of strengths, including the contributions of procedural flexibility. But informants also noted problems from their perspectives, including concerns about low uptake, the capacity of the IRB to act as a third-party neutral, frustration with available resources, and the potential for inconsistency across participants or time.

This Article now moves from a descriptive to normative view to provide a critical appraisal of IRBs' IDR systems. Strengths include the ability of these processes to consider both rights and interests, as well as the voluntary nature of participation and the continued access to litigation where participants choose to

file claims. But among system weaknesses, I echo some of the informants' concerns, focusing more specifically on participant non-consultation, low uptake, and IRBs' institutional capacities to behave neutrally and skillfully in the dispute resolution role. This Part will first describe informants' appraisals, followed by my own.

A. Informants' Appraisals

Beyond the strengths afforded by procedural flexibility, informants described many other advantages of their complaint resolution processes. The institutions that compensated subjects who sustained research-related injuries—either by institutional insurance or by de facto provision of medical treatment—viewed the availability of a financial remedy as a particular strength. (In contrast, institutions with "fuzzy" language on injuries or policies of nonpayment were a source of great frustration to informants, who would prefer to have the option to make subjects whole for physical injuries.) Many informants noted that their process functions well to give both subjects and investigators the opportunity to be heard and respected, and those with a written or standardized process were more likely to describe consistency and transparency as system strengths. The personal qualities of individuals involved in the process—such as substantive knowledge of the regulations, experience handling complaints and investigators, personal experience in the investigator role, interpersonal or counseling skills, and (sometimes) dispute resolution training—were also viewed as strengths. Informants appraised decision quality in terms of accuracy about facts, finality and non-recurrence of the dispute, parties' satisfaction, and the ability to enact system-level change for disputes that indicate a systemic problem. Most institutions believed their processes functioned well on these measures, and believed that they had struck the best possible balance between protecting participants, treating investigators fairly, and safeguarding the interests of the institution.

1. Access and Uptake

Despite the perceived strengths of their processes, informants believed the frequency of complaints was surprising low, and many were puzzled by the shortage. As one informant said, "*I've always felt that the number of complaints we get is remarkably small for the size of our research operation . . . . The information [about our IRB] is really prominent in our consent forms, but . . . it just seems odd to me, um, that we don't have more.*" Informants who sought to explain this shortage of complaints offered different explanations for the scarcity. Some noted that research staff are likely the first port of call for a subject complaint, and these informants emphasized the need for IRBs to train investigators and staff to respond thoroughly to subject concerns. Several

institutions that primarily conducted social and behavioral research suggested that complaints are infrequent because their research portfolio tended to be minimal risk, or excluded clinical trials. One research hospital informant noted that complaints are likely low because all hospital patients know they are receiving research-related services, giving them a different set of expectations about their care. Others suggested that participants enrolled in therapeutic research are less likely to complain, compared to healthy individuals who participate in research for financial reasons and may have more complaints related to compensation.

But as interviews continued, many informants suggested that research participants may be unable or unwilling to call the IRBs with complaints. Participants may not understand research protocols, making it difficult for them to form expectations – and thus, difficult to identify when they have experienced a wrong. Even if participants are aware that the IRB provides a venue for dispute resolution, they may be fearful of the consequences of complaining. Subjects enrolled in ongoing protocols or clinical care may also fear retaliation or stigma after lodging a complaint.

> It's probably the tip of the iceberg underneath that one [complaint] in two years is people that were frustrated and wanted to complain but they talked themselves out of it . . . I think there's some stigma attached to, um, calling up somebody that works for the university . . . I think the person would be uncomfortable to call the university.

> We have a low number, and I'd like to think that's because everyone's so excellent at what they do . . . . [But] I fear that sometimes there's people that might want to share something or talk through something, and they don't share because . . . [they] are also patients . . . and the research study might even be headed by the person who also provides their clinical care . . . . We definitely try to set up a system of being anonymous and we keep them separated from the investigator and all that good stuff, but even with all those protections I feel people might hesitate to say, or they might not even be sure what to complain about. You know what I mean, they're not always 100% positive of how a consent process should really be executed. Did they have enough time to think through it and ask their questions? They might not even feel confident, if they've had a bad experience, that they had a bad experience. I'm always very surprised that we have the small number that we do.

> Sometimes the researcher is also their physician that they have

> known for years and maybe the complaint is about some aspect of the study, but they don't want to sour the relationship that they've had with a certain specialist or something like that.

Finally, informants also noted that subjects may be uncertain about the process for dispute resolution, and this uncertainty may make the process inaccessible. Although the consent forms consistently directed participants to the IRBs, informants expressed concern that this information was not prominent or clear enough to empower subjects to use the system.

> I'm sort of surprised that more people don't call us or ask questions . . . I just think people don't necessarily think to call us, you know? . . . . I've often thought maybe we should, it would be interesting to do a study about putting the IRB's phone number first on the consent form to see if we got more calls. Because I think with that many protocols . . . I think we'd have more calls.

> I think that people probably don't report it enough, and I don't know if that has to do with, maybe perception of research compliance, or if our participants really are just not aware that they can report . . . . I definitely think that there has been . . . some instances where a student or participant could complaint, but they just don't . . . because they just brush it off, or because they, you know, are really not aware of the procedure, or if they just don't understand the importance of reporting.

Some may argue that low uptake of a complaint resolution is appropriate for research-related complaints; in a setting where many complaints may entail non-justiciable or minor harms, lumping the complaint or exiting the relationship[123] may be more efficient for many subjects and institutions. Institutions certainly benefit from the comparatively low administrative costs of a seldom-used complaint procedure. But the low frequency of complaints may be problematic in this context for several instrumental reasons, even without considering inherent value of dispute resolution for subjects. First, silence on minor complaints obscure systemic problems that eventually expose institutions to significant legal risks, such as deficiencies in informed consent procedures. Second, dissatisfied subjects who feel they must lump their disputes can contribute to difficult relationships between institutions and their surrounding communities, which can spill over into other conflicts. Third, when subjects choose to exit scientific research or decline

123 Jost, *supra* note 29, at 314; William L.F. Felstiner, *Influences of Social Organization on Dispute Processing*, 9 L. & SOC'Y REV. 63, 81 (1974).

to reenroll in future protocols, the institution must divert more resources to study recruitment and retention, thus increasing the costs of research and reducing the feasibility of human subjects protocols. The disproportionately low frequency of complaints, therefore, may not be fully in institutions' best interests at present.

Fully explaining the low uptake of institutions' dispute resolution processes requires more research with participants themselves, in order to explore perceptions of research experiences that give rise to complaints, their awareness of the availability and content of a complaint resolution process, and their expectations and perceptions of these institutionally controlled ADR systems. But my research with the designers and implementers of IRBs' processes suggests that research subjects do not receive sufficient information to make the complaint resolution process accessible—perhaps because they do not understand or believe that they have grounds to complain, because they are unaware of the forum, or because they are unaware of the procedural safeguards the forum provides. And moreover, even if participants are aware of mistreatment and the venue for complaint resolution, they may nonetheless be deterred by fears that complaining will result in stigma, retaliation, deterioration of relationships with care providers, or loss of access to services. The low uptake of these processes suggests that many subjects do not currently view them as meaningful options for complaint resolution.

2. Neutrality

Despite agreeing that IRBs had authority to resolve disputes, some informants expressed discomfort with placing the IRB in the role of a neutral third party. The most visible stakeholders in complaints are the subject and the investigator, but complaints also implicate the institution, the broader communities of which subjects are a part, and the legitimacy and progress of science as a greater social good. The federal regulations task IRBs with protecting subject welfare, and some informants suggested that this biased their judgments to favor subjects. As one noted, "*Because of the way the staff then would view their roles here . . . they're more participant, uh, oriented. And I always just have to point out to them . . . you need to give the investigator an equal chance.*" Another concurred: "I do think we need to remain neutral though in before until we get all of the facts . . . . But our end and ultimate goal is to protect the rights um of the participant to make sure they are treated correctly." Some informants even suggested that placing participants first was the best way to serve institutional interests: "I have to follow the regulations to protect the institution, as well as to protect the participant . . . in that way they're kind of woven together . . . . follow the regulations, be accurate, and honor the subject's complaint."

But some informants also noted that ties to investigators and institutions can

complicate these loyalties. As one reflected, "You're here as an IRB staff. You need to work for the subject. You you're protecting the subject, not the PI . . . . But the PI is a colleague . . . . So you need to have balance between both discussions." As institutional dispute resolution scholars would note, IRBs are institutional arms, staffed by institutional employees, and IRB professionals are aware of their role in protecting their institution throughout the complaint resolution process. As one informant noted, "I think [neutrality] is important, but I think it's very difficult to achieve . . . for us to be impartial . . . . I do think we're biased toward the institution because of our employment status." Or as another noted, "The first, you know, line of protection needs to be the participants but . . . as university officials there's a, a responsibility to the university as well." The burden of neutrality and pressure from the institution can make these dispute resolution processes highly stressful for IRB personnel, as one informant described:

> When our office has to engage in a very kind of intense uh investigation and follow up for a complaint . . . it's pretty stressful on our resources and on our personnel. There have definitely been times when we have uh feared for our safety because an investigator feels their um their career is on the line, and when the institution feels that you know their reputation is on the line. And [when] we're trying to pursue um you know an investigation that may have some implications for the institution . . . we might feel our job is in jeopardy . . . . It's personally very stressful . . . . We've been . . . trying to understand the reasons for burnout and turnover . . . in our office. And any compliance office I think, um, has similar issues because it's just the nature of this kind of work, compliance work . . . our turnover is pretty high . . . . Our biggest weakness is dealing with institutional oversight, and kind of being able to make our determinations in an autonomous way.

These concerns did not arise in all institutions; some informants reported little difficulty viewing their role as a third-party neutral. As one informant said, "I'm not representing or defending the role of the investigator or any institution, that I'm neutral because our goal . . . our goal is human subject protection and that [resolving disputes is] part, it's part of it so [I'm] definitely neutral." But it is important to note that neutrality may not be perfectly secured through an internal process, and IRBs are aware of these tensions.

3. Resources and Training

In part due to the rarity of participant complaints, many informants noted that they had not received extensive training or professional development to handle

disputes directly. A small minority of informants had completed complaint resolution or mediation training, but they had done so for other purposes, such as institution-wide HR initiatives or training for previous employment. Although many informants noted that they felt comfortable handling most subject complaints due to their institutional mandate to protect participants, they also reported uncertainty about how to manage complaints that may involve mental illness, threats of violence, and volatile interpersonal dynamics. When asked what resources could improve their processes, informants were most likely to mention the need for dispute resolution skills building, mediation training, or counseling training throughout the IRB office.

Informants sometimes noted struggling with the manpower and time needed to handle complex complaints, particularly given other IRB functions such as initial and ongoing protocol review. Multiple institutions also reported difficulties documenting complaints in a helpful way, and as noted, most did not document complaints in a manner that would allow for systematic analysis over time. Again, many described this as the result of rare complaints, since there may not be enough for a helpful analysis of systemic problems. As one informant noted, "I would be interested in a little more formal feedback loop . . . if we had data that would show if . . . there's a lot of complaints in a certain area then we could increase, redirect our education program . . . . It would be, you know, allocation of resources to prevent [problems]."

Informants also reported having little or no information about other institutions' processes, making it difficult to appraise and improve their systems. This arises in part from the nonpublic nature of these ADR systems, but also from a general lack of professional attention because complaints are currently rare. Many suggested that PRIM&R, the organization for IRB professionals, could build capacity by focusing on this issue in annual conferences or continuing education, such as providing case studies or an aggregation of best practices across institutions.

4. Consistency and Monitoring

As the previous section noted, some informants expressed concerns about consistency and predictability. In large part, this reflected the procedural flexibility that they viewed as essential to achieving participation and legitimacy goals. But many also suggested that the rareness of complaints may undermine consistency, since the procedures are not invoked often enough to become routine: "I know that we can all improve our processes. It's one of those areas that we don't see a lot of them . . . since it's infrequent and it's, each case is individual, it's hard to come up with, you know, systematic processes."

Some also noted that it was difficult to gauge whether their processes were in

fact consistent or successful, because they did not have enough complaints to assess how the system functioned as a whole. "[The process] hasn't really been tested . . . . with all our policies, even in writing, they were in draft form for quite a while. You really don't know, have you covered everything, until . . . the scenario arises and you pull the policy and you're ready to walk those steps out . . . You never know the holes until you find them." Institutions with larger research portfolios with a larger absolute number of complaints are less likely to have this problem, but informants from such institutions still noted difficulties with documenting complaints in a way that allows them to monitor for consistency and systemic problems.

B. A Critical Appraisal of IDR Processes

Taken as a whole, this study has revealed a set of institutional dispute resolution systems with broad procedural flexibility, institutional discretion, and management by institutional employees who perceive an ethical and regulatory imperative to protect subjects—but who also note conflicting loyalties to investigators and the institution as a whole. The system design typically matches the priority that informants placed on values of participation, revelation, and privacy; subjects and investigators have a full opportunity to communicate facts, these parties have some opportunities to shape the process and remedy, the system accommodates both justiciable and non-justiciable claims, decisions are reasoned and almost always written, decisions are enforceable within the scope of IRBs' regulatory authority, and the systems aim for party satisfaction as a primary proximate goal. To the extent that participation directly shapes party acceptance of the system, the processes serve legitimacy values as well, both for parties and the broader project of scientific advancement.

In relation to a recent framework of preferred design elements for ADR systems,[124] these systems also have several key strengths: they offer multiple process options (e.g., facilitated negotiation, quasi-arbitration), and accommodate both interests and legal rights. They provide flexibility for complaining subjects to have input on the process, although the processes made little distinction between rights and interests. Participation is voluntary and confidential for subjects (although less voluntary for investigators, who are subject to IRB authority), and the system aimed for transparency of process while parties were engaged in the dispute. Parties may also pursue litigation even after the conclusion of these IDR processes, in most cases.

Despite these advantages, this case study also reveals several key deficiencies of the systems. This Section will consider three problems in particular: (1) lack of participant input on system design; (2) potential underutilization; and (3)

124 Smith & Martinez, *supra* note 27, at 128.

challenges to IRB neutrality and resources for dispute resolution. This Part will conclude with a set of recommendations to improve on existing practices.

1. Exclusion of Participants from System Design

The origins of IRBs' IDR systems are largely stories of "muddling through."[125] Across all institutions, IDR processes arose informally as a set of departmental practices when IRBs responded to unexpected complaints, and those practices were responsive to institutional resources and IRBs' perceived role. At some institutions, practices for complaint resolution remain informal, and even unwritten. Other institutions have codified their practices, but most did not do so until prompted by the AAHRPP accreditation process. Where IRBs consulted external resources during process development, they were likely to ask other IRBs for guidance, rather than developing a new process with input from institutional and external stakeholders. IRBs typically described small modifications over time in response to institutional constraints and learning, but few to none had undertaken a wholesale examination of their complaint resolution practices. As noted above, AAHRPP requires a written policy for the resolution of complaints, but does not set requirements for how these systems are designed and operated.

In light of these origins, all the IDR systems in this Article were uniformly designed without the input of participant representatives. Literature on dispute system design emphasizes the importance of involving all stakeholders—all those who are "affected either by the problem/conflict or by a potential solution."[126] This can allow dispute system designers to account for parties' interests in process design, and to build in elements of procedural justice from the earliest opportunity.[127] The informal nature of procedure development clarifies why this has not happened, but it is plausible, ethical, and practical for IRBs to remedy the issue when there is an opportunity to reconsider their current policies.

Two factors may mitigate the exclusion of participants from the development of these IDR processes, but these are incomplete remedies for non-consultation. First, some might classify the IRB itself as a participant representative—it is, after all, bound to ensure the protection of research subjects. But IRBs are composed of members who are dissimilar, in most ways, from research participants. Per the Common Rule, IRBs must include at least five members "with varying backgrounds," with efforts made to avoid discrimination by race and gender, and must include at least one scientist, one nonscientist, someone from outside the

125 *See* Charles E. Lindblom, *The Science of "Muddling Through,"* 19 PUB. ADMIN. REV. 79 (1959).

126 NANCY H. RODGERS, ROBERT C. BORDONE, FRANK E.A. SANDER & CRAIG A. MCEWEN, DESIGNING SYSTEMS AND PROCESSES FOR MANAGING DISPUTES 72 (2013).

127 *Id.* at 75.

institution, and someone knowledgeable about applicable laws and standards of professional practice.[128] IRBs reviewing research with vulnerable populations (e.g., children, pregnant women, prisoners, people with mental disabilities) must also include individuals who are "knowledgeable" and "experienced" in working with these groups.[129] Experience in working with subjects, however, does not mean that IRBs understand how participants may experience research complaints, nor how they would prefer to seek redress at the institution. Moreover, many IDR procedures have developed within IRB administrative offices, rather than being considered by the full IRB.

Secondly, IRBs give participants some say over procedural options, such as electing anonymity, choosing between mediation or an arbitration-like process, or bringing disputes to a trusted local authority for protocols that have provided that choice. Giving participants choices at the time of the dispute alleviates the problem of non-consultation at the outset. But participant feedback is nonetheless important at the time of system design. Having a say in process development is important in part as a matter of procedural justice, but also as a matter of improving system accessibility, the durability of resolutions, and perceived legitimacy of the process (and the research institution more generally).

Consulting participant groups is daunting and complex. Institutions have enormous research portfolios, and it is impossible to consult a representative from every participant constituency. Research changes over time, and current participants may not be well-placed to represent future participants' needs. The difficulty of incorporating participant perspectives may be one reason why these views are so frequently omitted from general discussions of research ethics.[130] Part V will consider potential pragmatic strategies for soliciting participants' views of the dispute resolution system, as well as outcomes that IRBs should consider in evaluating whether system changes have led to improvement.[131]

2. Process Underutilization

It is difficult to know what an "optimal" number of participant complaints may be. We do not know the frequency of actual or experienced misconduct in research, nor do we know the frequency of physical injury. Moreover, we do not know the number of complaints that participants would deem sufficiently serious to seek resolution, rather than lumping or dismissing the problem. Of this number, we also do not know how many complaints are already addressed by investigators and their staffs, without escalating to the level of an IRB report. If the number of

128 45 C.F.R. § 46.107 (2018).
129 *Id.*
130 DRESSER, *supra* note 11.
131 See Section V.A.

complaints made to IRBs rose sharply, it may be practically impossible for existing institutions to resolve each complaint with the full complement of processes described here—intake, consultation, fact-finding, deliberation, decision, and appeal. Substituting an abbreviated process for the sake of inefficiency could disadvantage complainants with more complex grievances; at the other end of the spectrum, scaling up dispute resolution resources to handle large numbers of complaints may divert resources that are currently used for other ends, such as medical treatment or research expenses. Without knowing the number of complaints that participants may have in aggregate—including those never brought to the IRB's attention—it is difficult if not impossible to measure important system outcomes such as participant access and uptake.

It is possible to argue that the number of complaints currently received by IRB dispute resolution systems is in fact optimal. But almost all the informants in this study believed that their processes were underutilized. Prior research on participant comprehension of research protocols at the time of informed consent suggests that there are frequent disparities between participants' expectations and the reality of clinical trials.[132] For example, research on the "therapeutic misconception" and "preventive misconception" shows that as much as 62% of participants may be expected to believe that medications are effective or have "unrealistic beliefs" about the likelihood of benefit, when those drugs are in fact unproven.[133] This is one example of experiences that may not match expectations; many other surprises and misadventures are possible. The numbers in Table 2 may also give us pause to reconsider utilization; a median complaint frequency of 2.2 per 1,000 protocols (which enroll far more than 1,000 subjects!) seems far lower than what might be expected.

Considering these facts, it is reasonable to believe that utilization of these IDR programs is low. Although low uptake may be immediately advantageous for institutions with limited human resources on their IRBs, leaving research-related disputes unresolved can expose research institutions to adverse consequences such as future litigation, future media exposure, poor reputation, and increased costs of future research.

Some of the causes of low system uptake may be difficult to remedy in health care systems that merge therapeutic research with clinical care. Subjects may not wish to jeopardize their care relationships by complaining about studies conducted

132 *See* Flory & Emanuel, *supra* note 7, at 1593 (citing studies, including one showing that 30% of participants in cancer trials believed that they were receiving a treatment already proven to be the best for their cancer).

133 Paul Appelbaum et al., *Therapeutic Misconception in Clinical Research: Frequency and Risk Factors*, 26 IRB: ETHICS & HUM. RES. 1 (2004); Charles W. Lidz et al., *Therapeutic Misconception and the Appreciation of Risks in Clinical Trials*, 58 SOC. SCI. & MED. 1689 (2004); Alan E. Simon et al., *Preventive Misconception: Its Nature, Presence, and Ethical Implications for Research*, 32 AM. J. PREV. MED. 370 (2007).

by their own clinicians. Subjects in all institutions and all types of protocols may also be skeptical of the neutrality of any forum offered by the institution, including the IRB itself, and past research abuses have created a legacy of institutional mistrust in many communities. The dispute resolution systems in this case study were designed exclusively by the institutions, and although subjects could select their desired level of involvement in the process, the institutions did not consult subjects or subject groups during the initial design stage. These barriers may persist regardless of dispute system design, even with an external third-party neutral and advance notice of procedural protections such as the ability to remain anonymous.

But low uptake also reflects a lack of information, particularly lack of awareness of the forum and the process for dispute resolution, and systems can seek to remedy these problems by better educating subjects during study enrollment and follow-up. Subjects' awareness and understanding of protocols and "subjects' rights"—and thus, their expectations of how they should be treated—will inform whether they recognize wrongs as actionable. More effective education about protocol design and clear enunciation of other interests—such as a right to be treated with dignity during the study, or a right to voice concerns about study processes—may help. The low uptake almost certainly reflects low subject awareness of IRB oversight, authority over studies, and availability to resolve subject complaints.

Where subjects *do* understand that a forum exists for the resolution of their complaints, they currently have no way of knowing what will happen when they contact that forum. IRBs do not provide advance notice of procedural protections such as anonymity or confidentiality, nor are subjects aware of how the IRB will proceed to address their concerns. Because procedures are so flexible, written processes may be imprecise or absent, and they are not made available to potential subjects in detail. Subjects do not know in advance, for example, that facilitated negotiation is available, that the IRB makes decisions independent of the research team, or that complaints can sometimes lead to changes in institutional policies that may benefit future subjects. A lack of information about the *process*, which in part derives from broad procedural flexibility, may undermine predictability and subjects' perception of control over their complaints.

3. IRB Neutrality and Capacity

As noted throughout this Article, IRBs have several interests that come into conflict when they manage research-related disputes. IRBs are required to prioritize subject welfare (which may disadvantage researchers); they are colleagues of researchers who are repeat players in IRB review (which may disadvantage participants); and they are also members of the institution and aware of institutional interests. Furthermore, IRBs who oversee disputes are also the very

institutional representatives who initially approved study protocols to proceed. If disputes escalate to litigation, IRBs themselves may be liable for negligent protocol approval and oversight,[134] giving them a direct stake in resolving disputes quickly and with minimal institutional exposure. A participant complaint about study procedures may also be viewed as a challenge to IRBs' original determination that the procedures were ethical, which asks IRBs to revisit these initial judgments at the moment of the complaint. This could compromise equality and accountability, despite IRBs' regulatory role and sincere commitment to the interests of the subject. A long history of scholarship vacillates between two poles: some characterize IRBs as intrusive and stifling to researchers,[135] while others have viewed IRBs as insufficiently protective, overworked, and vulnerable to capture by researchers.[136] From the view of IRB personnel themselves, this study suggests sincere efforts at neutrality, but informants acknowledged that multiple interests—and the salience of institutional interests in particular—made this challenging.

The lack of neutrality of a third-party decision-maker can be inimical to all process values in dispute resolution,[137] including participation, accountability, and legitimacy. Participants skeptical of neutrality may decline to use IDR processes, or they may disengage if their experience with the process does not fulfill their expectations of fairness. Neutrality problems can also impair accountability if the decision-maker favors one disputing party, either due to conscious or unconscious bias. A lack of neutrality can also impair legitimacy, if disputing parties do not accept the process or the outcome as fair; this can challenge the durability of resolutions and lead to more disengagement from the process over time. Importantly, however, although these are potential problems, we do not have evidence yet that they are occurring. The study in this Article conducted interviews with IRBs themselves, rather than disputing parties. The broader literature on complaints in human subjects research is also thin, and although there are many records of researcher discontent with IRB decisions (particularly on protocol approval and disapprovals), there is little evidence specific to the participant complaint context.

There are also compelling advantages to using IRBs to manage research-related disputes. IRBs have enforceable authority to suspend research protocols, to require revisions or remedies internal to research protocols, or to cancel protocols

134 Mello et al., *supra* note 19.

135 Philip Hamburger, *The New Censorship: Institutional Review Boards*, 2004 SUP. CT. REV. 271 (2004); Philip Hamburger, *Getting Permission*, 101 NW. U.L. REV. 405 (2007).

136 Hazel Glenn Beh, *The Role of Institutional Review Boards in Protecting Human Subjects: Are We Really Ready To Fix a Broken System?*, 26 L. & PSYCHOL. REV. 1 (2002); Donna Shalala, *Protecting Research Subjects – What Must Be Done*, 343 N. ENG. J. MED. 808 (2000); Ezekiel J. Emanuel et al., *Oversight of Human Participants Research: Identifying Problems to Evaluate Reform Proposals*, 141 ANNALS INTERNAL MED. 282 (2004).

137 Redish & Marshall, *supra* note 37.

entirely, IRBs already have the scientific expertise to understand protocols and potential deviations, and they are familiar with each of the protocols from which disputes arise. IRBs' regulatory role may partially mitigate the lack of neutrality from the participant perspective (although not from the researcher perspective). IRBs within the institution can quickly mobilize other institutional actors, such as department chairs, legal counsel, human resources, and compliance departments that may assist in fact-finding. Moreover, there institutional role as the guardian of participant welfare means that IRBs should be involved, somehow, in any IDR process for research-related disputes. In light of the low frequency of complaints, institutions may also find it inefficient to invest in a separate IDR process for research-related disputes.

The balance of advantages and disadvantages shifted somewhat in multi-site studies under the 2018 revisions to the Common Rule, which requires that multi-site studies use a single IRB of record.[138] For these studies, the IRB that approved the study may be at a different institution from where the complaint arises. Presumably, these studies could refer complaints either to the local IRB at their site, or to the IRB of record. Local IRB may be somewhat less familiar with study procedures, but they may also have less concern for their own interest (in the event that the dispute escalates to litigation involving the approving IRB). Referring all complaints to the IRB of record presents other advantages, such as familiarity with the protocol and potentially less concern about liability of their own research institution. The revised Common Rule does not specify how complaints or injuries arising from such study should be resolved, leaving this an open question.[139]

Without evidence of current harm, and given the structural advantages of using IRBs for resolving research-related complaints, it is sensible to leave these dispute resolution processes within the IRB. But this raises questions of institutional support and IRB training for dispute resolution tasks. Informants in this study described burdens in implementing the IDR process, including substantial human resources, time necessary for deliberation on both process and outcome, emotional strain and fatigue, and a lack of skills training in relevant areas such as mediation or conflict resolution. IRBs are already (and have long been) overtaxed in time and resources, and they navigate an increasingly complex set of federal, state, and institutional policies. Particularly if the number of complaints were to increase, IRBs currently lack some expertise and resources needed for an effective response to complex or emotionally fraught complaints.

This discussion raises the question of IRBs' capacity and motivation to make changes to their IDR systems. To that end, IRBs have some advantages that make

138 82 Fed. Reg. 7149 (Jan. 19, 2017); 45 C.F.R. § 46.114(b)(1) (2018).

139 Stark & Greene, *Clinical Trials*, *supra* note 72 (noting that centralized IRB review raises questions about allocating institutional liability).

them well-positioned to improve these processes. Human research protection programs are fairly small and self-contained within their institutions, and they have a great deal of discretion over their internal procedures and their interpretations of federal regulations. IRBs or the heads of human research protection programs often report directly to institutional presidents or vice presidents for research, and IRBs' independent federal mandate to protect research participants gives them a separate source of authority to make changes that they deem necessary for that goal. IRB chairs and administrative staff are extremely well educated, as noted in this study, and they are attentive to their federal mandate, as this study has suggested. The informants in this study often expressed the motivation to improve their processes, including asking about other institutions' best practices, and many noted that this was the first time they had the opportunity to reflect on this institutional function. In their institutional capacity, moreover, these informants had power to make or credibly suggest changes to existing policies. It appears, therefore, that there would be high capacity and perhaps high motivation to change these systems given awareness of the need. But to date, IRBs have experienced a low frequency of complaints, creating few opportunities to reconsider their processes or to evaluate their effectiveness. IRBs also may lack the time, financial resources, and manpower to study this issue or to make resource-intensive changes. Some of the suggestions below, such as compensating injured participants, may be beyond the power of the IRB, and more properly suggested to institutional presidents or general counsels. But where changes are inexpensive and fairly straightforward, there is good reason for optimism about IRBs' capacity and motivation to improve their IDR processes.

V. IMPROVING IDR FOR RESEARCH-RELATED INJURIES

The previous Part describes a number of drawbacks of current IDR processes for resolving research-related disputes. This Part will conclude with recommendations for improving the functioning and fairness of these dispute resolution systems.

As noted above, I stop short of recommending that these IDR systems be relocated outside the IRB. To be sure, the Department of Health and Human Services and the FDA could *require* the use of a neutral third-party mediator or arbitrator through the federal regulations governing human subjects research. This could also be achieved by federal or state statute, by professional accreditation standards set by AAHRPP, or by changes in institution-level policies. But the costs of this choice may well outweigh the gains for most disputes, particularly those that do not allege physical injury or a legal claim against the institution (and even for these claims, the use of a neutral third party may still pose the problem of the

institution as a repeat player).[140] The structural advantages of having IRBs involved in dispute resolution for research-related injuries are great, and although non-neutrality is problematic, it is inherent to all IDR systems, and it is partially offset (from the subject perspective) by the IRB mandate to protect subjects. Imposing the requirement of a third-party neutral from outside the institution would also scale up the costs of disputes and could impose inefficient levels of process for minor complaints. Requiring subjects to bear these costs would impair access to the forum, as most subjects would be unable or unwilling to pay. Institutions could bear the costs, but this may impair neutrality of the forum for third-party decision-makers that were repeatedly retained. Requiring research sponsors to bear the costs would increase the expense of research more generally, posing tradeoffs between paying for more research or more administrative costs.

I will also stop short of recommending changes to the Common Rule to structure or constrain IDR as implemented by IRBs. This is for a similar reason; although we now have evidence from IRBs about how their processes currently work, including some likely deficiencies, we do not have systemic evidence that these deficiencies are experienced by subjects or researchers as harmful. IRBs described uses of procedural flexibility in order to promote participant priorities, such as voice and access. Mandating and monitoring IRB compliance with new regulatory requirements for complaint resolution, especially when the frequency of complaints may be low, is likely to increase inefficiencies in the current system. It may also discourage innovation, such as institutions that began using local trusted authorities in culturally or linguistically distinct participant populations to assist in handling disputes. Changing federal regulations may also not be necessary to improve IDR practices in IRBs; there are numerous examples of internal changes in IRBs that did not require a regulatory nudge.[141] Interviews with these informants suggested that many institutions were open to guidance and an opportunity to revisit their IDR procedures, and the informal nature of many of these IDR systems may facilitate the incorporation of new ideas without a regulatory requirement.

A. Consult Research Participants During System Design

First, IRBs should make efforts to consult participants at the moment of system design, or during periodic reevaluation of procedures. As noted above, this is not entirely straightforward, given that institutions often have many thousands of research portfolios representing a large number of different participant groups.

As a practical matter, consultation of participants' perspectives on dispute system design could either occur on a protocol-by-protocol basis or at the level of

140 Galanter, *supra* note 12.

141 Stark, *Victims*, *supra* note 72 (citing examples).

the IRB. On a per-protocol basis, IRBs could ask researchers to consult with representatives from participants or the larger community—such as through the use of a community advisory board[142]—to ascertain participant preferences for dispute resolution in the individual study. Or similarly, IRBs could ask researchers to disclose more information about the dispute resolution process, and to ask for informal feedback at the time of informed consent or the conclusion of studies.[143] Researchers could then report this information in aggregate back to the IRB for consideration. Another strategy may be for institutions to randomly select a small number of ongoing protocols and invite subjects enrolled in these protocols to give feedback on the dispute resolution procedure at the time of informed consent.

At the level of the IRB, the easiest (and least representative) method for soliciting feedback on the IDR system would be to ask participants for feedback while they are using the process, or perhaps after their issue is resolved. This may yield a biased perspective, however, because it will only capture the views of participants who have already chosen to use the system in its current form. IRBs could collect more representative feedback by soliciting comments from all participants in approved protocols at a given point in time—such as by allowing anonymous comments through a web portal, using a process akin to notice-and-comment rulemaking, or a series of public meetings.[144] Researchers could publicize this comment process to their current research participants. Or IRBs could prospectively identify the most common participant populations in their approved studies, and conduct focus groups sampling from these groups. This would be the most resource-intensive option, however, and it would likely be beyond the capacity of most IRBs.

The opportunity for subject participation in the design of these IDR processes may assist in improving access, procedural options, participation, and perceived legitimacy of the process. Where comments suggest potential improvements, IRBs could make provisional changes to their policies and assess the impact of these changes. These impacts should include outcomes such as complaint type and frequency, participant satisfaction, perceived legitimacy of the process,

142 A community advisory board (CAB) is a small group of community stakeholders in a research project that provides meaningful input on the design and implementation of a research protocol. *See, e.g.*, Stephen F. Morin et al., *Community Consultation in HIV Prevention Research: A Study of Community Advisory Boards at 6 Research Sites*, 33 J. ACQUIRED IMMUNE DEFICIENCY SYNDROMES 513 (2003); Sandra Crouse Quinn, *Protecting Human Subjects*: The Role of Community Advisory Boards, 6 AM. J. PUBLIC HEALTH 918 (2004).

143 Another strategy would be to require a representative for particular participant groups to be on the IRB, as is currently done for research with prisoners, 45 C.F.R. § 46.304(b) (2018)—but this may be more burdensome in practice.

144 Gathering these data would not count as "research" for IRB purposes, because it is not intended to contribute to "generalizable knowledge"—it would be solely for the purposes of improving internal operations. 45 C.F.R. § 46.102(l) (2018).

participants' perception of the institution's accountability during research, and participant awareness of the dispute resolution forum.

B. Increase Disclosure and Involve Participant Community Leaders

Second, IRBs should consider a range of other options to increase uptake and process utilization by participants. Although as a practical matter, no IRB wants to add to its workload, the informants in this study were convinced that low complaint frequencies indicated a problem with awareness and access. The remedy for lack of awareness is, of course, disclosure. IRBs can publicize their IDR processes on their websites, but it would be more useful to disclose more information at the time of informed consent. Several issues complicate disclosure. First, when processes are highly informal or malleable, there may be no formulate procedure to publicize; IRBs may therefore choose to highlight several process options, such as the option to make an anonymous complaint or the option of having an IRB staff member mediate communication with the investigator. Next, most investigators know little about the complaint resolution process, which means that institutions must educate not only subjects, but also investigators about this IRB function. Furthermore, adding elements to informed consent is not costless. Informed consent forms can be long and complex, and recent changes to the Common Rule reflect some of these problems.[145] Adding information about dispute resolution systems can compete for subject attention and extend the duration and complexity of the informed consent process. It may also attune participants to the possibility that they *could* be harmed, which could hinder enrollment or increase mistrust. But this is unlikely to be a substantial barrier; according to a recent study, even when participants are aware of the *death* of a healthy subject at the same institution, only 17% said this changed their thoughts about joining research, and only 4% said it would change their future participation.[146]

None of these drawbacks should hinder greater disclosure of institutions' processes for resolving research-related complaints. Meaningful consent to research must be predicated on "essential information that a reasonable person would want to know in order to make an informed decision about whether to participate"[147]—and the availability and quality of a forum to resolve research-related disputes and injuries may be essential for many participants. IRBs could potentially improve the effectiveness of these disclosures by asking investigators to convey this information verbally. Several reviews of informed consent strategies have shown that verbal disclosure and discussion is the most effective means of

145 80 Fed. Reg. 53970 (Sept. 8, 2015).

146 Caitlin E. Kennedy et al., *When a Serious Adverse Event in Research Occurs, How Do Other Volunteers React?*, 6 J. EMPIRICAL RES. HUM. RES. ETHICS 47 (2011).

147 82 Fed. Reg. 7149 (Jan. 19, 2017).

communicating with research participants,[148] and this would be an appropriate and efficient means of disclosing subjects' options in the event that complaints arise.

Another strategy for increasing process uptake may be to use a practice that several institutions have pioneered: asking investigators to identify a trusted member of the community to receive complaints and represent participant interests in communicating them to the IRB. Several institutions reported using trusted local authorities to help process complaints in research with distinctive populations, such as Native American tribes. One advantage of this process is that it outsources part of the responsibility to investigators to build stronger relationships with local subject communities; investigators must identify someone who can be familiar with the protocol and accept complaints, and then convey those complaints to the investigator or to the IRB. Investigators can then disclose this information to subjects as part of the informed consent process. Of course, subjects should keep the ability to complain to the IRB directly, in case the trusted local authority is unfamiliar or an inappropriate resource for them personally. But this may have additional benefits of improving investigators' engagement with participant populations, while also increasing the accessibility of the process to subjects. Another variation on this theme may be to add a member of the participant population as a temporary consultant to the IRB during deliberations about subject complaints arising from that protocol.

C. Compensate Participants for Physical Injuries

The informants in the study who expressed the greatest comfort with their IDR processes were at institutions that had agreed—either explicitly or as a de facto matter—to compensate participants for physical injuries sustained during human subjects research. There have been repeated calls and detailed proposals for U.S. research institutions to compensate participants for injuries, but this is not yet federally required.[149] Indeed, the NIH does not compensate participants for injuries, and there is no requirement that U.S. research institutions carry insurance for this purpose.[150] Many institutions had an unwritten practice of compensating injured participants, often by providing treatment themselves (e.g., at their own hospital) and waiving participant costs or cost-sharing. But nearly half of the institutions had a policy of never compensating subjects for physical injury (17%), or only compensating subjects when the research funders would agree to it up front (30%).

Compensation policies clearly facilitate dispute resolution of research-related complaints. IRB personnel who knew that their institution would ultimately pay

148 Flory & Emanuel, *supra* note 7; Nishimura et al., *supra* note 7.
149 Pike, *supra* note 5; Elliott, *supra* note 19.
150 *Id.*

participants for injuries sustained reported far greater confidence in managing disputes, less defensiveness, less concern about institutional liability and escalation of the dispute, and a greater sense that the system was operating ethically under Belmont Report principles for protecting human subjects. Although compensation was rarely if ever offered for non-physical injury, allowing compensation in cases of tangible harm was viewed as an essential procedural option. Informants at institutions that disallowed payments for injuries noted their frustration with this practice, and some commented that they wish their institution would institute more flexible policies.

This Article therefore echoes prior calls for institutions to compensate participants for tangible injuries sustained over the course of research, either by self-funding or purchasing insurance for this purpose. In addition to the ethical rationale for paying for research harms, allowing these payments has a highly pragmatic function of facilitating all dispute resolution in this context.

D. Build IRB Capacity for Conflict Resolution

The previous Part outlined some of the deficiencies of IRBs in expertise and resources for conflict resolution. The remedy is straightforward. In order to improve IDR processes—or to continue current processes in the event that process uptake increases—research institutions may need to devote additional personnel and training to IRB offices, or add administrative staff members who have prior training in conflict resolution. Very few of the personnel responsible for resolving complaints had training in dispute resolution; approximately 6% were trained as J.D.s, but even informants with law degrees noted that they lacked training on the interpersonal elements of conflict resolution or ADR. Research institutions could help meet these expertise needs by running workshops for IRB personnel—particularly managers and administrators, rather than members—or by considering conflict resolution training during hiring. Another method of increasing this expertise is to add modules to the Certified IRB Professional (C.I.P.) course run by the Council for Certification of IRB Professionals. More than 50% of informants in the study had obtained this qualification, suggesting that training modules on conflict management would be a good means of disseminating this information. Although AAHRPP accreditation was frequently described as complex and somewhat burdensome, an AAHRPP recommendation of having conflict resolution training would be another means of encouraging expertise-building among IRBs.

Human resources may be another need—again, particularly if the frequency of complaints increases. Complex complaints, although rare, were highly resource-intensive for IRB personnel. Many have called on research institutions to invest more in IRBs, and in human resource protection programs more generally, to

improve the speed and quality of protocol review. Improving IRB responses to participant complaints may be another reason to expand this area of the institution, if the frequency or complexity of complaints increases.

E. Use Records Effectively

IRBs can also improve their IDR systems through their practices for record-keeping and systematic examination of those records over time. Many IRBs did not record complaints in a manner that would allow for comparison across protocols, or over time. Making these comparisons at regular intervals, such as one- or two-year periods, could help IRBs identify recurring issues; they could address these through investigator training and protocol review instead of piecemeal responses to complaints. Creating a way to view complaints together would also improve institutional memory and consistency, particularly at times of personnel turnover, which may be essential for highly informal processes. IRBs are sensitive to local precedent,[151] and they may welcome opportunities to ensure that their responses to subject issues are consistent over time.

F. Provide for (Advisory) Third-Party Review

Instead of requiring the use of a third-party neutral for the initial resolution of every complaint, it may be more feasible and efficient to provide for appeals to an external reviewer or internal ombudsman to review IRBs' final decisions about complaints. At present, IRBs usually give investigators a written decision once a complaint is resolved. Investigators have an opportunity to appeal for reconsideration, but IRBs typically do not give or publicize to participants the possibility of an appeal. In some ways this lopsided procedure makes intuitive sense; IRBs can sanction investigators, but not participants, as part of the resolution, so investigators may make more use of this appeal mechanism. But from the participant's perspective, someone dissatisfied with the IRB's decision may feel that they have experienced harm without remedy, and some may want the same appeal option to demonstrate that they are being treated equally in the process.

An IRB could, therefore, address both concerns about neutrality and lopsided appeals by providing for an independent reviewer, which could be requested by the investigator, the participant, or even perhaps by the IRB itself if they seek a second opinion or fear institutional interference. This could function similarly to the external review mandated by state and federal law for coverage disputes in private health insurance, but would likely be far smaller in scope.[152] Given IRBs'

151 STARK, *supra* note 72

152 Hunter, *supra* note 32.

current goal of subject satisfaction with the process—and their view that most participants are in fact satisfied—the uptake (and therefore costs) of this external review are likely to be fairly low. Institutions could collaborate with one another to develop the infrastructure for this independent reviewer—for example, research institutions in each state could contribute to the costs of maintaining an ad hoc independent external reviewer for the state or region. When a subject or researcher invokes independent review, the IRB would then send the reviewer any reports of the complaint investigation and decision for their independent analysis and written opinion.

Although this independent review process may resemble the process of external review for health insurance coverage decisions, the process will necessarily be weaker. In external review for health insurance coverage disputes, the decision of the external review process is binding on the insurance company. But for structural reasons, binding external review is complex and likely not viable here. The federal regulations delegate authority for research protocol approval, disapproval, and oversight to IRBs; the rules specifically provide that research "may be subject to further appropriate review and approval or disapproval by officials of the institution. However, those officials may not approve the research if it has not been approved by an IRB."[153] The Common Rule does not permit institutions to delegate this authority outside the IRB (although using an external, paid IRB that is subject to federal regulation is permitted). Moreover, if the reviewer were an ombudsman within the institution, he or she could require *more stringent* protocol restrictions or termination, but could not *lift* protocol restrictions or reverse a study termination required by the IRB. This would make binding review of little use to investigators facing sanctions. Some complaints may also raise issues outside the IRB's purview, such as complaints of investigator harassment, which are typically referred to human resources and handled as legal matters.

For this reason, *binding* review by an independent party, or even binding review by an internal ombudsman who is not part of the IRB, is likely unavailable here; review will be advisory rather than binding. But even an *advisory* review of IRB decisions would be useful in alleviating concerns about neutrality and the inequality of the current appeals process. IRBs will have the opportunity to reconsider their findings in light of the third-party reviewer's recommendation, and then to adjust any protocol sanctions or remedies provided. The availability of a third-party advisory review may also shape IRBs' actions even when it is not invoked. IRBs that know a third party will evaluate their decision may take greater care in their analysis and written decisions, and they may produce (and subsequently use) better records of their process. All of these changes may help

153 45 C.F.R. § 46.112 (2018).

produce fairer and more effective decision-making throughout the dispute resolution process.

VI. CONCLUSION

The empirical study in this Article was the first in-depth look at the highly flexible systems that research institutions have established to mediate and, at times, adjudicate disputes involving human subjects. Disputes in this area are characterized in part by high stakes for investigators and institutional exposure to liability, but also by disparities in socioeconomic power and sophistication between participants and research institutions. Attention to fair process is therefore an ethical and practical imperative for functioning systems. At present, institutions' IDR systems take advantage of IRBs' mandate and authority to protect subjects, and IRBs have instituted highly flexible procedures to maximize the voice and satisfaction of research subjects who bring grievances. But notwithstanding these strengths, IDR systems for research-related complaints also pose problems of inclusion, access, neutrality, resources, and expertise. Changes to the Common Rule, such as the requirement that multisite studies designate one IRB of record, may continue to bring changes to how research-related disputes are resolved.

In light of these findings, this Article has recommended a number of structural changes to how IRBs handle research-related grievances. These include suggestions for considering participant input on system design; increasing publicity and accessibility through informed consent procedures and integration of participant community leaders; compensating participants for physical injuries; building IRB expertise and resources for conflict resolution; using records to identify recurring complaints and improve consistency; and providing for advisory third-party review and reconsideration of decisions, even if that review is not binding. Institutions dedicated to protecting the welfare of human subjects may well make these changes without being prompted by a change in federal or state regulations; with the exception of the suggestion that institutions compensate injured participants (which has repeatedly been ignored), these ideas build on existing systems and do not require large resource outlays. The practical rewards of a functioning IDR system may be great, including reduced institutional exposure, improved community relations, and increased legitimacy of research at the institution. But most importantly, these adjustments to IDR processes for research-related harms are ethically warranted. The Belmont Report and other ethical guidelines have spoken widely on the need to minimizing subject harm, but have said little about how institutions can (and should) offer redress when they fail to do. Participants in human subjects research take on many burdens in the interests of scientific progress; when they experience unintended harms, they should not

bear the additional burden of unfair process. This Article is a start toward that goal.

The Facts of Stigma: What's Missing from the Procedural Due Process of Mental Health Commitment

Alexandra S. Bornstein*

Abstract:

This is the first systematic review of federal, judicial opinions that engage the stigma of mental health commitment in the context of procedural due process. In 1979, in *Addington v. Texas*, the Supreme Court held that the stigma, or adverse social consequences, of civil commitment is relevant to the procedural due process analysis. The following year, in *Vitek v. Jones*, the Court held that the stigmatizing consequences of a transfer from a prison to a mental health facility, coupled with mandatory treatment, triggered procedural protections. While these cases importantly suggested a role for stigma in procedural due process, they left many questions related to the implementation of these standards unanswered. As a result, across the cases analyzed in this review, judges expressed different views of this stigma and consistently underestimated the real impact of this stigma. This in turn resulted in judges consistently underestimating the liberty interest created by commitment and the need for procedural due process. In order to properly protect individuals against the risk of erroneous commitment, judges must engage in further fact finding to determine the real harm that results from the stigma of mental health commitment.

* Columbia Law School, J.D. 2018; Middlebury College, B.A. 2011. Many thanks to Professor Kristen Underhill, for her guidance through every stage of this process; to the editors of the *Yale Journal of Health Policy, Law, & Ethics*, for their excellent feedback and editorial assistance; and to my family, Susan, Mitch, Matt, and Tim, for their constant support.

I. INTRODUCTION

A study published in 2000 found that 54 percent of respondents believed an individual with any mental illness was a danger to others. That same study found that 58 percent of respondents would not want an individual with any mental illness as a coworker and that 68 percent would not want that same individual marrying into their family.[1] Research suggests that the stigma associated with serious mental illness, mental illness that might require either voluntary or involuntary inpatient hospitalization, is even more profound. A 2008 survey on the public perception of one serious mental illness, schizophrenia, found that 77 percent of people would feel uncomfortable and 80 percent would fear for their safety around a person with untreated schizophrenia; 77 percent would feel uncomfortable working with that person; and 80 percent expressed discomfort related to dating that person.[2] Such stigma is unsurprising when viewed in light of how serious mental illness and mental hospitals are portrayed in popular culture—think *One Flew Over the Cuckoo's Nest*, the more recent *American Horror Story: Asylum* (a hospital physician experiments on patients and then leaves them to feed on other patients), or the mental-hospital-themed haunted houses that pop up all over the country for Halloween.[3]

This Note examines how that stigma affects the procedural due process afforded to individuals subject to involuntary hospitalization for mental illness. Nearly forty years ago, the Supreme Court recognized the severity of the stigma associated with involuntary commitment to a mental health hospital in a pair of cases related to civil and criminal mental health commitment, respectively. In *Addington v. Texas*, the Court considered the appropriate standard of proof to be applied in civil commitment hearings. Justice Burger, writing for the Court, stated that civil commitment constitutes a deprivation of liberty in part because

---

1 Jack K. Martin et al., *Of Fear and Loathing: The Role of 'Disturbing Behavior,' Labels, and Causal Attributions in Shaping Public Attitudes toward People with Mental Illness*, 41 J. HEALTH & SOC. BEHAV. 208, 216 (2000).

2 *Schizophrenia: Public Attitudes, Personal Needs*, NAT'L ALL. ON MENTAL ILLNESS 8 (May 13, 2008), https://www.nami.org/schizophreniasurvey.

3 Colby Itkowitz, *Halloween Attractions Use Mental Illness to Scare Us. Here's Why Advocates Say It Must Stop*, WASH. POST (Oct. 25, 2016), https://www.washingtonpost.com/news/inspired-life/wp/2016/10/25/this-halloween-mental-health-advocates-are-taking-a-powerful-stand-against-attractions-depicting-asylums/?utm_term=.80e98b67420a. An amusement park on the border of North and South Carolina includes this description for its "7th Ward Asylum": "You would be crazy to tour this twisted asylum. Lost and tortured souls are all that remain, but you'll see plenty that will make you question your sanity . . . . The 7th Ward was home to the Carolina's most chronically insane. From murderers to crazed psychopaths, many of the poor souls trapped behind the Gothic walls would spend their entire lives there. As you walk these halls today, be sure to stay with your group. This is one place you don't want to be committed." *Id.*

commitment creates "adverse social consequences . . . whether we label this phenomena 'stigma' or choose to call it something else is less important than that we recognize that it can occur and that it can have a very significant impact on the individual."[4] In the following year, in *Vitek v. Jones*, the Court considered a procedural due process challenge to a Nebraska statute that gave the Director of Correctional Services the authority to transfer an incarcerated individual to a mental health facility without notice or hearing. The plaintiff argued that the Due Process clause entitled him to procedural protections before commitment because he had a liberty interest in not being stigmatized by commitment to a mental health facility. The Court agreed, to the extent that this stigma existed and was relevant to the procedural due process analysis. Justice Burger wrote "the stigmatizing consequences of a transfer to a mental hospital for involuntary psychiatric treatment, coupled with the subjection of the prisoner to mandatory behavior modification as a treatment for mental illness, constitute the kind of deprivations of liberty that requires procedural protections."[5]

These two cases suggested, for the first time, a role for stigma in procedural due process analysis with respect to mental health commitment. Yet in so doing, the Supreme Court provided only minimal explanation for how it arrived at the underlying conclusion that stigma results from commitment or how this conclusion fits into the broader procedural due process analysis. In the nearly forty years since these holdings, the Supreme Court has offered little clarification. Instead, it has left it to lower court judges to determine how to engage with the stigma of mental health commitment in the context of procedural due process.

This study seeks to determine how judges have applied the holdings in *Addington* and *Vitek* to measure their real impact on the procedural due process protections afforded to individuals facing mental health commitment proceedings. A systematic review was conducted of all federal, judicial opinions that discussed the stigma of mental health commitment in the context of procedural due process analysis since the Supreme Court decided these two cases. This methodology was utilized for its application in analyzing the variability in how judges have interpreted these standards across all opinions that have engaged with them.[6]

---

4 Addington v. Texas, 441 U.S. 418, 425–26 (1979).

5 Vitek v. Jones, 445 U.S. 480, 494 (1980).

6 For further discussion of the advantages of systematic review in the context of legal doctrinal analysis, see William Baude et al., *Making Doctrinal Work More Rigorous: Lessons from Systematic Reviews*, 84 U. CHI. L. REV. 37, 42 (2017) (arguing that systematic review reduces the need for the reader to rely on the author's credibility to believe her claims, makes it easier for the reader to access the uncertainty associated with a claim, creates more complete documentation which can support progress in the field, decreases real or perceived bias, and can reduce error); Mark A. Hall & Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CAL. L.

The results of the following analysis suggest that there is immense variation in how judges have engaged with the stigma of mental health commitment in the context of procedural due process, since *Addington* and *Vitek*. Among some opinions, judges determined that the presence of this stigma clearly required procedural protection. Among others, it was much more difficult to ascertain what role this stigma played in the procedural due process analysis. In general, across all opinions, judges spent very little time discussing the stigma of mental health commitment. The result is that judges seem to have profoundly different understandings of stigma in this context and its role in the procedural due process analysis. Additionally, judges consistently fail to engage with current empirical evidence related to the real consequences of this stigma. The variability in judges' treatment of stigma, as well as their anemic understanding of the many psychological, social, and economic consequences of stigma, has amounted to a systematic bias against plaintiffs seeking due process protection in commitment proceedings. While *Addington* and *Vitek* importantly included stigma in the procedural due process analysis in the context of mental health commitment, in many cases, judges have not implemented these standards properly. In these cases, judges must engage in their own fact finding to address those questions left unanswered by the Supreme Court. Given the limited fact-finding resources available to lower courts, this Note argues that a resource that aggregates relevant information on this subject, almost like a publicly available amicus brief, could assist judges in appropriately considering this issue while ensuring that judges engage with current research on the subject.

The following section will provide a brief overview of mental health commitment laws in the United States, research related to the stigma associated with mental illness and mental health commitment, and a discussion of cases that have established stigma as relevant to the procedural due process analysis, including *Addington* and *Vitek*.

II. BACKGROUND ON MENTAL HEALTH COMMITMENT AND THE ROLE OF STIGMA IN PROCEDURAL DUE PROCESS ANALYSIS

a. Mental health commitment laws

Involuntary, mental health commitment is the process by which the government compels an individual to receive mental health treatment in an inpatient, mental health facility.[7] There are different mechanisms and standards

REV. 63 (2008) (discussing other advantages of systematic review of judicial opinions).

7 Most states also have outpatient commitment or assisted outpatient treatment laws, which give judges the authority to compel individuals to receive outpatient, community-based, mental health treatment. See *What is AOT?*, TREATMENT ADVOCACY CTR. (Apr. 12, 2017), http://www.treatmentadvocacycenter.org/storage/documents/aot-one-pager.pdf. Outpatient

by which a person may be involuntarily committed. Although largely the province of state governments, there are also federal laws that dictate mental health commitment for certain populations.

i. Civil commitment laws

All fifty states and the District of Columbia have civil commitment laws. States have grounded their authority to enact such laws in two powers: the police power, to protect the state's citizens from potentially dangerous people, and the *patriae parens* power, to protect potentially dangerous people from themselves. Although the specifics of these laws vary across states, most reflect these dual purposes. These laws generally require some showing that the individual is in fact mentally ill and that the individual is either a danger to themselves or to others.[8] While the standards for dangerousness to others is and has been relatively consistent across states, the standard for dangerousness to one's self varies across states and has varied over time. Previous standards limited the consideration to whether an individual presented an immediate, intentional, violent threat to themselves, specifically whether an individual had attempted suicide or engaged in self-mutilation, and whether such behavior would likely result in serious harm or death.[9] Current standards still consider these factors but vary in what else they consider. For example, the Treatment Advocacy Center, a group that advocates for comprehensive mental health treatment including, when appropriate, mental health commitment, gives Pennsylvania's commitment law a failing grade for its limited definition of dangerousness to one's self.[10] In addition to considering the likelihood of intentional, violent self-harm, judges also consider whether there is evidence that an individual is unable to "to satisfy . . . [their] need for nourishment, personal or medical care, shelter, or self-protection and safety" and that their inability to do so creates a "reasonable probability that death, serious bodily injury or serious physical debilitation would ensue within 30 days unless adequate treatment were afforded" through commitment.[11]. This standard is known as "grave disability."

By contrast, the Treatment Advocacy Center gives Illinois's civil commitment law its highest possible grade due to its expansive definition of

---

commitment laws are not discussed in this Note and, to simplify, inpatient commitment is referred to as simply "mental health commitment."

8 See Megan Testa & Sara G. West, *Civil Commitment in the United States*, 10 PSYCHIATRY 30, 33 (2010).

9 *Mental Health Commitment Laws A Survey of the States*, TREATMENT ADVOCACY CTR. (Feb. 2014), http://www.treatmentadvocacycenter.org/storage/documents/2014-state-survey-abridged.pdf.

10 *Id.*

11 50 PA CONS. STAT. § 7301(b)(2)(i).

danger to one's self.[12] Illinois' law applies a "need-for-treatment" standard such that judges also consider whether an individual refuses to comply with treatment or cannot understand the need for treatment and as a result will likely suffer "mental and emotional deterioration."[13] As such, an individual may be committed under Illinois' law before they become gravely disabled. Laws also vary across states with respect to who may commence proceedings; in some states, any party may commence proceedings, such as an individual's family member,[14] while in other states proceedings may only be commenced by mental health professionals.[15] The federal government does not have a civil commitment law, but federal courts may of course consider procedural due process challenges to state civil commitment laws.

ii. Criminal commitment

Mental health commitment can also occur within state and federal prison systems. Prior to 1820, most people deemed mentally ill were imprisoned, not as a means of punishment but to remove them from the larger population.[16] In the 1820s, activists began protesting conditions and the lack of adequate mental health treatment in prisons. These activists advocated for the building of hospitals dedicated to the proper treatment of individuals with mental health conditions. By 1880, there were seventy-five public mental health hospitals and the majority of people diagnosed with mental health conditions had been transferred from prisons to these hospitals. The census in that year reported that, of all "insane people," less than one percent were still residing in prisons or jails, while the remaining ninety-nine percent (nearly 59,000 people) were in public mental health facilities.[17]

Eventually, this system broke down as well. By the 1960s, the poor conditions of these facilities created a backlash known as the "deinstitutionalization" movement.[18] The deinstitutionalization movement called for and eventually succeeded in reducing the number of people confined to residential, mental health facilities. While seemingly well intentioned, this movement removed people from their residential treatment without providing

12 405 ILL. COMP. STAT. 5/1-119.

13 405 ILL. COMP. STAT. 5/1-119.

14 For example, any "responsible party" may commence the process of involuntary commitment in a Pennsylvania trial court. 50 PA CONS. STAT. § 7304(c)(1).

15 See, e.g., New York's inpatient commitment law. N.Y. MENTAL HYG. LAW § 9.27(a).

16 *The Treatment of Persons with Mental Illness in Prisons and Jails: A State Survey*, TREATMENT ADVOCACY CTR. 9–11 (Apr. 8, 2014), http://www.treatmentadvocacycenter.org/storage/documents/treatment-behind-bars/treatment-behind-bars.pdf.

17 *Id.*

18 *Id.* at 11.

adequate alternative treatment. Without treatment, people were unable to successfully reincorporate into society and many committed crimes for which they were arrested and imprisoned. A prison psychologist was quoted in a seminal 1972 article saying, "[w]e are literally drowning in patients."[19] This trend has continued. [20] According to surveys done by the Department of Justice in 2002 and 2004, forty-four percent of all federal prisoners, fifty-six percent of all state prisoners, and sixty-four percent of all individuals in local jails reported experiencing mental health symptoms or receiving treatment from a mental health professional in the previous twelve months.[21] These estimates compare to roughly eighteen percent of the general population, according to a 2014 study done by the National Institute of Mental Health.[22]

Although many people with mental health conditions that are convicted of crimes are incarcerated and remain incarcerated, there are both state and federal laws that allow for commitment to mental health facilities within the criminal justice system. 18 U.S.C. §§ 4243–4246 provide procedure by which a federal criminal offender may be either initially placed in or transferred into a mental health facility.[23]

If an individual is found not guilty of an offense for reason of insanity, 18 U.S.C. § 4243 provides that that individual will be committed to a mental health facility unless it can be shown, by clear and convincing evidence, that their "release would not create a substantial risk of bodily injury to another person or serious damage of property of another."[24] If an individual is convicted of an offense and suffers from a mental health condition, but does not bring an insanity defense, 18 U.S.C. § 4244 provides that they may still be committed to a mental health facility rather than being incarcerated.[25] In this case, the Attorney General may request a hearing to demonstrate that that individual should still be committed to a mental health facility prior to sentencing.[26] Per 18 U.S.C. § 4245, if an individual was convicted of a crime, incarcerated, and then later determined

---

19 *Id.*

20 See HUMAN RIGHTS WATCH, ILL-EQUIPPED: US PRISONS AND OFFENDERS WITH MENTAL ILLNESS 24 (2001), www.hrw.org/reports/2003/usa1003 ("Thousands of mentally ill are left untreated and unhelped until they have deteriorated so greatly that they wind up arrested and prosecuted for crimes they might never have committed had they been able to access therapy, medication, and assisted living facilities in the community.").

21 DEP'T OF JUST., NCJ 213600, MENTAL HEALTH PROBLEMS OF PRISON AND JAIL INMATES 1 (2006), https://www.bjs.gov/content/pub/pdf/mhppji.pdf.

22 *Any Mental Illness (AMI) Among U.S. Adults*, NATIONAL INSTITUTE FOR MENTAL HEALTH (2014), https://www.nimh.nih.gov/health/statistics/prevalence/any-mental-illness-ami-among-us-adults.shtml.

23 18 U.S.C. § 4243–4246 (2012).

24 18 U.S.C. § 4243(d) (2012).

25 18 U.S.C. § 4244 (2012).

26 *Id.*

to require inpatient treatment, they may be transferred to a mental health facility after a hearing is held.[27] The Nebraska analogue to this federal law, which was at issue in *Vitek v. Jones*, did not require a hearing prior to transfer. This law will be discussed further in Part II.c, *infra*. Finally, 18 U.S.C. § 4246 provides the procedure by which an individual may continue to be committed even after his initial sentence has elapsed.[28] All fifty states and the District of Columbia have similar laws allowing for commitment within the prison system.

b. Stigma associated with mental illness and mental health commitment

The classical sociological literature defines stigma as an "'attribute that is deeply discrediting' that reduces the bearer 'from a whole and usual person to a tainted, discounted one.'"[29]A more recent review of the literature provides several definitions of the term: "[a] deeply discrediting attribute; 'mark of shame'; 'mark of oppression'; devalued social identity."[30] The authors go on to describe four essential components of stigma. These elements include: "(a) distinguishing and labeling differences, (b) associating human differences with negative attributions or stereotypes, (c) separating 'us' from 'them,' and (d) experiencing status loss and discrimination."[31]

Both Justice Burger in *Addington* and Justice White in *Vitek* focused on the consequences of the stigma associated with mental health commitment. Much research has been done on this topic. The relevant literature in fact identifies two related though distinct types of stigma that can have different consequences for individuals: public stigma and internalized stigma.[32] Public stigma is "the phenomenon of large social groups endorsing stereotypes about and acting against a stigmatized group."[33] Studies have identified numerous consequences correlated with the public stigma associated with mental illness. These consequences include, for example, underemployment, joblessness, and the inability to live independently.[34] While mental illness itself can affect these

27 18 U.S.C. § 4245 (2012).

28 18 U.S.C. § 4246 (2012).

29 Paula Abrams, *The Scarlet Letter: The Supreme Court and the Language of Abortion Stigma*, 19 MICH. J. GENDER & L. 293, 299 (2013) (quoting ERVING GOFFMAN, STIGMA: NOTES ON THE MANAGEMENT OF SPOILED IDENTITY 3 (1963)).

30 Bernice A. Pescosolido & Jack K. Martin, The Stigma Complex, 41 ANN. REV. SOC. 87, 92 (2015).

31 *Id.* at 91.

32 J.D. Livingston & J.E. Boyd, *Correlates and Consequences of Internalized Stigma for People Living with Mental Illness: A Systematic Review and Meta-Analysis*, 71 SOC. SCI. & MED. 2150, 2151 (2010).

33 *See, e.g.*, Patrick W. Corrigan et al., *The Stigma of Mental Illness: Explanatory Models and Methods for Change*, 11 APPLIED & PREVENTIVE PSYCHOL. 179 (2005).

34 *Id.*

outcomes directly, these studies demonstrate that stigma has an independent, additional effect. Other studies have also found public stigma to be associated with social isolation and a lower likelihood of seeking treatment.[35]

Internalized stigma can affect how individuals view themselves. Individuals may come to believe that they do in fact possess the negative attributes that are ascribed to their broader stigmatized group. Individuals with mental illness may come to believe, for example, that they are dangerous or incompetent.[36] Studies have shown internalized stigma to be associated with negative consequences, including increased symptom severity and poorer treatment adherence.[37]

Although less frequently studied, involuntary commitment and hospitalization generally have been found to have an even greater stigmatizing effect than being perceived as mentally ill or receiving outpatient treatment.[38] A recent study of several hundred individuals with serious mental illness who had been involuntarily hospitalized found that hospitalization created additional internalized stigma. Specifically, the study found greater incidence of feelings of shame and self-contempt, which in turn was found to lead to lower self-esteem and lower quality of life.[39] Another qualitative study found that individuals reported higher levels of discrimination following hospitalization.[40] A Brazilian study conducted among a hundred and sixty individuals with a history of involuntary commitment found that individuals with families with more biased views towards mental illness were more likely to be re-committed.[41]

In the prison context, there are a number of negative consequences associated with being committed and being perceived as mentally ill. Prisoners, unsurprisingly, often possess the same biases against people with mental illness as do the general population. Prisoners labeled as mentally ill, experience social

---

35 Deborah A. Perlick et al., *Stigma as a Barrier to Recovery: Adverse Effects of Perceived Stigma on Social Adaptation of Persons Diagnosed with Bipolar Affective Disorder*, 52 PSYCHIATRIC SERVICES 1627 (2001); 2003 National Survey on Drug Use & Health: Results, Dep't of Health & Human Servs., Substance Abuse & Mental Health Servs. Admin., & Office of Applied Studies, (June 3, 2008), http:// www.oas.samhsa.gov/nhsda/2k3nsduh/2k3Results.htm.

36 *See, e.g.*, Corrigan, P.W. et al., *The Internalized Stigma of Mental Illness: Implications for Self-Esteem and Self-Efficacy*, 25 J. SOC. & CLINICAL PSYCHOL. 875 (2006); Jennifer Boyd Ritsher & Jo C. Phelan, *Internalized Stigma Predicts Erosion of Morale Among Psychiatric Outpatients*, 129 PSYCHIATRY RES. 257 (2004); Philip T. Yanos et al., *The Impact of Illness Identity on Recovery from Severe Mental Illness*, 13 AMER. J. PSYCHOL. REHABILITATION 73 (2010).

37 *See* Livingston & Boyd, *supra* note 32.

38 Nicolas Rüsch et al., *Emotional Reactions to Involuntary Psychiatric Hospitalization and Stigma-Related Stress Among People with Mental Illness*, 264 EUR. ARCHIVES PSYCHIATRY CLINICAL NEUROSCIENCE 35 (2014).

39 *Id.*

40 Ingrid Sibitz et al., *Impact of Coercive Measures on Life Stories: Qualitative Study*, 199 BRIT. J. PSYCHIATRY 239 (2011).

41 Alexandre Andrade Loch, *Stigma and Higher Rates of Psychiatric Re-hospitalization: São Paulo Public Mental Health System*, 34 REVISTA BRASILEIRA DE PSIQUIATRIA 185 (2012).

isolation and additional stigmatization.[42] One account of prison life by Victor Hassine, a formerly incarcerated person, described individuals perceived as mentally ill as fundamentally disruptive to prison life. He wrote, "Their helplessness often made them the favorite victims of predatory inmates. Worst of all, their special needs and peculiar behavior destroyed the stability of the prison system."[43] It has been found that mentally ill prisoners are disproportionately victims of physical and sexual violence while in prison. A 2007 study of over 7,500 prisoners (randomly sampled from a population of roughly 20,000 prisoners) found that the number of incarcerated men that reported being victims of sexual violence was three times higher among men with mental health conditions than among men without diagnosed mental health conditions (one in twelve compared to one in thirty-three).[44] The study also found a higher likelihood of reported sexual victimization among women with mental health conditions than among women without mental health conditions.[45] It has also been found that women diagnosed with mental illness are less likely to receive parole.[46]

c. Supreme Court jurisprudence on stigma in the procedural due process analysis

Procedural due process guarantees that no state nor the federal government "shall . . . deprive any person of life, liberty, or property, without due process of law."[47] State-imposed stigma has for a long time been considered relevant to the existence of a liberty interest. Prior to 1976, several cases decided by the Supreme Court suggested that stigma, or reputational harm, created by the state was enough to implicate a liberty interest, thereby triggering due process protection.[48] Yet in 1976, in *Paul v. Davis*, the Supreme Court reversed course, holding that reputational harm created by a state-imposed label was relevant but not sufficient to trigger procedural protection under the Due Process clause of the

42 HUMAN RIGHTS WATCH, ILL-EQUIPPED: US PRISONS AND OFFENDERS WITH MENTAL ILLNESS 24 (2001) (citing TERRY KUPERS, PRISON MADNESS: THE MENTAL HEALTH CRISIS BEHIND BARS AND WHAT WE MUST DO ABOUT IT 20 (1999)).

43 VICTOR HASSINE, LIFE WITHOUT PAROLE: LIVING IN PRISON TODAY 29 (1996).

44 Nancy Wolff et al., *Rates of Sexual Victimization in Prison for Inmates With and Without Mental Disorders*, 58 PSYCHIATRIC SERVICES 1087, 1090 (2007).

45 *Id.* at 1091.

46 Kelly Hannah-Moffat, *Losing Ground: Gendered Knowledges, Parole Risk, and Responsibility*, 11 SOC. POL. 363 (2004).

47 U.S. Const. amends. V, XIV.

48 Eric J. Mitnick, *Procedural Due Process and Reputational Harm: Liberty as Self-Invention*, 43 U.C. DAVIS L. REV. 79, 83–86 (2009) (citing Jenkins v. McKeithen, 395 U.S. 411, 429 (1969); Wisconsin v. Constantineau, 400 U.S. 433, 437 (1971); Bd. of Regents v. Roth, 408 U.S. 564, 573–74 (1972)).

Fourteenth Amendment.[49] The Court held that state-created stigma only triggers procedural due process protection when it is accompanied by the abridgement of some "right or status previously recognized by state law" or "guaranteed in one of the provisions of the Bill of Rights."[50]

In September 1971, Edward Charles Davis III was arrested in Louisville, Kentucky for shoplifting. The charge was later dismissed. A year later, the chief of police of Louisville, acting in his official capacity, distributed a flyer identifying "Active Shoplifters."[51] A photo of Davis along with his name was included on the flyer. When Davis's employer found out that he had been listed in this flyer, he was not fired but was told that another arrest could lead to his termination. Although not actually fired, Davis stated that he felt "humiliation and ridicule" from members of his department and he ultimately left the job.[52] After leaving this job, he found it difficult to find new employment. At the time of the lawsuit, he was unemployed.[53]

Davis filed a 42 U.S.C. § 1983 claim, arguing that his inclusion on the flyer by the police chief without appropriate procedural protections violated his right to procedural due process.[54] The District Court found for the police chief, but when Davis appealed, the Sixth Circuit reversed. The Supreme Court granted certiorari and held that stigma was relevant but insufficient to garner procedural protections.[55] The court explained that due process protection was intended to protect those rights guaranteed through either state law or the Constitution. Reputation alone, without some additional harm, was not protected by either.[56] This standard, that stigma coupled with some tangible harm recognized by law, such as loss of employment or property, triggers due process protection, became known as the "stigma plus" standard.[57]

Three years later, in *Addington v. Texas*, the court considered how stigma that results from a state-imposed label affects the procedural due process analysis in the context of civil commitment proceedings.[58] Appellant, Frank O'Neal Addington, had been temporarily committed several times from 1969–1975. After he was arrested for "assault by threat" against his mother, she filed a

---

49 Paul v. Davis, 424 U.S. 693 (1976).

50 *Paul*, 424 U.S. at 710 n.5, 711.

51 *Id.* at 695; Mitnick, *supra* note 48 at 87.

52 Mitnick, *supra* note 48 at 88 (citing Edward Charles Davis III, *A "Keep Out" Sign on the Courthouse Doors?*, JURIS DR., (1976)).

53 *Id.*

54 *Paul*, 424 U.S. at 694.

55 *Id.* at 696–97.

56 *Id.* at 708.

57 See Lindsey Webb, The Procedural Due Process Rights of the Stigmatized Prisoner, 15 U. PA. J. CONST. L. 1055, 1069 (2013).

58 *Addington*, 441 U.S. at 418.

petition to have him committed indefinitely.[59] At trial, the judge instructed the jury that to commit Addington, their findings must be substantiated by clear and convincing evidence. Following the jury's finding that Addington should be committed, Addington appealed on procedural due process grounds. He argued that because civil commitment results in the same deprivation of liberty as imprisonment, due process requires the application of the higher, beyond a reasonable doubt evidentiary standard.[60] The state appellate court agreed and reversed, but on appeal, the Texas Supreme Court reversed again. The Texas Supreme Court found that procedural due process only required proof based on a preponderance of the evidence, an even lower standard than the trial court had initially required. Addington appealed to the Supreme Court and the Court granted *certiorari*.

Ultimately, the Supreme Court held that, in fact, while the highest standard was not required, the intermediate clear and convincing evidence standard was appropriate, because civil commitment "constitutes a significant deprivation of liberty that requires due process protection."[61] In reaching this conclusion, Justice Burger, writing for the Court,[62] stated that civil commitment following the determination that an individual is dangerous (which was required by the Texas law) creates "adverse social consequences" for the committed individual."[63] He further elaborated: "whether we label this phenomena [sic] 'stigma' or choose to call it something else is less important than that we recognize that it can occur and that it can have a very significant impact on the individual."[64]

Vitek v. Jones was decided the following year.[65] On May 31, 1974, appellant Larry D. Jones was convicted of robbery and sentenced to three to nine years in Nebraska state prison. Nine months later he was transferred to the prison hospital and then placed in solitary confinement. While in solitary confinement, he burned his mattress and burned himself in the process. After being treated for the resulting burns, he was transferred to a state mental hospital.[66] The transfer was authorized by a Nebraska statute, which stated that: "[w]hen a designated physician or psychologist finds that a prisoner 'suffers from a mental disease or defect' and 'cannot be given proper treatment in that facility,'" the Director of Correctional Services may transfer that prisoner to any suitable facility within or outside of the correctional system.[67]

---

59 *Id.* at 419.
60 *Id.* at 421.
61 *Id.* at 425.
62 Justice Powell took no part in the consideration of the case or the decision.
63 *Id.* at 426.
64 *Addington*, 441 U.S. at 425–26.
65 *Vitek*, 445 U.S. at 480.
66 *Id.* at 484.
67 Id. at 483 (citing NEBRASKA REV. STAT. § 83–180(1) (1976)).

Following his transfer, Jones joined a suit challenging the constitutionality of the Nebraska statute. Although people lose many freedoms upon incarceration, the Supreme Court has held that "[p]risoners may . . . claim the protections of the Due Process Clause. They may not be deprived of life, liberty, or property without due process of law."[68] A three-judge District Court, empaneled pursuant to 28 U.S.C. § 2281 (1970) (now repealed), found for Jones and his fellow plaintiffs, determining that the statute was unconstitutional because a transfer to a mental health facility invoked a liberty interest that requires additional procedural protections."[69] The District Court enjoined the state from transferring Jones to the mental hospital without appropriate due process.[70] The state appealed to the Supreme Court directly.[71] The Supreme Court upheld the judgment of the District Court.[72] Justice White, writing for the majority, stated its holding:

> the stigmatizing consequences of a transfer to a mental hospital for involuntary psychiatric treatment, coupled with the subjection of the prisoner to mandatory behavior modification as a treatment for mental illness, constitute the kind of deprivations of liberty that requires procedural protections.[73]

As in Paul, the Court held that stigma was insufficient alone to create a liberty interest, but that stigma that resulted from a transfer coupled with mandated treatment implicated a liberty interest and therefore required procedural protections.[74] This was the first time the Supreme Court explicitly

---

68 Wolff v. McDonnell, 418 U.S. 539, 556 (1974).

69 *Id.* at 488.

70 28 U.S.C. § 2281 provided: "An interlocutory or permanent injunction restraining the enforcement, operation or execution of any State statute by restraining the action of any officer of such State in the enforcement or execution of such statute or of an order made by an administrative board or commission acting under State statutes, shall not be granted by any district court or judge thereof upon the ground of the unconstitutionality of such statute unless the application therefor is heard and determined by a district court of three judges under section 2284 of this title. 28 U.S.C. § 2281 (1970).

71 28 U.S.C. § 1253 provided for direct appeal to the Supreme Court of this type of injunction ("Except as otherwise provided by law, any party may appeal to the Supreme Court from an order granting or denying, after notice and hearing, an interlocutory or permanent injunction in any civil action, suit or proceeding required by an Act of Congress to be heard and determined by a district court of three judges."). 28 U.S.C. § 1253 (1970).

72 *Vitek*, 445 U.S. at 485 (citing Vitek v. Miller, 434 U.S. 1060 (1978)). While it was ultimately a 5–4 decision, those writing in concurrence and dissent did not disagree with the court's holding that this type of transfer required due process protections. Rather, these justices disagreed with respect to the appropriate level of procedural protections and whether the Court could hear the case at all. See *id.* at 497 (Powell, J., concurring in part); *id.* at 501 (Stewart, J., dissenting); *id.* at 501 (Blackmun, J., dissenting).

73 *Id.* at 494.

74 Webb, supra note 57 at 1073–74 (quoting Vitek, 445 U.S. at 494) ("The Vitek Court, like Paul, found a liberty interest in the

included stigma in the due process analysis associated with transfer from a prison to a mental health facility, or any involuntary commitment in the prison context.

In holding the Nebraska statute unconstitutional the District Court had based its conclusion in part on the fact that commitment creates stigmatizing consequences. Justice White agreed with this conclusion, stating that "commitment to a mental hospital" has "adverse social consequences."[75] He offered two case citations to support this assertion. First, he quoted Justice Burger's consequences language in *Addington*.[76] Second, he cited to a statement in a case decided by the Supreme Court earlier that year, to be discussed more in Part IV.c.ii, *infra*.[77] In this case, the Court stated that commitment, in this case the commitment of a child, might trigger some negative, social consequences "because of the reaction of some to the discovery that the child has received psychiatric care."[78] To substantiate this conclusion, the Supreme Court in that case had cited to the same "adverse social consequences" language in *Addington*.

Addington and *Vitek* were landmark decisions in mental health law. For the first time the Supreme Court held that the stigma of mental health commitment, in both the civil and criminal contexts, is real and so damaging to liberty that it was to be considered in procedural due process analysis.

d. Stigma of mental health commitment in procedural due process since Addington and Vitek

While *Addington* and *Vitek* importantly clarified that the stigma of mental health commitment was relevant to procedural due process analysis, these cases left a number of questions related to the application of these standards unanswered. First, the Court did not clarify which consequences of stigma were relevant to the analysis. In *Addington*, Justice Burger referred to the "adverse social consequences" that result from commitment, but then went on to say such consequences may accurately be labeled 'stigma' generally.[79] Justice White merely referred to "the stigmatizing consequences of a transfer to a mental hospital for involuntary psychiatric treatment" without additional clarification. As discussed in Part II.b, *supra*, there is both public and internalized stigma which can result in various, negative consequences. The justices did not specify

---

combination of stigma and a specific type of consequence–the 'mandatory behavior modification' involved in mental health treatment–associated with that stigma. As under Paul, stigma must accompany the condition, just as a particular type of condition must accompany the stigma, in order for a liberty interest to exist. In Vitek, the Court noted that the conditions that Mr. Vitek experienced in the mental institution in which he was confined, considered alone, 'might not constitute the deprivation of a liberty interest retained by a prisoner.'")

75 *Vitek*, 445 U.S. at 492.

76 *Id.* (citing *Addington*, 441 U.S. at 425–26).

77 Parham v. J. R., 442 U.S. 584 (1979).

78 *Id.* at 600 (citing *Addington*, 441 U.S. at 425–26).

79 *Addington*, 441 U.S. at 425–26.

which of these consequences judges are to consider in the procedural due process analysis because they did not engage in any discussion of these specific consequences.

What is more, because both justices spent very little time discussing how they arrived at their conclusions that commitment causes stigma, it was left unclear how broadly these conclusions apply. Neither case explained, for example, whether a commitment order for several days would result in the same stigma as a commitment order for a longer period of time. In *Vitek*, Justice White did not clarify whether the stigma to which he referred was that in the eyes of the general public or that in the eyes of prison population. He did not clarify whether this stigma only attached because the plaintiff was transferred to a facility outside of the prison system or whether it would attach if transferred to any mental health facility.

Finally, it is not obvious from either decision how stigma fits into the overall procedural due process analysis. In Addington, Justice Burger noted the existence of the stigma and consequently upheld the use of an intermediate evidentiary standard but did not state explicitly what role stigma should play in the procedural due process analysis. In Vitek, Justice White held that stigma coupled with mandatory treatment implicated a liberty interest, akin to the "stigma plus" standard established in Paul. Yet it is not clear from Vitek whether any plus factor, such as demonstrable proof of any of the stigmatizing consequences of stigma would be sufficient to implicate a liberty interest, or whether under this standard, standard mandatory treatment is necessary to trigger procedural due process protections.

III. METHODOLOGY AND DESCRIPTION OF THE SAMPLE

a. Opinion collection and selection

To determine how federal judges have treated the stigma of mental health commitment in procedural due process analysis since *Addington* and *Vitek*, a systematic review was conducted of all federal judicial opinions that have discussed this topic since *Addington* was decided on April 29, 1979. Specifically, the search identified all federal cases, both published and unpublished, that discussed: stigma and related concepts (such as social consequences and shame), involuntary commitment and related concepts (such as involuntary treatment and inpatient commitment), and mental health and related concepts, within a single paragraph.[80] Search criteria were developed through reading case law, to

80 To find these opinions, a search was conducted in Westlaw, limiting to all federal jurisdictions, using the following search criteria: ((psychol! psychiat! personalit! mental!) /3 (disorder! ill! health! disabil! disease! diagnos!)) /p stigma! "social costs" "social consequences"

determine the terms judges use in this context, as well as literature related to mental illness, mental health commitment, and stigma. This search yielded 206 opinions.

From these 206 opinions, the study sample was selected based on four criteria. First, the sample was limited to those opinions issued after *Addington*. Second, opinions that included all of the search terms but did not actually discuss stigma in the context of mental health commitment were removed. These opinions might have, for example, discussed the mental health history of defendants, "commitment" of certain crimes, and the stigma of arrest. Or, these opinions may have presented issues related to the stigma of mental health commitment, say in a background section, but ultimately did not discuss the substance of the issues because they were decided on procedural grounds. These opinions may have even cited to the holdings in *Addington* and *Vitek* but did not include any larger discussion of mental health commitment. Many of these opinions were 42 U.S.C. § 1983 actions unrelated to mental health commitment brought by prisoners who merely analogized their situations to that described in Vitek, often in a footnote.[81] Ultimately, these opinions were all removed.

Third, opinions discussing issues related to sex offender treatment and labeling were removed. These opinions contained the search terms, because a number of Circuits have extended the holding in *Vitek* to apply to prisoners labeled as sex offenders. Although this topic is related to the issue of mental health commitment, these opinions were removed from the study sample to simplify analysis.

Fourth and finally, the sample was limited to those opinions that discussed procedural due process. Although the search criteria yielded opinions that discussed the stigma of mental health commitment in a variety of legal contexts, including substantive due process, equal protection, the Americans with Disabilities Act, and the Second Amendment, for this Note, the scope was limited to those opinions that discuss this issue in the context of procedural due process. Following these exclusions, the study sample consisted of fifty-three opinions. Table 1 shows how many opinions were excluded at each step of the opinion selection.

"scarlet letter" shame embarrassment disgrace curse /p commitment hospitalization (commit! /3 (civil! inpatient mental involun!)) "compelled treatment" "involun! treat!" "inpatient treatment" "mental hospital!" "involuntarily admit!".

81 Twelve (unpublished) opinions that included the following language as their only discussion of the relevant issue were excluded: "Vitek v. Jones, 445 U.S. 480, 493-94 (1980) (prisoner possesses liberty interest under the Due Process Clause in freedom from involuntary transfer to state mental hospital coupled with mandatory treatment for mental illness, a punishment carrying 'stigmatizing consequences' and 'qualitatively different' from punishment characteristically suffered by one convicted of a crime)."

Table 1. Opinion Selection

	No. of Opinions
1. All federal opinions containing search criteria,	206
2. Decided after Addington v. Texas (April 29, 1979),	185
3. That discuss stigma in the context of mental health commitment or sex offender treatment,	96
4. Limited to mental health commitment,	61
5. Limited to discussion in the context of procedural due process.	53

b. Content analysis

The remaining fifty-three opinions[82] were reviewed using ethnographic content analysis. This method required reviewing opinions without particular categories in mind, developing categories, and then re-reading the opinions to categorize them by the themes that emerged. To implement this methodology, all discussion of stigma of mental health commitment in the context of procedural due process from the opinions was identified and collected. Once this information was collected from all the opinions in the sample, it was reviewed to determine what similarities and differences existed between the opinions. These findings emerged into themes and each of the fifty-three opinions was assigned one or more of these themes, as will be discussed further in Part IV, *supra*.

There are of course limitations to this study. While this study focuses on federal courts, much of civil and criminal commitment occurs in state courts. This study does not account for how state court judges engage with the stigma of mental health commitment. Additionally, the information collected concerns judges' discussion of stigma in the context of procedural due process rather than case outcomes. While in general judges seemed to deny plaintiffs procedural due process protections, this information was not recorded systematically, because there are so many variables that could affect this outcome. As discussed above, this study was limited to the context of procedural due process. Findings do not necessarily translate to how judges engage the stigma of mental health commitment in other legal contexts. Finally, while the sample includes both published and unpublished opinions, it does not account for those cases in which judges have chosen not to write opinions at all.

82 *Vitek* is among the fifty-three opinions included in the sample since it was decided roughly a year after *Addington*.

c. Description of the sample

The fifty-three opinions analyzed were decided over the years 1979 to 2015. The first case was decided on June 20, 1979 and the last on February 10, 2015. The number of cases was relatively evenly distributed over time, although fewer seem to have been decided in the 1990's than in the other three decades in the sample. Figure 1 shows the number of opinions in the sample decided by year.

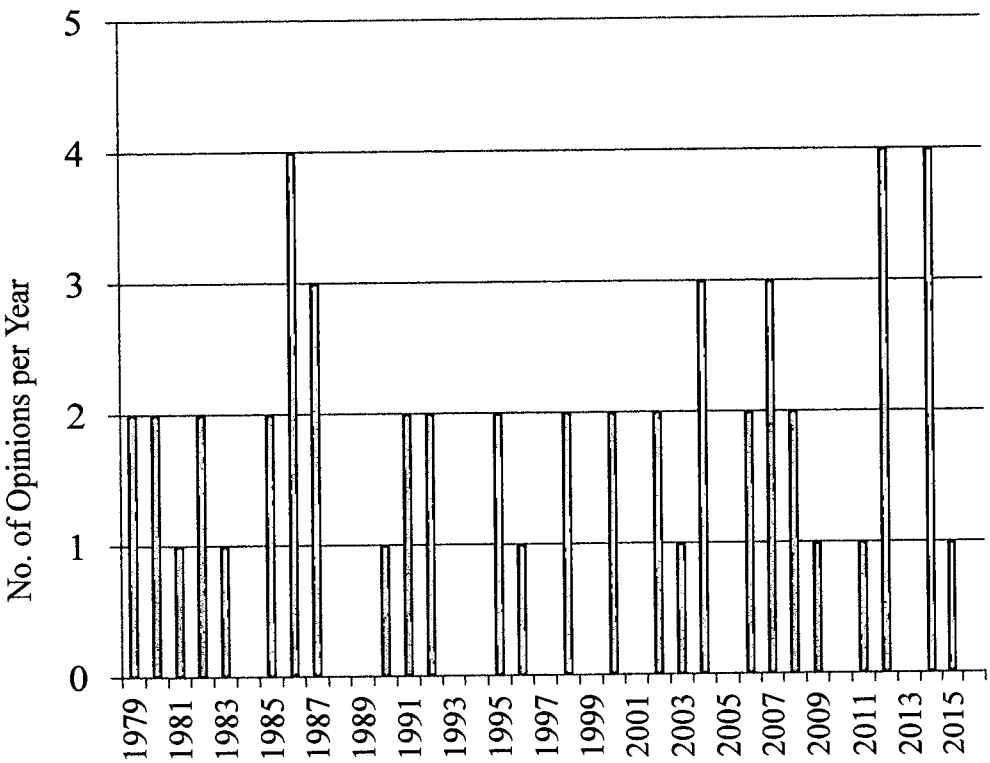

Figure 1. No. of Opinions by Year

The sample includes at least one opinion from every circuit as well as four Supreme Court cases, including *Vitek*. The opinions were also relatively evenly distributed by court type: roughly half were trial-court opinions and half appellate opinions. Table 2 shows the number of opinions decided by circuit, in total and broken out by whether the case was decided by a District Court or the Court of Appeals for that circuit. The court information provided by Westlaw was used to determine the circuit from which each opinion came. The sample includes several cases that were appealed and heard in multiple courts in the

sample, so there are multiple opinions in the sample for the same case.

Table 2. No. of Opinions by Circuit and Court Type

Circuit	Total	Court of Appeals	District Courts
First Circuit	2	1	1
Second Circuit	8	3	5
Third Circuit	9	2	7
Fourth Circuit	4	2	2
Fifth Circuit	4	2	2
Sixth Circuit	6	3	3
Seventh Circuit	1	1	0
Eighth Circuit	4	2	2
Ninth Circuit	3	1	2
Tenth Circuit	6	4	2
Eleventh Circuit	1	1	0
D.C. Circuit	1	0	1
Supreme Court	4	-	-
Total	53	26	27

Finally, cases in the sample pertained to both civil commitment and criminal commitment. Cases were occasionally difficult to categorize. For example, if a person was detained by the police for psychiatric evaluation, this was categorized as a case relating to civil law, because the detainment is not considered an arrest. On the other hand, cases related to criminal defendants pleading not guilty for reason of insanity were categorized as criminal, even though, in some states, such defendants are subsequently committed under civil commitment laws.

Table 3. No. of Opinions in Civil and Criminal Cases

Case type	No. of Opinions
Civil	18
Criminal	35
Total	53

IV. RESULTS

Across the fifty-three opinions reviewed, judges consistently spent very little time discussing the stigma of mental health commitment relative to other issues. Within this limited discussion of mental health stigma, four main themes emerged. First, many opinions in the sample did not discuss stigma beyond restating the conclusions drawn in *Addington* and *Vitek* and either applying their holdings or distinguishing the facts at bar from those in *Addington* and *Vitek*. Among these opinions it was often unclear what role stigma played in the overall procedural due process analysis. There were also opinions that contained somewhat more extended discussion of stigma, and among these discussions, three themes emerged. First, there were opinions that compared the stigma of mental health commitment with stigma resulting from other circumstances. Second, there were opinions that contained more involved discussion of either the consequences or causes of stigma. Finally, some opinions stated explicitly that the stigma of mental health commitment and related issues were so obvious, there was no need to discuss them more broadly - despite the supposed obviousness of the stigma, many of these opinions found procedural due process was not required. None of even those opinions with a somewhat expanded discussion of stigma engaged the full scope of the harm caused by the consequences of stigma.

a. Discussion of stigma limited to quoting Addington and Vitek

Many of the opinions in the sample contained almost no discussion of stigma other than to cite to the consequences language in *Addington* and *Vitek*.[83] These cases arose in both the civil and prison contexts across the entire time period covered by the sample, although more seem to have been filed more recently. Among some of these cases, it was clear that the presence of stigma triggered or would trigger additional procedural protections.[84]

In many other cases, because there was so little additional discussion, it was not clear what role the stigma ultimately played in the judge's decision to grant or deny due process protections.[85] For example, in an opinion by the then Northern District of New York, the court denied defendants' motion for summary judgment against a prisoner claiming that his due process rights were violated when transferred to the mental health treatment wing of the prison without due

83 *Addington*, 441 U.S. at 426; *Vitek*, 445 U.S. at 494.

84 *E.g.*, United States v. Visinaiz, 96 F. App'x 594, 597 (10th Cir. 2004); Bucano v. Sibum, No. 3:12-CV-606, 2012 WL 2395262 (M.D. Pa. June 25, 2012).

85 *E.g.*, Doe v. Gallinot, 657 F.2d 1017 (9th Cir. 1981); United States v. Barajas, 331 F.3d 1141, 1147 (10th Cir. 2003); Cummings v. Darsey, No. CIV.A. 06-5925 (RBK), 2007 WL 174159, at *4 (D.N.J. Jan. 16, 2007).

process. Yet while the court ultimately decided that there were outstanding questions of fact that made summary judgment inappropriate, it is not clear what role stigma played in this decision or what role the judge believed stigma plays in procedural due process analysis more generally. The judge referenced stigma in two places in the opinion. First, the judge discussed *Vitek* but cited the case for the proposition that the plaintiff was entitled to prove he had a mental illness before "suffering the stigmatizing effects of transfer to a mental institution" rather than to discuss the stigma of the transfer itself.[86] Later in the discussion, the judge referenced stigma again in addressing the defendant's contention that being transferred to the mental health wing was better than being placed in protective custody and similar to remaining in the general population. He stated: "certainly from the plaintiff's point of view, the APPU [the mental health treatment wing] is less desirable than the general population, and it is claimed it has stigma attached to it by the general population inmates."[87] Yet, rather than suggesting that this stigma, coupled with mandatory treatment, implicated a liberty interest, per *Vitek*, the judge went on to undercut the defendants' point on other grounds. Based on this brief discussion, it was not clear how, in the judge's view, stigma fit within the procedural due process analysis in general.

Many other opinions also confined discussion of stigma to references to *Addington* and *Vitek*, but ultimately held procedural protections were inappropriate by distinguishing the facts of the case at bar from those in those two cases. In these cases, judges generally distinguished from *Addington* and *Vitek* without going into whether the facts of the instant cases could in themselves result in stigmatic consequences or, if they did not, why they did not.[88] For example, in a case before the District Court of Idaho, plaintiff David Tyler Hill, who had been incarcerated by the Idaho Department of Corrections (IDOC) at the Idaho Maximum Security Institution (IDSI), brought a 42 U.S.C. § 1983 action against the IDOC and its chief psychologist.[89] Specifically, Mr. Hill challenged his transfer to an area of the IDSI designated for mental health treatment without a hearing.[90] In considering whether his transfer implicated procedural due process, the judge cited to the *Vitek* "stigmatizing consequences" language but then distinguished Mr. Hill's situation from that in *Vitek*. He explained that Mr. Hill's transfer was different than the transfer in *Vitek*, because Mr. Hill never left IDOC facilities, whereas in *Vitek* the plaintiff was transferred

86 Flowers v. Coughlin, 551 F. Supp. 911, 915 (N.D.N.Y. 1982).

87 *Id.* at 916.

88 *See, e.g.*, Pierce v. Blaine, 467 F.3d 362, 371 (3rd Cir. 2006) (distinguishing from *Vitek*, because the judge determined that the plaintiff in *Vitek* had been transferred for an indefinite period of time while the plaintiff in the instant case was transferred for several weeks for psychiatric evaluation); Green v. Dormire, 691 F.3d 917, 922 (8th Cir. 2012) (same).

89 Hill v. Reinke, No. 1:13-CV-00038-BLW, 2014 WL 7272939 (D. Idaho Dec. 18, 2014).

90 *Id.* at *2.

out of a facility run by the Department of Corrections and into a "state agency run hospital."[91] The district judge did not explain why this type of transfer would be less stigmatizing nor did he examine the potentially stigmatizing consequences of Mr. Hill's transfer.[92] Ultimately, the judge concluded that the transfer did not implicate a liberty interest and the court granted defendants' motion for summary judgment.[93]

Judges distinguished from *Vitek* on other grounds and did not discuss why or if these distinguishing factors affected the stigma of the commitment. One such factor was length of commitment. According to these opinions, the plaintiff in *Vitek* was transferred to a state run hospital for an indefinite period of time[94] and so judges did not apply *Vitek* in situations in which plaintiffs were committed for finite amounts of time, for example, for several weeks for psychiatric evaluation.[95] Judges did so without discussing why this type of commitment would be less stigmatizing than commitment for an indefinite amount of time.

Among these opinions, there were some with very minimal discussion of stigma that found that procedural due process protections were or would be required, but more often judges distinguished from the facts in *Addington* and *Vitek* and determined that procedural due process protections were not appropriate with little discussion.

b. Comparison to stigma created by other circumstances

In a number of opinions in the sample, the discussion analogized the stigma of mental health commitment to the stigma associated with other circumstances. Judges examined a number of other potentially stigmatizing circumstances. This section begins with an extended discussion of the comparison made to the stigma of insanity pleas, because the issue split two Circuits and was ultimately decided by the Supreme Court, in one of the four opinions in the sample. The Supreme Court subsequently applied its ruling on this issue in another one of the four opinions in the sample.

i. Insanity pleas

In 1980, the Second Circuit considered a due process challenge to commitment proceedings following a determination that a defendant was not

---

91 *Id.* at *18.

92 *Id.*

93 *Id.* at *1.

94 In fact, the statute at issue in *Vitek* provided that in order to keep a prisoner committed after their sentence has elapsed, the hospital must hold a civil commitment hearing. *Vitek*, 445 U.S. at 484.

95 *E.g.*, *Pierce*, 467 F.3d at 371; *Green*, 691 F.3d at 922.

guilty of criminal charges by reason of insanity.[96] Per Connecticut law, after the defendant, Mr. Warren, was acquitted by reason of insanity, a hearing was held to determine whether he was a danger to himself or others and therefore should be committed to a mental health facility. At the hearing, it was determined, based on a preponderance of the evidence, that he was a danger and he was committed.[97] He petitioned the court for his release because he argued that this evidentiary standard, used at both his commitment hearing and subsequent release hearings, violated procedural due process.[98]

In considering the challenge, the Second Circuit took up the liberty interest and specifically the issue of stigma associated with mental health commitment in this situation. The court determined that commitment that follows from a pleading of not guilty by reason of insanity does not result in stigma, because the person is already stigmatized. The Court seemed to suggest that the defendant had reached a sort of stigma ceiling. The Second Circuit wrote: "[a]ny stigma resulting from the label 'mentally ill and dangerous' certainly attached at the time the accused was found not guilty by reason of insanity. Additional stigma which might result from subsequent commitment to a mental hospital must be regarded as minimal, if any."[99] The Court did not provide any explanation for this conclusion.

Two years later, the Fifth Circuit considered the same question but disagreed with the Second Circuit, holding that a defendant that pleads not guilty by reason of insanity can become further stigmatized through commitment.[100] The Fifth Circuit interpreted the Second Circuit's holding as stating that the initial stigma that results from pleading not guilty by reason of insanity results from the "judicial determination . . . that they [the defendants] committed a crime and that no additional stigma attaches upon commitment."[101] This conclusion, the Fifth Circuit stated, was inconsistent with the holding in *Vitek*, because there the Supreme Court determined that a prisoner, an individual that has been convicted of a crime, can still face additional stigma upon transfer to a mental hospital.[102] It is possible the Fifth Circuit misinterpreted the Second Circuit's holding. The initial stigma referred to by the Second Circuit seems to have been that which results from the judicial determination that a defendant is not responsible for a crime because he is insane, rather than that from a judicial determination that an individual committed a crime. While this seems to be the more likely

96 Warren v. Harvey, 632 F.2d 925 (2d Cir. 1980).

97 *Id.* at 929.

98 *Id.* at 931.

99 *Id.* at 931–32.

100 Benham v. Edwards, 678 F.2d 511, 524–25 (5th Cir. 1982), *cert. granted, judgment vacated sub nom.* Ledbetter v. Benham, 463 U.S. 1222 (1983).

101 *Id.*

102 *Id.*

interpretation, it is hard to be sure since the Second Circuit spent so little time on the discussion, and, regardless, the Fifth Circuit clearly thought otherwise. Ultimately, the Fifth Circuit held that additional stigma could result from a transfer from prison to a mental health facility after pleading not guilty by reason of insanity.

The Supreme Court addressed the issue a year later.[103] The Court considered the issue of whether additional stigma could result from commitment following an insanity plea, in a footnote, and agreed with the Second Circuit. Footnote sixteen of Justice Powell's opinion stated only that: "[a] criminal defendant who successfully raises the insanity defense necessarily is stigmatized by the verdict itself, and thus the commitment causes little additional harm in this respect."[104] The Court seemed to endorse this idea of a stigma ceiling in this context, although it did so without reference to case law or external evidence. Justice Brennan, in dissent, commented on this conclusion, but did not disagree with it.[105] He stated only that Justice Powell put too much emphasis on the lack of additional stigma in his due process analysis and in fact there should be more emphasis place on the physical intrusion and restraint placed on committed individuals.[106] This was the first time since *Vitek* the Supreme Court directly addressed the role of the stigma of mental health commitment in procedural due process.

Shortly after this case was decided, the Fifth Circuit case discussed above, was remanded and vacated.[107] This issue arose in two other cases in the sample. Nearly a decade later, the Supreme Court took up another case related to commitment following an insanity plea and again held that no additional stigma resulted from commitment.[108] A decade after that, in a case before the Tenth Circuit, the Court also applied the Supreme Court's conclusion.[109]

ii. Criminal Conviction

Judges also compared the stigma of mental health commitment with that of criminal conviction, separately from pleading insanity. One of these opinions provides an example of a judge looking to cases beyond *Addington* and *Vitek* to inform a conclusion related to the stigma of mental health commitment. In a case before the Fourth Circuit, plaintiff Theresa Gooden brought a 42 U.S.C. § 1983

103 Jones v. United States, 463 U.S. 354 (1983).

104 *Id.* at 367 n.16.

105 *Id.* at 371 (Brennan, J., dissenting).

106 *Id.*

107 *Ledbetter*, 463 U.S. at 1222.

108 Foucha v. Louisiana, 504 U.S. 71, 114 (1992).

109 United States v. Weed, 389 F.3d 1060, 1068 (10th Cir. 2004) (citing Jones, 463 U.S. at 367 n.16).

action against police officers and her county after she was taken from her apartment to a hospital for emergency mental health commitment.[110] In discussing the potential harm that may arise from a seizure for civil commitment, the judge quoted a district court's assessment from 1979 that "such a deprivation can create 'a stigma of mental illness which can be as debilitating as that of criminal conviction.'"[111] This quotation, in turn cited to a 1973 D.C. Circuit case and a 1963 hearing before a Senate Subcommittee.[112] In grappling with this question, whether the stigma of involuntary civil commitment is as "as severe" as criminal conviction, the judge in the 1973 D.C. Circuit case looked to then current studies in addition to then current news stories and Congressional hearings from the previous decade on the issue.[113]

In relying on this case law, the judge was in fact relying on conclusions the judges in those cases drew based on external sources of information on stigma, including studies, news stories, and Congressional hearings. Yet, these sources of evidence, relied upon in 1990, were from the 1960's and 1970's. While it is possible that the stigma of both of these circumstances remained constant in the intervening twenty to thirty years, it is not clear why the judge did not just rely on similar, more current sources.

iii. History of Mental Illness

In other opinions, judges opined on whether a long history of mental illness erases any additional stigma that may be created by commitment. In one such opinion, the judge looked to how juries had thought about stigma in the past to inform his determination of whether the jury's damages award for a six-day commitment without adequate procedural protection was reasonable.[114] The jury had awarded the plaintiff, Robert Marion, $750,000 in compensatory damages for the deprivation of liberty he suffered over the course of his six-day commitment. In determining what amount of compensatory damages were appropriate, the judge compared Mr. Marion's situation to three cases in which

110 Gooden v. Howard Cty., Md., 917 F.2d 1355 (4th Cir. 1990), *opinion superseded on reh'g*, 954 F.2d 960 (4th Cir. 1992).

111 *Id.* at 1363 (quoting Gross v. Pomerleau, 465 F. Supp. 1167, 1173 (D. Md. 1979) (quoting Stamus v. Leonhardt, 414 F.Supp. 439, 444 (S.D.Iowa 1976)). Although, according to Westlaw, the relevant quotation is at *Stamus*, 414 F.Supp. at 449.

112 Stamus v. Leonhardt, 414 F.Supp. 439, 449 (S.D.Iowa 1976)) (" . . . the legal and social consequences of commitment constitute a stigma of mental illness which can be as debilitating as that of a criminal conviction. *See* In re Ballay, 157 U.S.App.D.C. 59, 482 F.2d 648, 668–69 (1973); Hearings on S. 935 Before the Subcomm. on Constitutional Rights of the Senate Comm. on the Judiciary, 88th Cong., 1st Sess., 38 (1963).").

113 In re Ballay, 482 F.2d 648, 668 (D.C. Cir. 1973).

114 Marion v. LaFargue, No. 00 CIV. 0840 (DFE), 2004 WL 330239 (S.D.N.Y. Feb. 23, 2004).

individuals had been committed without having previously been diagnosed as mentally ill. The judge concluded that the cases were distinguishable, and that "Marion's case for damages was significantly weaker."[115] He explained:

> It was undisputed that Marion has had serious mental illness for many years. It seems clear that the other three juries were convinced that the plaintiffs never had any mental illness. . . . Accordingly, the amounts that those plaintiffs received for emotional damages are attributable only in part to the days of confinement, and in large part to the lingering stigma that unfortunately attaches to findings of mental illness. . . .[116]

Consequently, the judge determined that Mr. Marion was entitled to less than the defendants in these cases and less than what the jury had awarded him. The judge reduced the award from $750,000 to $150,000. It seems his decision was driven in part by his conclusion (based on past jury behavior) that stigma attaches when an individual is labeled as mentally ill and if already labeled, additional stigma does not occur upon commitment. Like the person who brings an insanity defense and now faces commitment, Mr. Marion had reached his stigma ceiling and, consequently, was entitled to far less damages for the violation of his procedural due process than if he had not had a history of mental illness. The judge came to this conclusion by considering past jury behavior rather than engaging in fact finding related to the current stigma of mental health commitment.

These opinions provide examples of judges either comparing the stigma of mental health commitment to other types of stigma. Because these types of comparisons were largely not addressed in either *Addington* or *Vitek*, judges were forced to consider other sources of information, or rely on personal opinion, in coming to conclusions on this matter. In general, judges favored looking to information from the past, such as prior case law or past jury behavior, rather than current sources of information, such as recent studies. Additionally, different judges relied on the same sources of information but came to very different conclusions. As discussed above, two appellate courts considered the same question, whether commitment following an insanity plea creates additional stigma, and relying on the same case law, came to entirely different results. The Supreme Court ultimately resolved this issue, but this is one of only a few issues the Court has addressed since it decided *Addington* and *Vitek*. There were many other inconsistencies in how judges were comparing this stigma to other forms of stigma that the Court has not addressed.

---

115 *Id.* at *10.
116 *Id.*

c. *Some discussion of the consequences and causes of the stigma of mental health commitment*

i. *Consequences of stigma*

There were several opinions in the sample that included a broader discussion of the consequences discussed in *Addington* and *Vitek*. In both *Addington* and *Vitek*, the Supreme Court focused on consequences, specifically the "adverse social consequences" and the "stigmatizing consequences" of commitment without additional discussion of specific consequences.[117] Although many opinions did not discuss these consequences any further, there were some in the sample that expanded upon this idea. Several of these opinions kept the discussion very general. For example, in a 1986 D.C. Circuit opinion, the judge stated that the: "personal and social consequences of commitment have a profound impact on a person long after he has been treated and released."[118] He substantiated this conclusion by citing to *Addington*.[119]

There were just a few other opinions that discussed the consequences of stigma in more specific terms, identifying the individual consequences that may result from the stigma associated with commitment. For example, in a 1985 North Carolina District Court case, the judge considered whether due process protections were required for a transfer to a mental health facility within the Department of Corrections. The case came to the court from a magistrate judge who had determined that this type of transfer did not implicate a liberty interest and therefore did not require procedural protections, because unlike in *Vitek* the plaintiff was not transferred outside of the Department of Corrections. The magistrate judge determined that the distinction was dispositive because, even though Judge White did not state so explicitly, the stigma at issue in *Vitek* was that in the eyes of the public rather than that among other inmates.

The District Court judge disagreed with the magistrate judge's interpretation of *Vitek*. He concluded that the transfer did implicate a liberty interest because it created stigma within the prison system, which was as harmful as stigma outside of the prison system. The judge went on to list specific consequences of a transfer to a mental health facility within the prison system: "[d]enial or delay of parole, study release, work release, and gain time jobs."[120] Additionally, he stated

---

117 *Addington*, 441 U.S. at 425–26; *Vitek*, 445 U.S. at 494.

118 Sanderlin v. United States, 794 F.2d 727, 736 (D.C. Cir. 1986).

119 *Id.*

120 Baugh v. Woodard, 604 F. Supp. 1529, 1535 (E.D.N.C. 1985), *aff'd in part, vacated in part*, 808 F.2d 333 (4th Cir. 1987) (the sole issue on appeal was the timing of the hearing required by due process: the District Court had held that such a hearing must take place prior to transfer whereas the Fourth Circuit concluded that the hearing could occur immediately after transfer but before admission to the mental health facility).

that "[t]here is also undisputed evidence that a prisoner returning to the general prison population from a mental health unit are viewed as 'bugs' by other inmates. These prisoners are ostracized and exploited by other prisoners."[121] While these assertions seem to be supported by the research discussed in Part II.b, *supra*, the judge made these assertions without reference to case law or any external evidence. Although many other opinions in the sample considered a transfer within the Department of Corrections, this is one of the few opinions that engaged in a more detailed analysis of the stigmatizing consequences in this context and one of the few to ultimately find that procedural protections were required. These opinions, particularly those that included a discussion of specific consequences, engaged the harm associated with this stigma more than did other opinions and found that procedural protections were appropriate more frequently than those opinions that did not engage this discussion.

ii. Causes of stigma

There were some opinions that included a discussion of the causes of the stigma of mental health commitment. Justice Burger, in *Addington*, did not directly address the causes of the stigma of mental health commitment but did discuss what he saw as causing the stigma associated with mental illness generally. He asserted: "[o]ne who is suffering from a debilitating mental illness and in need of treatment is neither wholly at liberty nor free of stigma." He cited to several articles in psychiatric publications to support this claim.[122] Justice White, in *Vitek*, did not engage in any discussion of the causes of the stigma associated with mental health commitment.

In a Third Circuit opinion, the judge considered an appeal from an award of attorney's fees in a class action brought by six named plaintiffs on behalf of all juveniles who had or would be committed to mental health facilities pursuant to Pennsylvania law by a parent or guardian.[123] In the underlying litigation, plaintiffs had alleged that this law violated both the due process and equal protection clauses of the Fourteenth Amendment. In discussing the liberty interest potentially affected by this type of commitment, the judge quoted another case in the sample, a 1979 Supreme Court opinion that identified at least one cause of the stigma of mental commitment for children: "'commitment sometimes produces adverse social consequences for the child because of the

121 *Id.*

122 *Addington*, 441 U.S. at 429 (citing Paul Chodoff, *The Case for Involuntary Hospitalization of the Mentally Ill*, 133 AM. J. PSYCHIATRY 496, 498 (1976); Carol C. Schwartz et al., *Psychiatric Labeling and the Rehabilitation of the Mental Patient*, 31 ARCHIVES GEN. PSYCHIATRY 329, 334 (1974)).

123 Institutionalized Juveniles v. Sec'y of Pub. Welfare, 758 F.2d 897, 901 (3rd Cir. 1985).

reaction of some of the discovery that the child has received psychiatric care.'"[124] This quotation, which comes from *Parham v. J.R.*, seems to suggest that the stigma associated with commitment is not unique to commitment but would in fact result from any type of mental health treatment. Justice Burger, writing for the Court, followed this assertion with a citation to the "adverse social consequences" language in *Addington*.[125]

In *Parham v. J.R*, Justice Burger elaborated still further on what in his mind causes stigma for individuals facing commitment. The Court was considering a procedural due process challenge to a state civil commitment law and ultimately upheld its constitutionality. In coming to this conclusion, Justice Burger stated that making it more difficult to commit individuals in need of treatment could be the real cause of stigma, because "what is truly 'stigmatizing' is the symptomatology of a mental or emotional illness."[126] To support this contention, he cited to the assertion in *Addington* that to be mentally ill is to never be wholly free from stigma.[127]

Very few opinions in the sample addressed the causes of the stigma of mental health commitment. Those that did seemed to suggest that there is nothing uniquely stigmatizing about commitment, but rather that it is the underlying mental illness or treatment more generally that causes stigma. This in turn prompted these judges to deem procedural protections unnecessary, because the individual would experience the stigma regardless of the commitment.

d. Discussion of the obviousness of issues related to the stigma of mental health commitment (and yet often holding such stigma does not trigger procedural protections).

Finally, some opinions mentioned stigma but used language suggesting that the conclusions related to this stigma were so obvious there was no need for further discussion. In some instances, the obviousness of this stigma would lead judges to require procedural protections, yet more often, judges used this language, engaged in very minimal discussion of stigma, and ultimately held that procedural protections were not required.

For example, in an opinion from the Southern District of New York, the court considered a class action brought by civilly committed individuals arguing that it was a violation of procedural due process that the state did not appoint psychiatrists to assist in retention hearings.[128] Plaintiffs argued that committed

124 *Id.* at 913 (citing *Parham*, 442 U.S. at 600).

125 *Parham*, 442 U.S. at 600 (citing *Addington*, 441 U.S. at 425–26).

126 *Id.* at 601 (citing *Addington*, 441 U.S. at 429).

127 *Id.*

128 Goetz v. Crosson, 769 F. Supp. 132, 133 (S.D.N.Y. 1991), *aff'd in part, rev'd in part*, 967 F.2d 29 (2d Cir. 1992) (affirming the District Court's holding to the extent that in most cases due

individuals were due the same level of procedural protections as were criminal defendants, but the judge was not convinced. The judge conceded "there is an obvious stigma attached to confinement in a mental hospital,"[129] but the interest of the criminal defendant is "almost uniquely compelling."[130] He went on to explain why he found the criminal defendant's interest more compelling than that of the committed individual: the criminal, he asserted, was being punished, not treated and the committed individual was committed to protect society but also to protect himself. Yet, after describing the stigma as obvious, the judge entertained no further discussion of it. He did not consider the specific consequences of the stigma associated with commitment (or, incarceration for that matter). While he recognized the existence of the stigma, he seemed to give it minimal weight in comparison to other factors, without discussing why.

For another example, in the only Supreme Court case in the sample yet to be discussed, the Court engaged the topic of whether additional stigma attaches upon the commitment of a person who had plead not guilty by reason of insanity, the subject of Part IV.b.i, *supra*.[131] In this discussion, Judge White, writing for the Court, used language to suggest the obviousness of the conclusion that additional stigma does not in fact attach. To begin, Judge White applied the conclusion previously drawn by the Court in a footnote in the case discussed above. He wrote, "[s]tigmatization (our concern in *Vitek*) is simply not a relevant consideration where insanity acquittees are involved."[132] Despite this dismissive language, he cited to the Supreme Court case and the Second Circuit Court case that were discussed above to support this assertion. Yet in addition to citing to the Court's own precedent and the Second Circuit case, he also offered his own opinion on the subject.[133] He wrote, "[i]t is implausible, in my view, that a person who chooses to plead not guilty by reason of insanity and then spends several years in a mental institution becomes unconstitutionally stigmatized by continued confinement in the institution after 'regaining' sanity."[134] While this particular question had been previously decided by the Court, Justice White's assertion seemed to bely something else: that there are some conclusions so obvious there

process does not require the state to appoint of a psychiatrist but reversing and remanding back to the District Court to determine whether there may be some cases that are so factually complicated that a psychiatrist expert may be necessary).

129 *Id.* at 135.

130 *Id.* (citing Ake v. Oklahoma, 470 U.S. 68, 78).

131 *Foucha*, 504 U.S. at 71.

132 *Id.* at 114.

133 *Id.* ("As we explained in *Jones*: 'A criminal defendant who successfully raises the insanity defense necessarily is stigmatized by the verdict itself, and thus the commitment causes little additional harm in this respect.' 463 U.S., at 367, n. 16, 103 S.Ct., at 3051, n. 16; see also Warren v. Harvey, 632 F.2d, at 931-932.")

134 *Id.*

is no need to consider them further, to look to external research to corroborate.

While opinions that used this obviousness language did so in different contexts, some judges referring to the obviousness of the stigma itself and other referring to the obviousness of related conclusions, in general, use of the language was associated with very little additional discussion of any of the questions that were left unaddressed by *Addington* and *Vitek*. Often in these opinions judges would go on to find that the presence of stigma did not require procedural protections.

V. DISCUSSION

a. Judges have an incomplete view of the stigma of mental health commitment.

Among the fifty-three cases analyzed, there was variability in how and whether each opinion discussed stigma. There were those opinions that merely re-stated or cited to the language in either *Addington* or *Vitek* without any further discussion. There were those that drew comparisons between this type of stigma and other stigma and therefore engaged in longer discussion. Others engaged in some discussion about specific consequences of stigma or a broader discussion of consequences of stigma generally and other traced possible sources for that stigma. Some stated explicitly that no discussion was required because the stigma that results from commitment and other related issues are so obvious.

Yet, despite this variability, among all fifty-three opinions, judges consistently failed to consider the full consequences of stigma associated with mental health commitment. As discussed in Part II.b, *supra*, there are many more consequences to the stigma of mental illness and commitment than are described in any of the opinions in the sample. *Addington*, for example, references "adverse social consequences," but it is not clear whether this was meant to include all harms that result from stigma, such as employment discrimination, reduced income, and decreased ability to live independently. *Vitek* may have expanded the analysis to include all "stigmatizing consequences" but did not go into a discussion of what those consequences were. Neither opinion stated explicitly what about commitment causes the stigma: whether is it the mental illness, the treatment, or, in the prison context, the nature of the transfer itself. Accordingly, judges frequently distinguished from both *Addington* and *Vitek* based on the facts of a particular situation. Judges, for example, distinguished from *Vitek*, by determining that stigma only attaches when an incarcerated person is physically transferred out of a prison facility into a mental hospital or when that person is transferred for an indefinite amount of time. In distinguishing in this way, these judges failed to consider the stigma created by other circumstances.

Those judges that did engage in broader discussions of the consequences and

causes of stigma related to mental health commitment still failed to engage the full extent of this stigma. In cases in which judges considered one type of stigma relative to another type of stigma, judges determined, for example, that a person who was already incarcerated could not face further stigmatization if committed, without providing evidence to support that claim. In other opinions, judges failed to adequately address what created the stigma associated with mental health commitment, some determining that mental illness itself is the cause, others the manifestation of symptoms, and most providing no explanation at all. Across all fifty-three cases, judges did not consider the full scope of the harm associated with the stigma of mental health commitment.

b. A systematic bias against committed people bringing due process challenges.

Judges' incomplete understanding of stigma has created a systematic bias against individuals bringing procedural due process claims in the mental health commitment context. In *Addington*, the Supreme Court stated that the adverse social consequences of mental health commitment were relevant to the procedural due process analysis. *Vitek* further clarified in stating that, in the criminal context, the stigmatizing consequences of a transfer to a mental health facility, coupled with mandatory treatment, implicated a liberty interest and therefore triggered due process protections.

Yet, as discussed above, when judges have applied these standards they have not considered the full scope of the harm associated with stigma of mental health commitment, because of an incomplete view of that stigma. While some judges found that the presence of stigma compelled procedural protections, many did not.

By systematically underestimating the stigmatizing consequences of mental health commitment, judges have systematically underestimated the liberty interest itself implicated by mental health commitment. This in turn has meant that judges have consistently required less rigorous procedural due process protections for individuals subject to commitment orders. By requiring less rigorous procedural protection, these individuals are at greater risk for erroneous commitment. By undervaluing the harm these individuals suffer as a result of the stigma of mental health commitment, judges have increased the likelihood that individuals are subject to inappropriate commitment orders.

c. Judges have been overly deferential to Supreme Court fact finding

Two related issues seem to drive judges' incomplete engagement with the stigma of mental health commitment: overreliance on case law and insufficiency of information. This first issue, more specifically put, is that judges seem to be

overly deferential to the conclusions drawn by the Supreme Court in *Addington* and *Vitek*. That is, most judges merely recited the Supreme Court's conclusion that commitment causes stigma or if they did engage in a broader discussion of stigma they did so without engaging in their own fact finding related to this stigma, as if to suggest that the Supreme Court has already done most of the work, no need to do too much more.

Lower court judges should of course adhere to stare decisis with respect to legal rules, yet the conclusion that commitment leads to stigma is not, per se, a legal rule. Scholar Allison Orr Larsen and others have described this type of conclusion as a legislative fact, that is, "a generalized fact . . . [that] provides descriptive information about the world that judges use as foundational building blocks to form and apply legal rules."[135] Judges draw these factual conclusions based on many different sources, including information provided by parties' briefs, amicus briefs, and their own knowledge and assumptions about the world.[136] Lower courts choosing to accept and apply these conclusions is what Larsen refers to as following "factual precedent"[137] and it is not clear in all cases that lower courts must in fact do so.

In some cases, those in which a legal rule is dependent upon a factual finding of the Court, it is clear that lower courts must accept and follow the Supreme Court's factual precedent. To illustrate this point, Larsen points to one of the Court's conclusions in Citizens United v. Federal Election Commission.[138] After considering the record in the that case as well as the companion case, McConnell v. Federal Election Commission, Justice Kennedy, writing for the majority, concluded that politics are not corrupted by corporate money in campaigns.[139] When the Court ultimately granted First Amendment protection to corporations for such speech, the protection was based in part on this conclusion. In a subsequent case, the Supreme Court of Montana was presented with different evidence and ultimately held that corporate spending could (and did) influence politics. The Supreme Court quickly granted certiorari and reversed the Supreme Court of Montana in a several-paragraph, per curium opinion.[140] Although the Supreme Court of Montana may have had different evidence that could have reasonably supported a different factual conclusion, the Supreme Court made clear that its conclusion was controlling.

Larsen concedes that it is necessary for lower courts to defer to factual

---

135 Allison Orr Larsen, *Factual Precedents*, 162 U. PA. L. REV. 59, 72 (2013).

136 Allison Orr Larsen, *Confronting Supreme Court Fact Finding*, 98 VA. L. REV. 1255, 1258–60 (2012).

137 Allison Orr Larsen, *Factual Precedents*, *supra* note 135 at 72.

138 Citizens United v. Fed. Election Comm'n, 558 U.S. 310 (2010).

139 Allison Orr Larsen, Factual Precedents, *supra* note 135 at 94.

140 *Id.*

precedent in cases such as Citizens United. If a legal rule is dependent upon the Court's factual finding, as it was in Citizens United, allowing lower courts to reconsider that conclusion would essentially re-litigate the entire issue and could "run the risk of chaos or at least a serious weak spot in the Supreme Court's authority."[141]

Like that espoused in Citizens United, the legal rules in *Addington* and *Vitek* are in one sense dependent upon a factual conclusion made by the Court. Generally put, the legal rule that stigma should be considered in the procedural due process analysis in the context of mental health commitment is based upon the factual conclusion that commitment creates stigmatizing consequences. If lower court judges did not accept this factual conclusion, they could then conclude that stigma need not be considered in procedural due process. This would lead to the chaos of which Larsen warns. As such, lower courts cannot and should not do as the Supreme Court of Montana did, and re-litigate the issue of whether mental health commitment causes stigma. And, based on my review, judges are not doing this, to the extent that they are not explicitly contradicting the premise.

Yet, Larsen also concludes that lower courts are overly deferential to the Supreme Court's factual findings in situations they really should not be. She argues that the Supreme Court is no better equipped than are lower courts to engage in legislative fact finding and that, in general, lower courts reconsidering legislative facts allows for more flexible legal rulings without disrupting legal precedent.

This too applies to *Addington* and *Vitek*. While the Supreme Court resolved the question of whether commitment has stigmatizing consequences, as discussed in Part II.d, *supra*, the Supreme Court did not resolve many other questions relevant to the application of the *Addington* and *Vitek* rules. The Supreme Court did not discuss what the consequences of stigma are or, relatedly, what weight to apply to stigma in the overall procedural due process analysis. The Supreme Court did not address what causes the stigma and therefore in what situations this stigma may or may not occur. In *Vitek* specifically, Justice White did not clarify whether the relevant stigma was that in the eyes of other prisoners or the public at large. Judges have deferred to the Supreme Court's factual findings with respect to all of these questions even though they did not in fact resolve them. The fact that the Supreme Court did not consider these questions does not mean that lower courts should not consider these questions. In fact, to properly apply this test, lower courts must consider these questions.

While *Addington* and *Vitek* clarified that judges must consider the stigma of mental health commitment in procedural due process, these rulings did not

141 *Id.* at 108.

properly clarify how to do so. In order to properly implement these standards, in order to properly account for the full harm associated with the stigmatizing consequences of the stigma of mental health commitment, judges must do more than rely on the Supreme Court's fact finding in *Addington* and *Vitek*. Instead, judges must engage in their own fact finding to determine the full harm associated with the stigmatizing consequences the Supreme Court has instructed must be considered in the procedural due process analysis.

d. Remedies to assist in judges' fact finding related to the stigma of mental health commitment.

Accepting that judges must do more to implement the legal rules espoused in *Addington* and *Vitek* by determining what consequences result from mental health commitment, highlights the second issue that seems to drive judges' incomplete view of information: that is, insufficiency of information. If judges are to engage in fact finding related to the consequences of stigma, judges need access to that information and the expertise to make sense of it.[142]

One potential solution could be to divert procedural due process challenges to commitment to courts with particular expertise in mental health. The Department of Justice works with the Substance Abuse and Mental Health Services Administration (SAMHSA) to administer the Mental Health Courts Program, an integrated system of judges, lawyers, and mental health professionals that deals specifically with nonviolent offenders with mental health diagnoses.[143] The purpose of this program is to better serve these individuals by requiring specialized training for all those involved in the program, offering voluntary treatment in exchange for adjusting sentencing or even dropping charges, and coordinating case management with a mental health professional. The program currently operates roughly forty courts around the country.[144] These courts' jurisdiction could be broadened to include constitutional challenges to commitment orders. There could certainly be some benefits created by requiring all procedural due process challenges to mental commitment to be deferred to mental courts. These judges would have more direct and consistent access to mental health experts and would therefore have more information about mental

142 As discussed in Part IV, *supra*, there were some judges in the sample that relied on sources of information on the stigma of mental health commitment other than *Addington* and *Vitek*, yet these sources, such as prior case law or a judges' opinions, generally did not reflect the current research on the subject. For a fuller discussion of courts' reliance on antiquated information related to mental illness, see Joanmarie Ilaria Davoli, *Still Stuck in the Cuckoo's Nest: Why Do Courts Continue to Rely on Antiquated Mental Illness Research?*, 69 TENN. L. REV. 987 (2002).

143 OFFICE OF JUSTICE PROGRAMS, *Mental Health Court Programs*, https://www.bja.gov/ProgramDetails.aspx?Program_ID=68.

144 *Id.*

health in general. These mental health professionals may also have more expertise with respect to the real consequences of stigma about which they could educate judges.

On the other hand, these courts have potential drawbacks. Tailoring sentencing to include treatment may in itself present separate procedural due process issues, given that individuals may feel compelled to accept treatment over jail time without appropriate procedural protections. Additionally, given that these judges would primarily be engaged in trial litigation and sentencing, they may be less-equipped to engage with constitutional matters such as due process analysis than would other federal judges that engage deal with more varied litigation. Second, separating these individuals from the general population of litigants may in fact perpetuate stigma.[145] Finally, mental health professionals may not necessarily have more information about mental health stigma and may therefore not provide judges with the necessary, additional information to adequately implement the *Addington* and *Vitek* standards.

Instead, in matters related to legislative fact finding related to the stigma of mental health commitment, courts could rely on an independently maintained resource, like that discussed by Allison Orr Larsen in her article, *Confronting Supreme Court Fact Finding*.[146] She proposed that rather than relying primarily on amicus curie briefs or in-house research, as the Supreme Court does currently, the Court could rely on resources that aggregate the type of information contained within amicus briefs but reflect a broader range of ideas than are typically reflected in those briefs. This, Larsen argues, would provide the Court information without biasing that information in favor of groups with the resources to compile amicus briefs.[147] This type of resource could be created for the stigma of mental health commitment through collaboration between legal groups, like the American Bar Association, and mental health organizations, such as the American Psychological Association, or even multiple interest groups with differing agendas. This type of resource could allow judges at all levels greater access to information, created and maintained by individuals with the relevant

145 See E. Lea Johnston, *Theorizing Mental Health Courts*, 89 WASH. U. L. REV. 519, 536 (2012) (arguing that mental health courts contribute to the stigmatization of mental illness by suggesting that offenders with mental illnesses lack the ability to control their actions, are so much more vulnerable to recidivism they should be isolated from the general population, and that they cannot be trusted to make their own health care treatment choices).

146 Allison Orr Larsen, *Confronting Supreme Court Fact Finding*, supra note 136 at 1311–12.

147 Larsen referenced a particular resource, the American Bar Association's The Citizen Amicus Project, which no longer exists. *Id.* at 1311. The Native Amicus Briefing Project (NAB) has a similar mission, but focuses on particularly on improving federal judges' understanding of federal Indian law. NAB tracks federal cases that deal with Indian law and drafts and submits amicus briefs in those cases. The organization is run by a small group of attorneys and works with other attorneys, Indian law scholars, law students, and Native organizations. *About Us*, NATIVE AMICUS BRIEFING PROJECT (NAB) (2018), http://nativebrief.sites.yale.edu/about-us.

expertise.

VI. CONCLUSION

This is the first systematic review of federal judicial opinions that discuss the stigma of mental health commitment in the context of procedural due process. Results show that many judges limit their discussion of the stigma of mental health commitment to citations to *Addington* or *Vitek*. While some opinions engaged in broader discussions of what specific consequences result from the stigma of commitment and the sources of that stigma, in general judges articulated an incomplete view of this stigma. This has led judges to consistently underestimate the stigma associated with mental health commitment, resulting in a systematic bias against plaintiffs bringing procedural due process challenges in the context mental health commitment. To address this bias, federal judges must engage in more fact finding about the real and complete consequences of stigma that results from mental health commitment.