ARTICLES

THE HARM PRINCIPLE AND FREE SPEECH

REBECCA L. BROWN*

TABLE OF CONTENTS

* The Rader Family Trustee Chair in Law, USC Gould School of Law. I am grateful to Tabatha Abu El-Haj, Scott Altman, David Cruz, Lee Epstein, Sam Erman, Omar Noureldin, Bob Rasmussen, Fred Schauer, Nomi Stolzenberg, Geoffrey Stone, and David Strauss for their generous engagement on this project. My thanks also go to research assistants Lauren Fishelman, Narain Kumar, Gus Paras, Alex Radisich, and librarian Cindy Guyer for their valuable contributions.

INTRODUCTION

The harm principle allows government to limit liberties as necessary to prevent harm.[1] Does the freedom of speech present an exception to the harm principle? Most American scholars say yes.[2] It is common practice to proclaim proudly that the U.S. Constitution protects speech even when it causes harm. But two tenets of the author of the harm principle himself suggest that, today, this answer may be too glib. For John Stuart Mill, the enhanced protection of speech is only a means to protect thought, and moreover, opinions lose their immunity if they cross over from thought into action.[3] Together, these two points invite us to consider the possibility that the special protection we have come to afford, even to a newly broadened range of speech that goes well beyond thought, may be misplaced. There are cases, I will argue, in which we should be slow to assume that society is necessarily without power to protect itself from harm that expression may cause.

Constitutional law has developed a firm rule prohibiting the regulation of speech based on its content, no matter what the alleged harm might be.[4] This rule, to which I will refer here as the "cardinal rule" of free speech, means that if a restriction turns on what is said or expressed, or on characteristics of an expression, then it is presumptively invalid.[5] The rule can be appreciated, perhaps, as an understandable reaction to a long period in our history in which political dissent was not protected rigorously

1. The harm principle holds "[t]hat the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others." JOHN STUART MILL, ON LIBERTY 80 (David Bromwich & George Kateb eds., Yale Univ. Press 2003) (1859).

2. *See, e.g.*, Andrew Koppelman, *Veil of Ignorance: Tunnel Constructivism in Free Speech Theory*, 107 NW. U. L. REV. 647, 696–97 (2013); Frederick Schauer, *Harm(s) and the First Amendment*, 2011 SUP. CT. REV. 81, 81–82 (discussing the tendency of discourse to downplay harm caused by speech).

3. MILL, *supra* note 1, at 95, 132. There is a rich literature addressing, in more depth, how best to understand John Stuart Mill's views on free speech. *See, e.g.*, GEOFFREY MARSHALL, CONSTITUTIONAL THEORY 157 (H. L. A. Hart ed., 1971) (attributing to Mill the position that speech is punishable if it is "so closely connected with the commission of the act as to be part of it," but never if it is "discussion"); Vincent Blasi, *Shouting "Fire!" in a Theater and Vilifying Corn Dealers*, 39 CAP. U. L. REV. 535, 542 (2011) ("[The harm principle] can justify regulations of speech that do not . . . amount to the total prohibition of the circulation of an opinion."); Frederick Schauer, *On the Relationship Between Chapters One and Two of John Stuart Mill's* On Liberty, 39 CAP. U. L. REV. 571, 592 (2011) (challenging Blasi's reading, arguing that, for Mill, liberty of thought protects even harmful communications).

4. *See infra* Part III.

5. R.A.V. v. City of St. Paul, 505 U.S. 377, 382 (1992).

enough.[6] But the cardinal rule has outgrown that core mission. After its articulation in 1972, the cardinal rule was "increasingly invoked,"[7] quickly playing a "significant role,"[8] a "central role,"[9] and by 2000 was recognized as "the central inquiry" in free speech law.[10] Its use increased, not only in frequency, but also in the scope of what was covered. As of 2016, there has been no case in which a majority of the Supreme Court has found a government interest sufficient to redeem a law that it had analyzed as content-based.[11] Like the Latin root from which its name derives, the cardinal rule is indeed that on which all things First Amendment now hinge.[12]

But the rule has taken on a life of its own, with no serious examination of whether its ever-expanding reach serves the values it was designed to protect. It goes well beyond the protection of ideas and applies even when there is no impact on the self-governing quest for truth. The absolute protection now extends to any restrictions involving subject matter, genre, or topic of expression, and most recently even immunizes the underlying nature of an expression, such as whether it is violent, from legislative

6. *See infra* Part III.A.

7. Geoffrey R. Stone, *Content Regulation and the First Amendment*, 25 WM. & MARY L. REV. 189, 189 (1983).

8. Paul B. Stephan III, *The First Amendment and Content Discrimination*, 68 VA. L. REV. 203, 204 (1982).

9. Geoffrey R. Stone, *Content-Neutral Restrictions*, 54 U. CHI. L. REV. 46, 46 (1987).

10. Erwin Chemerinsky, *Content Neutrality as a Central Problem of Freedom of Speech: Problems in the Supreme Court's Application*, 74 S. CAL. L. REV. 49, 49 (2000). *See also* Barry P. McDonald, *Speech and Distrust: Rethinking the Content Approach to Protecting the Freedom of Expression*, 81 NOTRE DAME L. REV. 1347, 1362 (2006) ("[The] principle of absolute neutrality has continued to exert a heavy influence on the Court's attitude towards the constitutionality of . . . [content-based] regulations.").

11. *See* McDonald, *supra* note 10, at 1365 & n.63. McDonald found that "[i]n the twenty cases commencing with C*arey* [*v. Brown*, 447 U.S. 455 (1980)], where a majority of the Court has applied a strict scrutiny standard *for reasons of content discrimination*, it has found *every one* to be unconstitutional." *Id.* Since McDonald's finding in 2006, the trend has continued. *See, e.g.*, United States v. Alvarez, 132 S. Ct. 2537 (2012) (holding unconstitutional federal law prohibiting persons from claiming falsely to have received military honors and quoting cardinal rule); Brown v. Entm't Merchs. Ass'n, 131 S. Ct. 2729 (2011) (holding unconstitutional state law prohibiting sale of violent video games to minors and quoting cardinal rule); United States v. Stevens, 559 U.S. 460 (2010) (striking federal law prohibiting sale of "crush videos" on overbreadth grounds and quoting cardinal rule). In one case, *Burson v. Freeman*, 504 U.S. 191, 193–211 (1992), a *plurality* of the Court upheld a law prohibiting campaigning within 100 feet of a voting site, finding that strict scrutiny was satisfied, with only three Justices willing to apply strict scrutiny to the regulation. *Holder v. Humanitarian Law Project*, 561 U.S. 1 (2010) may or may not be a counterexample. *See infra* notes 21–27 and accompanying text.

12. *Cardinal*, THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE 314 (2d ed. 1987) (meaning "of prime importance," from the Latin *cardo*, meaning "hinge, hence, something on which other things hinge").

scrutiny.[13]

Yet an absolute rule against all regulation targeting the broadest notion of content does not square with any sensible vision of free speech jurisprudence, with our national history, or with a nuanced concept of ordered liberty. All agree that some kinds of speech can and should be subject to regulation even though they are identified by their content. Obvious examples include prosecutions for perjury,[14] fraud,[15] and threats.[16] In tacit recognition, the Court has used various techniques, in a wide variety of cases, to avoid application of the cardinal rule rather than apply it straightforwardly to invalidate such laws. The cases are full of exceptions, mischaracterizations, and contortions, resulting in a well-documented array of ad hoc and inconsistent results.[17] But yet to emerge is a principled theory to rationalize the ad hoc judgments about when the draconian force of the cardinal rule is called for and when it is not.[18] The Court has sometimes responded by making the presumptuous proclamation that some speech is simply "not protected" by the command that there be no law abridging the freedom of speech, and it has embarked on the hubristic project of assessing relative values of different categories of speech.[19] At other times, it retreats to misty-eyed platitudes about the importance of speech to a democracy.[20]

13. There is inconsistency in the cases. *See infra* Part III.B.

14. Perjury undermines the function and province of the law and threatens the integrity of judgments that are the basis of the legal system. *See* United States v. Dunnigan, 507 U.S. 87, 97 (1993) ("To uphold the integrity of our trial system . . . the constitutionality of perjury statutes is unquestioned.").

15. *See, e.g.*, Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748, 771 (1976) (fraudulent speech generally falls outside the protections of the First Amendment).

16. *See, e.g.*, Virginia v. Black, 538 U.S. 343, 359–60 (2003).

17. *See, e.g.*, Clay Calvert, *Free Speech and Content-Neutrality: Inconsistent Applications of an Increasingly Malleable Doctrine*, 29 MCGEORGE L. REV. 69 (1997); Erwin Chemerinsky, *The First Amendment: When the Government Must Make Content-Based Choices*, 42 CLEV. ST. L. REV. 199 (1994); McDonald, *supra* note 10; Martin H. Redish, *The Content Distinction in First Amendment Analysis*, 34 STAN. L. REV. 113 (1981); Stephan, *supra* note 8, at 205; Cass R. Sunstein, *Half-Truths of the First Amendment*, 1993 U. CHI. LEGAL F. 25 (1993).

18. There may be an important distinction between what utterances are even covered by the First Amendment and how they are treated once found to be covered. The best case for this distinction is presented in Frederick Schauer, *The Boundaries of the First Amendment: A Preliminary Exploration of Constitutional Salience*, 117 HARV. L. REV. 1765 (2004). While there are many kinds of expression that never even receive direct First Amendment analysis, the examples I am interested in here are those that have been considered and have been explicitly found to be exceptions to First Amendment coverage.

19. *See* THOMAS I. EMERSON, THE SYSTEM OF FREEDOM OF EXPRESSION 326 (1970) (objecting to the value judgments necessitated by the doctrine).

20. *See* United States v. Alvarez, 132 S. Ct. 2537, 2547 (2012) ("Our constitutional tradition stands against the idea that we need Oceania's Ministry of Truth."); Snyder v. Phelps, 562 U.S. 443,

What is interesting about all of the facially incoherent results is their illumination of a consistent, deep-seated instinct of the Court that has been suppressed, denied, and disavowed at every turn. That pariah instinct whispers that sometimes government has good reason to regulate content to prevent harm; that principles of democracy, equality, and liberty can be promoted, rather than offended, by allowing it to do so. The feeble voice of this heretical intuition has been increasingly drowned out in the last half century by a crescendo of a free-speech drumbeat in our culture and our courts. The project of this Article is to bring that tenacious intuition out of the shadows and see what can be learned from it in unmasking the harm principle at work.

A close look reveals that the Supreme Court has responded to the power of the harm principle over the years. In the history of the Court's treatment of speech regulation since the early twentieth century, we can see seeds of that principle, although free-speech doctrine has tried hard to disguise it. When government seeks to prevent genuine social harm that is not associated with the spread of an idea, the Court has found ways to allow it. Examination of the strategies for achieving that result reveals a coherent and defensible, but never acknowledged, approach to free-speech protection that over the years has mostly endorsed the cardinal rule when ideas are at stake, but has departed from it when core First Amendment values are felt not to be threatened. The modern Court, however, has demonstrated an increasing unwillingness to engage in that type of discerning analysis. This Court has acted on what precedent has said, rather than on what it has done. This departure from the traditional, unspoken approach has led to mistakes in both directions, both over-protecting and under-protecting speech.

We see in the literature applause for a rule that, in its claim to absolutism, intentionally over-protects speech because that is the safer course for liberty.[21] This Article will challenge that rationale for the cardinal rule: over-protection does hurt liberty. It results in the disabling of government from redressing real social harm at or beyond the outer periphery of First Amendment concern, while at the same time posing a serious risk of under-protection at its core. The reasons for worrying that

452 (2011) ("[S]peech concerning public affairs is more than self-expression; it is the essence of self-government." (internal quotations omitted)); Citizens United v. FEC, 558 U.S. 310, 344 (2010) ("Under our Constitution it is We The People who are sovereign . . . . It is therefore important—vitally important—that all channels of communications be open to them during every election . . . ." (internal quotations omitted)).

21. *See* Stone, *supra* note 9, at 74.

the cardinal rule over-protects speech are clear: the absolute nature of the rule dismisses consideration of justifications. But under-protection is also a serious risk because the cardinal rule is so draconian that, when faced with a regulation with which the Court feels some instinctual sympathy, it has made moves worthy of Cirque du Soleil to avoid characterizing such regulations as content-based in the first place, knowing that if it does so the law is doomed. This allows the upholding of speech-restricting regulations without serious engagement with their true effect on democratic values, and camouflages the Court's tacit sympathies for state interests that have not been acknowledged or defended openly. This blank check for lack of candor is a foe to the freedom of speech.

Although few challenges to free political dissent have come to the Court since the paradigm-shifting terrorist attack on the World Trade Center in 2001, in the few cases presenting issues skirting close to that character, the Court—even with the cardinal rule in full swing—has not stepped up in defense of free speech. Indeed, the clamorous judicial protestations about the indispensable value of free speech, which we hear so strongly proclaimed in other cases, have been strangely frail or absent in the opinions when authoritarian interests of government may be at risk.[22]

At the same time, claims of free speech of other kinds—not involving challenge to government, but all kinds of other speech, like entertainment, aids to sexual gratification, hateful harangues against individuals or groups—those kinds of more peripheral speech have garnered a near-absolute degree of protection from regulation. In repeatedly striking down regulation of these kinds of expression, the Court has offered some of its most vehement protestations of the deified nature of the freedom of speech, whose essential role in the preservation of our democracy must never be impaired. This seems backwards.

A contrast between two recent cases in the Supreme Court illustrates this problem. First, consider the human rights advocate who wished to give counsel to the revolutionary Liberation Tamil Tigers as to how to avail themselves of lawful international remedies for their political grievances against Sri Lanka, instead of using violence.[23] The Court held that the advisor, offering political assistance in the form of pure speech, could

22. *See* Erwin Chemerinsky, *Not a Free Speech Court*, 53 ARIZ. L. REV. 723, 725 (2011) (Under the Roberts Court, "when the government is functioning as an authoritarian institution, freedom of speech always loses.").

23. Holder v. Humanitarian Law Project, 561 U.S. 1 (2010).

constitutionally be punished as a criminal for offering material support to a terrorist organization.[24] At the same time, another Court decision precluded any criminal punishment for a purveyor of commercial videos depicting the slow, brutal crushing of a kitten to death under the stiletto heel of an anonymous woman urging the animal to scream louder and more piteously for the sexual gratification of those who purchase such videos.[25] The "crush video" seller is protected from punishment by the First Amendment,[26] while the political advisor is not. The cardinal rule produced both results.

Although the first case, *Holder v. Humanitarian Law Project*, involved a law imposing a content-based restriction on political speech, the Court did not approach it as such, choosing instead to emphasize that the statute was carefully drawn to cover only a narrow category of speech, that which aids terrorist organizations either directly or indirectly. This evasion freed the Court to ignore the cardinal rule altogether and uphold the statute.[27] The second statute sought to reduce horrific brutality by drying up the market for the videographic fruits of that illegal conduct. Yet the Court in *United States v. Stevens* insisted that the cardinal rule would apply in full force to that conduct-based restriction.[28]

While democratic legitimacy and liberty depend on the robust protection of individual rights, including, centrally, the freedom of speech, it is important to remember also that governments are instituted for the purpose of protecting liberty and the pursuit of happiness for the people as well. Legislatures, although deserving of skepticism for their potential for

24. *Id.* at 16–18.

25. United States v. Stevens, 559 U.S. 460, 491 (2010) (Alito, J., dissenting) (describing type of crush videos federal statute was designed to prohibit). The statute was struck down on overbreadth grounds. *Id.* at 482 (majority opinion). The actual film at issue in the case depicted dogfighting. *Id.* at 466.

26. *Id.* at 482. The theory of harm involved in the case is discussed in greater detail below. *See infra* Part IV.C.

27. The Court equated the giving of advice to other nonspeech forms of support because they are "fungible." *Holder*, 561 U.S. at 36–38. It did find a government interest conspicuously characterized as "of the highest order," which could perhaps be understood as the application of strict scrutiny, but the analysis did not resemble the scrutiny usually afforded content-based restrictions. *Id.* at 28. In *United States v. Alvarez*, 132 S. Ct. 2537, 2549 (2012), for example, the Court found a "compelling" interest (a term not used in *Holder*), and still struck down the Stolen Valor Act because it was not "actually necessary" to achieve the government's interest. No such analysis appeared in *Holder*, 561 U.S. at 35–36 (affording Congress's judgment about the efficacy of the statute "significant weight").

28. *Stevens*, 559 U.S. at 468. Although the case was actually resolved on overbreadth grounds, the Court made clear that, as protected speech, regulation of the videos would be subjected to strict scrutiny. *Id. See also* Schauer, *supra* note 2, at 103–11 (analyzing the distinct kinds of harms present in three cases, including *Stevens*, and suggesting that a serious consideration of harm has been missing from First Amendment doctrine).

tyranny, were an expected source of justice for the nation.[29] In our own times, for example, while some rights are indeed protected directly by the Constitution through judicial enforcement, it is statutes that guarantee a wider range of liberties, such as the freedom from private racial, gender, or religious discrimination, as in the workplace or in housing. Likewise it is statutes, state and federal, that are responsible for our protection from a panoply of evils such as consumer fraud, unsafe food and drugs, abject poverty, unfair wages and hours, exploitation of children, dangerous products, crushing medical costs, political corruption, and literally countless other aspects of life that most people would deem to be key components to equality, personal liberty, and flourishing. Some of these laws operate on expression; yet they seek to prevent real harm to the common good.[30] Recall the Declaration of Independence:

> We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness.—*That to secure these rights, Governments are instituted* among Men, deriving their just powers from the consent of the governed.[31]

The iconic passage reminds us that the just *acts* of government have a role to play in liberty even as *restrictions* on government may. This strand of our constitutional intellectual history offers a cautionary tale, suggesting that we should be wary of a constitutional default rule that categorically disables government from legislating for the common good. It is important that limitations on government—as well as limitations on individual liberties—be justified. The cardinal rule has precluded any serious engagement with that project of justification by ruling all such legislation out of bounds.

This Article will argue for a more rigorous engagement with the harm principle. It is true that the special role of speech in our democracy requires more than just a balancing of quantitative variables like the incommensurables of rights versus state interests, as some have suggested.[32] Such a bald balancing, I agree with the absolutists, is doomed

29. *Cf.* David A. Strauss, *The Ubiquity of Prophylactic Rules*, 55 U. CHI. L. REV. 190, 201 (1988) ("[I]t is a commonplace that a legislative judgment that a statute is constitutional is generally entitled to some deference from a court.").

30. For example, there is a growing body of literature arguing that Title VII's prohibition of sexual harassment in the workplace—when the violation rests in part on expression—violates the cardinal rule against content regulation. *See infra* notes 238–34 and accompanying text.

31. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776) (emphasis added).

32. *See Alvarez*, 132 S. Ct. at 2551 (Breyer, J., concurring) ("Ultimately the Court has had to

to under-protect speech and overvalue the objectives that government might offer to support restrictions.[33] Instead, it is the *kind* of alleged harm, and especially the process by which that harm is produced, that provides a meaningful criterion for assessing the legitimacy of a regulation. This is not based on weighing rights against state interests and thus does not pose a threat that rights will be balanced away in the press of events.

The key distinction turns on scrutiny of the government's theory for how the restricted speech causes harm.[34] If the means for causing harm involve the dissemination of ideas, then the government is offering a censorial theory,[35] and the heaviest negative presumption is appropriate.[36] A censorial theory of harm exists, for example, if the government justifies a restriction on expression by arguing that people would be offended or induced to make bad choices by the message communicated. For those kinds of restrictions, I do not challenge free-speech orthodoxy, which would invalidate such laws.

A different kind of theory of harm, however, should give rise to a more deferential analysis. Sometimes the government articulates a non-

determine whether the statute works speech-related harm that is out of proportion to its justifications."); Ashcroft v. ACLU, 542 U.S. 656, 677–78 (2004) (Breyer, J., dissenting) (noting necessity of "examining both the extent to which the Act regulates protected expression and the nature of the burdens it imposes on that expression").

33. *See* Stone, *supra* note 7, at 225–27 (discussing persuasive arguments about the difficulty of making judgments about the two values in the balance, the burden on the individual and the interests of the government); Eugene Volokh, *Speech as Conduct: Generally Applicable Laws, Illegal Courses of Conduct, "Situation-Altering Utterance," and the Uncharted Zones*, 90 CORNELL L. REV. 1277, 1307–10 (2005) (same).

34. *See* Frederick Schauer, *On the Distinction Between Speech and Action*, 65 EMORY L.J. 427, 453–54 (2015) (arguing that autonomy cannot support a distinction between speech that causes harm and autonomous action that causes harm).

35. I have chosen the perhaps inelegant word "censorial" to capture a theory of harm that involves censorship of an idea or message, derived from "censor," originally the person charged with counting the people (census) and supervising their morals. *Censor*, THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE 334 (2d ed. 1987). I rejected "censorious" because it connotes severe criticism or disapproval, more like "censure," which is not the meaning I want to convey here. *See Censorious*, THE RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE 334–35 (2d ed. 1987).

36. Some say this power simply was not part of the police power to begin with, and so such a restriction would be ultra vires even without a First Amendment. *See* Howard Gillman, *Preferred Freedoms: The Progressive Expansion of State Power and the Rise of Modern Civil Liberties Jurisprudence*, 47 POL. RES. Q. 623, 639 (1994) (discussing constitutional tradition of addressing free speech rights as a limit on legislative power rather than as a set of immunities). Others suggest that the First Amendment did away with any kind of sedition-like offenses, and use the extrajudicial fate of the Alien and Sedition Acts in the late eighteenth century as support. *See* Abrams v. United States, 250 U.S. 616, 630–31 (1919) (Holmes, J., dissenting). *See also* ZECHARIAH CHAFEE, FREE SPEECH IN THE UNITED STATES 221–22 (1941); William T. Mayton, *Seditious Libel and the Lost Guarantee of a Freedom of Expression*, 84 COLUM. L. REV. 91, 94–95 (1984).

censorial reason for restricting speech. The argument in such a case is not that the audience must be protected from a message or idea, but rather that something in the nature of the expression works harm in a way separate from any message or idea propounded. For example, a restriction on making outrageous personal attacks at a funeral might be based on a government interest in protecting the privacy of vulnerable and grieving family members, rather than on suppressing any message, although by narrowing the restriction only to an expression that a jury could find "outrageous," such a rule would qualify as content based, and be subject to the cardinal rule under current law.[37] In such a case, the nearly insuperable presumption of invalidity should not apply and the government should have a meaningful opportunity to demonstrate to the Court that it is reasonably protecting against social harms validly within the police power, and that operating upon content of expression may be the best way to achieve its legitimate goals. Laws shown to act non-censorially to address a real harm should receive the lesser scrutiny afforded to content-neutral laws. This would be consistent with the understanding of the freedom of speech as derivative of the freedom of thought; by not implicating freedom of thought, a non-censorial curtailment of speech would not threaten the abridgment that the Constitution prohibits.

When speech threatens harm because it is too persuasive, too provocative, or too offensive, the foundational principle of Millian freedom of thought is implicated. David Strauss has identified a centrally important "persuasion principle" in constitutional law, which protects speech that may cause harm by being too effective in influencing the rational process of its audience.[38] Indeed, we value speech precisely because "it persuades people to do or believe things."[39] But when speech causes harm by means not involving the rational process of its audience, that central principle of free speech is not implicated; the extraordinary protective measures the law employs to protect expression because of its uniquely indispensable relationship to democracy and autonomy may not be justified in limiting the police powers.

There is some overlap between what I am calling these non-censorial

37. The example is merely illustrative, although it shares some similarities with *Snyder v. Phelps*, 562 U.S. 443, 451, 458–59 (2011) (invalidating a jury award for intentional infliction of emotional distress based on "outrageous conduct" involved in picketing near a funeral).

38. David A. Strauss, *Persuasion, Autonomy, and Freedom of Expression*, 91 COLUM. L. REV. 334, 337 (1991).

39. *Id.*

restrictions and some types of speech that the Court has called "unprotected." For example, so-called "fighting words," which the Court has ruled to be completely outside the protection of the First Amendment, could qualify under a non-censorial theory if the theory by which the words cause harm is that they operate directly to produce a response unmediated by the processing of an idea or message.[40] But my emphasis is on the theory of harm, not on the nature or value of the words spoken. This approach, therefore, avoids the presumptuous judicial exercise of assessing the value of speech.

The scope of my disagreement with the Court, the cardinal rule, and their supporters is not vast. There is no dispute that "the right to free speech is at its core the right to communicate—to persuade and to inform people through the content of one's message."[41] Of course, the "central meaning" of the First Amendment is freedom from official sanction for criticism of the government.[42] The First Amendment is fundamentally concerned with "distortion of public debate, improper motivation, and communicative impact."[43] And I acknowledge that the roots of the rule against content regulation lie in a basic interpretation of the First Amendment itself, a characteristic that could, in theory, justify a judge-made rule.[44] But my claim is that early interpretive judgment about content regulation never focused on the kinds of expansive implications that the cardinal rule has taken on in recent decades. Scholars, as well as justices, have been too quick to embrace an absolutism that is justified neither by precedent nor by democratic theory. Where I take issue with the dominant thread of jurisprudence and scholarship is in its claim that drawing a bright line between all content-based and all content-neutral restrictions is the right way to protect the values enshrined in the First Amendment. Indeed, I argue, in its unjustified abandonment of the harm principle it is counterproductive of that goal, as exemplified in several recent cases from

40. Different harms, even if non-censorial, may still be very different from one another in the way that they cause harm. *See* Frederick Schauer, *The Phenomenology of Speech and Harm*, 103 ETHICS 635, 640–53 (1993) (addressing to what degree harms caused by speech are lesser than other harms). *See also* Erica Goldberg, *Free Speech Consequentialism*, 116 COLUM. L. REV. 687 (2016) (arguing that some speech harms are equivalent to conduct harms).

41. Volokh, *supra* note 33, at 1304.

42. N.Y. Times Co. v. Sullivan, 376 U.S. 254, 273 (1964).

43. Stone, *supra* note 9, at 56–57 (defining communicative impact as fear of how people will react to what the speaker is saying). *See also* ALEXANDER MEIKLEJOHN, POLITICAL FREEDOM: THE CONSTITUTIONAL POWERS OF THE PEOPLE 27 (1960); *infra* notes 176–71 and accompanying text (discussing communicative impact).

44. *See* Strauss, *supra* note 29, at 199–200.

the Roberts Court.[45] The difference, while not vast, can be dispositive in assessing the constitutionality of efforts to address critical and evolving social issues such as sexual harassment, hate speech, violent pornography, child sex abuse, and even campaign finance restrictions.[46] It reflects a very different orientation with regard to the relationship between the Court and the legislature in addressing newly recognized social harm.

During the *Lochner* era, Progressives deplored the Court's straitjacketing of government, as the Court diminished societal liberty by disabling government from addressing the pressing social issues of the day under an expansive view of a constitutional right to contract.[47] While it has become quite popular, in and out of academia, to draw a parallel between the Roberts Court's enforcement of first amendment rights and *Lochner*,[48]

45. The approach suggested in this Article calls into question the analysis used in several recent cases, which are discussed throughout. *See* United States v. Alvarez, 132 S. Ct. 2537, 2551 (2012) (striking down Stolen Valor Act); Brown v. Entm't Merchs. Ass'n, 131 S. Ct. 2729, 2742 (2011) (striking down state law restricting sale of violent video games to minors); Holder v. Humanitarian Law Project, 561 U.S. 1, 40 (2010) (validating constitutionality of federal "material support" statute as applied to speech); United States v. Stevens, 559 U.S. 460, 481–82 (2010) (striking down federal law restricting sale of videos showing extreme animal cruelty).

46. Given the broad definition of content that has been accepted in the recent cases, the regulation of political speech or electioneering at issue in *Citizens United v. FEC*, 558 U.S. 310 (2010), qualifies as a content-based law. However, neither the *Citizens United* Court nor the earlier Court in the case that it overruled, *Austin v. Michigan Chamber of Commerce*, 494 U.S. 652 (1990), *overruled by Citizens United*, 558 U.S. 310 (2010), characterized it that way. This illuminates that a very important difference leading to the overruling of *Austin* had to do with whether the Court viewed the law as censorial or not. That is, the later Court seemed to suggest that the campaign finance laws acted to suppress ideas and even viewpoints. It went so far as to describe the differential treatment of corporate expenditures as "based on a speaker's corporate identity." *Citizens United*, 558 U.S. at 347. Speech restrictions based on identity, the Court warned, "are all too often simply a means to control content." *Id.* at 340. The *Austin* Court, in contrast, saw the restrictions as a "burden [on] expressive activity," but one that "also allow[ed] corporations to express their political views." *Austin*, 494 U.S. at 658, 660. This unacknowledged difference in whether the restriction was censorial or non-censorial in its theory of harm was highly important, if not critical, to the opposite outcomes that the two decisions reached. *See* Strauss, *supra* note 38, at 340 (showing that "expenditure of money . . . may be [regulated] to avoid corruption . . . [b]ut . . . not . . . to reduce the effectiveness of the speech").

47. The term "*Lochner* era" refers to a period around the turn of the twentieth century, in which the Court repeatedly invalidated state labor and market regulations on constitutional grounds. *See* HOWARD GILLMAN, THE CONSTITUTION BESIEGED: THE RISE AND DEMISE OF LOCHNER ERA POLICE POWERS JURISPRUDENCE 1–5 (1993). *See also id.* at 148–50 (describing Progressive attack on formalism of the judiciary).

48. *See, e.g.*, JAMIE RASKIN, THE SUPREME COURT IN THE CITIZENS UNITED ERA: A CENTURY AFTER THE *LOCHNER* ERA, THE ROBERTS COURT IMPOSES A STARTLING NEW CORPORATISM ON AMERICA (2014), http://www.pfaw.org/media-center/publications/supreme-court-citizens-united-era-century-after-lochner-era-roberts-court-; Thomas H. Jackson & John Calvin Jeffries, Jr., *Commercial Speech: Economic Due Process and the First Amendment*, 65 VA. L. REV. 1, 30–32 (1979) (arguing that commercial speech doctrine is reminiscent of *Lochner*); Linda Greenhouse, *Linda Greenhouse on*

my claim is a narrower and more pointed one: that the specific reason *Lochner* was a mistake is the same reason that the cardinal rule is a mistake. That mistake is in identifying a too-narrow conception of the legitimate sphere of police-power regulation, the common good. I have argued that the *Lochner* period in our history, a century ago, reflected the Court's impoverished understanding of what it meant for a state to act in the common good, leading to a desiccated view of the police power, the classic power of a state to prevent harm to its people.[49] That deficient understanding of harm in a world radically changed by the Industrial Revolution led the Court to rule off limits, in the name of liberty, to many legislative efforts to remedy serious social problems.[50] The New Deal saw a new conception of societal harm, which included a new appreciation of government's obligation to protect persons from exploitation at the hands of other private individuals, for the sake of the common good.[51] Similarly, today's cardinal rule has led the modern Supreme Court to disable the government from addressing many pressing social issues of our day, again in the name of liberty. It has done so because it undervalues the legislature's role in recognizing new kinds of societal harm and redressing them. Like the *Lochner* era before it, this laissez-faire era in Supreme Court free-speech analysis misunderstands the role of the harm principle and, at the same time, disserves the precious liberty right that it seeks to protect.

The stakes of getting this right are high. Geoffrey Stone demonstrates in his important book that, in each of six episodes during our history in which the nation was in great fear, the government overreacted by punishing political dissent.[52] Thus, on the eve of war with France in 1798, during the Civil War, during World War I, World War II, the Cold War, and the Vietnam War, Stone shows that the extraordinary pressures to suppress dissent led to a sacrifice of civil liberties, particularly the freedom of speech. The Supreme Court was part of that regrettable history, either actively or by looking the other way. The tone has changed in the past half

the Roberts Project, HARV. C.R.–C.L. L. REV. (Oct. 10, 2013), http://harvardcrcl.org/linda-greenhouse-on-the-roberts-project/.

49. *See* Rebecca L. Brown, *The Art of Reading* Lochner, 1 N.Y.U. J.L. & LIBERTY 570, 584 (2005); Rebecca L. Brown, *The Fragmented Liberty Clause*, 41 WM. & MARY L. REV. 65, 82–83 (1999).

50. *See, e.g.*, Adkins v. Children's Hosp., 261 U.S. 525, 561–62 (1923) (striking down minimum-wage law for women); Lochner v. New York, 198 U.S. 45, 64 (1905) (striking down maximum-hour law for bakers).

51. *See, e.g.*, W. Coast Hotel Co. v. Parrish, 300 U.S. 379, 398–400 (1937) (overruling *Adkins* to uphold working-hour regulations for women as benefiting the common good).

52. GEOFFREY R. STONE, PERILOUS TIMES: FREE SPEECH IN WARTIME FROM THE SEDITION ACT OF 1798 TO THE WAR ON TERRORISM 12–14 (2004).

century, but it is important to recognize that the second half of the twentieth century was a time largely unchallenged by serious society-wide fear, and to the extent the last decade has brought an increased fear of terrorism, it has so far remained quite remote. We do not know for sure, accordingly, how committed we as a people are to the protection of political dissent when we are deeply afraid. A rule that too bluntly dictates a wooden solution to nuanced issues is inadequate to the task of protecting our deepest values in the face of new and destabilizing fears. Moreover, by ignoring the importance of the harm principle to ordered liberty, the cardinal rule undermines the quest for a well-ordered society, regulated by laws that genuinely seek to promote important constitutional values like equality and liberty.

My argument will proceed as follows. First, I will offer a brief overview of the cardinal rule and how current law applies it, followed, in Part II, by a discussion of the theoretical and jurisprudential problems that it creates. Part III will look back at the intellectual history leading to this doctrine and offer a revisionist account suggesting that, all along, there was an understanding that speech produces harm in different ways, with different implications for freedom of speech. A close examination of the Court's early identification of "unprotected" categories of speech will reveal that the important distinction has not depended on whether a regulation involves the content of speech, but rather on whether the harm is caused by the promulgation of a thought or idea. Part IV will show that, without examination, this important understanding was lost, starting in 1972, and the effects of that error have ballooned since. Finally, in Part V, I will show how the harm principle, properly understood, can provide the foundation for a better approach that has the value of consistency with both the major philosophical underpinnings of the First Amendment and the constitutional traditions underlying fundamental rights more generally.

My aim is not to provide a comprehensive analysis of specific cases, which others have accomplished artfully.[53] Rather, I suggest a deep structure running through the apparently aberrant holdings, by demonstrating that, like a black hole, an unseen harm principle can explain the phenomena. This Article seeks to bring it out of the darkness.

53. *See generally* McDonald, *supra* note 10; Stephan, *supra* note 8; Stone, *supra* note 7.

I. A SHORT SUMMARY OF CURRENT LAW

The most salient pivot in free-speech jurisprudence today is the distinction between laws that are based on the content of speech and those that are neutral with respect to content.[54] While content-neutral laws are generally tolerated, current law proclaims that government may not regulate speech based on its content.[55] This grand division, recognized explicitly since 1972,[56] aims to isolate the most worrisome kinds of regulation, those in which government singles out particular messages for approval or disapproval.[57] Under the aegis of the cardinal rule, a regulation targeting content can survive only if one of two things is true. In theory, the law could survive if the government could supply a sufficiently compelling interest. In practice, however, the Court has been leery of finding any government interest sufficient to overcome the presumption against content regulation, and so in practice the rule has been virtually absolute for laws recognized as content based.[58] The other way for government to vindicate a content-based restriction, under current law, is to show that the speech in question falls into a category that was recognized in 1942 as "unprotected," and therefore may be regulated freely without First Amendment constraint.[59]

The Court has declared a short list of certain categories of speech to be completely unprotected.[60] If the narrow definitional conditions for the category are fulfilled, then the government can regulate or punish these expressions.[61] The unprotected classes of speech are often identified as

54. *See* Regan v. Time, Inc., 468 U.S. 641, 648 (1984). *See also* Republican Party of Minn. v. White, 536 U.S. 765, 774 (2002); McIntyre v. Ohio Elections Comm'n, 514 U.S. 334, 345–46 (1995); Ark. Writers' Project, Inc. v. Ragland, 481 U.S. 221, 230 (1987); Carey v. Brown, 447 U.S. 455, 464 (1980). Generally, a law is content neutral if it applies to all speech regardless of the message, while a content-based law turns on what was sa*id*. *See* ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 965 (4th ed. 2011).

55. CHEMERINSKY, *supra* note 54, at 960. *Accord* R.A.V. v. City of St. Paul, 505 U.S. 377, 382 (1992); DANIEL A. FARBER, THE FIRST AMENDMENT 20 (3d ed. 2010).

56. *See infra* Part IV (discussing Police Dep't of Chi. v. Mosley, 408 U.S. 92 (1972)).

57. *See* Stephan, *supra* note 8, at 217–18 (describing concerns of viewpoint discrimination).

58. *See* sources cited *supra* note 10. Perhaps *Holder v. Humanitarian Law Project*, 561 U.S. 1 (2010) could be considered an exception. There, the Court upheld a content-based law without mentioning the cardinal rule or explicitly applying strict scrutiny. *See* sources cited *supra* note 27.

59. *See* CHEMERINSKY, *supra* note 54, at 1017.

60. *See* United States v. Stevens, 559 U.S. 460, 468–69 (2010) (quoting Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942)). *But see R.A.V.*, 505 U.S. at 383–84 (discussing that even in unprotected speech categories, regulations cannot engage in content discrimination by prohibiting only speech that embraces a certain viewpoint).

61. There are also some classes of speech that are not viewed as unprotected, but that may be regulated upon a lesser showing by the government. *See, e.g.*, Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 566 (1980) (establishing four-part test for commercial speech regulation

incitement to illegal activity, fighting words, obscenity, defamation, fraud, and speech integral to criminal conduct, although the exact contours of this list vary among incantations and have changed over time.[62] The commentators have described this as a "two-level" theory of speech under the First Amendment: protected speech and unprotected speech.[63]

Many leading free-speech scholars have defended this structure, sometimes passionately[64] and sometimes reluctantly.[65] The principal advantage of setting out, in advance, specific categories that are subject to regulation, while prohibiting all other efforts to restrict on the basis of content, is said to be that it protects against individualized consideration of particular cases, which can lead to the inflation of governmental interests, and a consequent slippery slope.[66] Cass Sunstein has called the two-level approach essential to "any well-functioning system of free expression" because it prevents the lowering of the burden of justification placed on government when it regulates high-value speech, and simultaneously allows the regulation of "speech that in all probability should be regulated."[67] "Categorical rules," supporters urge, "are more likely to work

resembling intermediate scrutiny); Young v. Am. Mini Theatres, Inc., 427 U.S. 50, 70–71 (1976) (stating that state may legitimately use content of nonobscene erotic films as a basis for special regulation).

62. *Stevens*, 559 U.S. at 468–69. Profanity was once included on the list of unprotected categories, but it quietly disappeared from most renditions of the list. *See Chaplinsky*, 315 U.S. at 572. Perhaps this was affected by the 1971 decision in *Cohen v. California*, 403 U.S. 15, 16, 26 (1971), involving a jacket bearing the words "fuck the draft," but the issue there was not presented as whether profanity is protected or not; rather, the Court saw a censorial theory of harm and found, accordingly, that the First Amendment barred the prosecution. A certain kind of child pornography was added to the list in *New York v. Ferber*, 458 U.S. 747, 773–74 (1982).

63. The term "two-level" is attributed to Harry Kalven Jr. *See* Harry Kalven Jr., *The Metaphysics of the Law of Obscenity*, 1960 SUP. CT. REV. 1, 10. *See also* GEOFFREY R. STONE ET AL., CONSTITUTIONAL LAW 1134 (7th ed. 2013) (discussing two-level theory). Geoffrey R. Stone includes within the category of "low value" speech both the categories that are unprotected and those that receive less-than-full protection. Stone, *supra* note 9, at 47 & n.4.

64. *See* Geoffrey R. Stone, *Restrictions of Speech Because of Its Content: The Peculiar Case of Subject-Matter Restrictions*, 46 U. CHI. L. REV. 81, 82–84 (1978).

65. *See* T.M. SCANLON, *Freedom of Expression and Categories of Expression*, *in* THE DIFFICULTY OF TOLERANCE 85 (2003) [hereinafter *Freedom of Expression*, *in* THE DIFFICULTY OF TOLERANCE] ("I find it hard to resist the idea that different categories of expression should to some degree be treated differently in a theory of freedom of expression."). Nonetheless, Scanlon was wary: "categories . . . rest on distinctions . . . that a partisan of freedom of expression will instinctively view with suspicion." *Id.* at 103. Overall, Scanlon seems to be a supporter. *See* T.M. Scanlon, *Freedom of Expression and Categories of Expression*, 40 U. PITT. L. REV. 519, 519–20 (1979) [hereinafter *Freedom of Expression*].

66. *See* Stone, *supra* note 9, at 73–74; Strauss, *supra* note 29, at 200–01.

67. CASS R. SUNSTEIN, THE PARTIAL CONSTITUTION 233–34 (1993). *See* Frederick Schauer, *Categories and the First Amendment: A Play in Three Acts*, 34 VAND. L. REV. 265, 286 (1981)

in spite of the defects in the human machinery on which we must rely to preserve fundamental liberties."[68] Some acknowledge that it also over-protects, but accept that as a cost of giving due regard to free speech.[69]

The strongest defense of the cardinal rule has been in the claim that it is needed to prevent the Court from having too much discretion, to tie its hands so that it does not succumb to any forms of viewpoint discrimination or value judgment of its own about the content of speech. The near refusal on the part of the Court even to entertain any justifications for content regulation has been hailed as a "fortress model" of speech protection, intentionally over-protecting "in order to minimize the potential harm from legislative and administrative abuse and judicial miscalculation."[70]

That defense, however, has had the legs cut out from under it. A recent empirical study of the Supreme Court's behavior in protecting free speech offers this disquieting conclusion: "[J]ustices are opportunistic free speechers . . . willing to turn back regulation of expression when the expression conforms to their values and uphold it when the expression and their preferences collide."[71] This study covers only the cases from 1953 to 2010,[72] overlapping substantially with the very period of time when we have been in the grip of the cardinal rule and have considered protection of free speech to be, more or less, absolute. This finding casts doubt on any confidence one might have that a bright-line rule can avoid bias.[73] Indeed, the cardinal rule actually may invite bias by excluding transparent, data-based assessment of the actual burdens and benefits of any particular regulation, let alone a frank excavation of democratic principles in tension. The cardinal rule insidiously undervalues candor and, with it, liberty.

The system has had a few detractors. Some have attacked the two-level structure, primarily on the ground that it has produced incoherent results.[74] Some worry that it allows too much speech to be regulated by

(branding entire categories as unprotected allows greater protection for the protected categories).

68. LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW 584 (1978).

69. *See* City of Ladue v. Gilleo, 512 U.S. 43, 60 (1994) (O'Connor, J., concurring) (acknowledging that reasonable regulations can fall to the cardinal rule).

70. Stone, *supra* note 9, at 74 & n.122 (citing LEE C. BOLLINGER, THE TOLERANT SOCIETY: FREEDOM OF SPEECH AND EXTREMIST SPEECH IN AMERICA 76 (1986)).

71. Lee Epstein et al., *Do Justices Defend the Speech They Hate? In-Group Bias, Opportunism, and the First Amendment*, AM. POL. SCI. ASS'N 2013 ANN. MEETING, Aug. 2013.

72. *See id.* at 2.

73. *See* Stone, *supra* note 9, at 73 (arguing that balancing presents danger "that judges and jurors may be influenced by their own conscious or unconscious biases").

74. *See* McDonald, *supra* note 10, at 1353; Redish, *supra* note 17, at 113; Stephan, *supra* note 8, at 206–07.

branding entire categories as unprotected.[75] Others have argued that this segregation violates the core principle of equality in the marketplace of ideas.[76] Still another critic warns that the categorical two-level system "marks for extinction even speech that is the core value of the First Amendment and a quintessential value of representative democracy."[77] Indeed, the lead case recognizing "fighting words" as unprotected involved verbal outbursts that could be considered political criticism at the heart of free-speech concern.[78]

The problem of under-protection has been mitigated substantially by the shrinking of the categories over time, and the Court's reluctance to expand the list to include any new categories.[79] Since the articulation of the "fighting words" doctrine in 1942, for example, no conviction has been upheld on this ground. But the criticisms tend not to go to the heart of the cardinal rule, focusing rather on its application.

II. THE PROBLEM

More important than the inconsistency that the cardinal rule has produced is a theoretical problem, that the process of reasoning through the constitutional question has been forced into a construct that is backwards. The common-law constitutional system generally used to recognize and develop the scope of rights begins with the identification of a principle in the Constitution,[80] with refinement of the rights coming in subsequent

75. *See* Stephan, *supra* note 8, at 206 (condemning the broad neutrality rule because it "requires that all speech receive the same treatment, regardless of the values implicated").

76. *See* Kenneth L. Karst, *Equality as a Central Principle in the First Amendment*, 43 U. CHI. L. REV. 20, 31 (1975).

77. Burton Caine, *The Trouble with "Fighting Words":* Chaplinsky v. New Hampshire *Is a Threat to First Amendment Values and Should Be Overruled*, 88 MARQ. L. REV 441, 456 (2004).

78. *Id.* at 456–57 & n.74 (discussing *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942) and its facts involving insults to police). Justice Kennedy, too, has voiced a criticism, in a slightly different context, of "a jurisprudence of categories rather than ideas" that mechanically grants government authority to restrict speech. Int'l Soc'y for Krishna Consciousness, Inc. v. Lee, 505 U.S. 672, 693–94, (1992) (Kennedy, J., concurring in part and dissenting in part) (addressing public forum doctrine).

79. *See* Stephan, *supra* note 8, at 213 ("For the most part, the concept of less-protected speech has been used to expand first amendment protection. . . ."); Geoffrey R. Stone, *Free Speech in the Twenty-First Century: Ten Lessons from the Twentieth Century*, 36 PEPP. L. REV. 273, 284 (2009) (explaining two factors tending to cabin the impact of low-value doctrine). One new category of unprotected speech was recognized in 1982. *See infra* Part IV.B (discussing New York v. Ferber, 458 U.S. 747 (1982)).

80. *See, e.g.*, Griswold v. Connecticut, 381 U.S. 479, 485–86 (1965) (identifying a right to privacy based on constitutional text, precedent, traditions, and practice, and applying it to contraceptive use by married couples).

cases.[81] This is what we usually think of as constitutional reasoning, or constitutional common-law decisionmaking.[82] There is a strong case to be made that the legitimacy of judicial review depends on it.[83]

In the application of the cardinal rule, however, the Court appears to have done just the opposite. The Court started in 1942 with a list of specific examples of speech that had historically been allowed to be regulated under state police power, notwithstanding the First Amendment.[84] It mused briefly on what might explain the examples of speech subject to regulation, as I will explain in further depth below.[85] But instead of using these cases to derive a general principle, and then in turn to apply that principle to later cases, the Court never moved past the specific examples, never engaged in the common-law endeavor of drawing out a thread of principle to extend to future cases. Over time, the list changed both in its content and in the characterization of what the list represented—but never with acknowledgment of the changes or discussion of reasons for them. The content of the list evolved, and the character of the list, which began as a set of examples of speech regulations that had historically been considered not to implicate the First Amendment, came to be referred to as a finite set of "exceptions" to the First Amendment.[86]

As the Court in time became more inclined to protect speech, it narrowed several of the categories of speech on the list, refining some specific "exceptions" to ensure appropriate application.[87] But still an animating principle eluded the discussions. The list slowly, and again without discussion, took on the character of exclusivity, such that if not on the 1942 list, then a kind of speech was not subject to any kind of content-

81. *See, e.g.*, Washington v. Glucksberg, 521 U.S. 702, 727–29 (1997) (rejecting claim that privacy protects assisted suicide); Roe v. Wade, 410 U.S. 113, 154 (1973) (applying privacy to abortion).

82. *See* RONALD DWORKIN, LAW'S EMPIRE 219–22 (1986) (discussing role of principle in integrity); Rebecca L. Brown, *Assisted Living for the Constitution*, 59 DRAKE L. REV. 985, 991 (2011) (arguing that principle is important to constitutional interpretation in a way that is different from its role in common-law reasoning). *See generally* DAVID A. STRAUSS, THE LIVING CONSTITUTION (2010) (arguing that constitutional law gains its legitimacy from the common-law reasoning that undergirds it).

83. *See* STRAUSS, *supra* note 82, at 47.

84. *See* Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942).

85. *See infra* Part III.A.

86. Genevieve Lakier demonstrates that this history was incorrect, and that the assessment of some speech as having low value came much later. *See* Genevieve Lakier, *The Invention of Low-Value Speech*, 128 HARV. L. REV. 2166 (2015).

87. *See* Watts v. United States, 394 U.S. 705, 708 (1969) (threats); Brandenburg v. Ohio, 395 U.S. 444, 447–49 (1969) (incitement); N.Y. Times Co. v. Sullivan, 376 U.S. 254, 292 (1964) (libel); Roth v. United States, 354 U.S. 476, 488–94 (1957) (obscenity).

based regulation.[88] Without any theory to justify or circumscribe the list, the Supreme Court in 2010 felt free to proclaim that the list was forever etched in stone, never to be expanded.[89] Indeed, not only did the Court abandon any effort at unifying the "exceptions" with a justifying principle, but it did so with an affect worthy of the classic film *Casablanca*'s Captain Renault, who was "shocked, shocked" to discover gambling at Rick's Cafe. The Supreme Court found it "startling and dangerous" that the government would urge the Court to recognize a new category of speech as unprotected, since of course the Court did not engage in such an endeavor (except when it did).[90] The Solicitor General had asked the Court to treat the video mentioned above—depicting horrific animal cruelty, created and sold for the purpose of sexual gratification of purchasers—similarly to obscenity, an item on the list. A stunning eight justices of the Court, achieving a rare degree of unanimity on a rights issue, agreed that the government's resort to the common-law process of analogy was "startling and dangerous."[91] This leaves the rule against content regulation utterly unresponsive to changing values and societal conceptions of harm[92]—the same fate that I claim doomed *Lochner*.

The flawed process of constitutional reasoning that has afflicted the cardinal rule and its accompanying list of unprotected categories of speech is not the only jurisprudential collateral damage of the cardinal rule. The converse problem is that during the same period there were numerous situations in which the cardinal rule would have required invalidation of a regulation, and the speech in question did not fall into one of the limited categories of exceptions—but in which the Court was inclined to uphold the law, unlike the case of the animal cruelty video. There are frequent examples of cases in which the Court found ways to uphold without ever choosing one of the only two options ostensibly available under the cardinal rule: strike the law or find the speech unprotected. Instead, the

88. The Court arguably added one type of speech to the list, when it upheld the power of New York to prohibit the sale of films showing the sexual abuse of children. New York v. Ferber, 458 U.S. 747, 775 (1982). It has been unclear whether the Court considered this an actual addition to the list of unprotected categories, or, as it later indicated in *United States v. Stevens*, 559 U.S. 460, 471 (2010), rather just an application of a previously recognized category of conduct integrally related to criminal conduct.

89. *See Stevens*, 559 U.S. at 470–72.

90. *Id.* at 470.

91. *Id.*

92. *See* SCANLON, *Freedom of Expression*, *in* THE DIFFICULTY OF TOLERANCE, *supra* note 65, at 87 (discussing that as peoples' values change, categories of interests designed to rank expression must change).

Court used an array of diversionary techniques that resulted in regulations being upheld, but without an honest assessment of the First Amendment principles at stake. In such cases, it has implausibly found laws to be content neutral,[93] it has sometimes assessed a law without analyzing whether it is content neutral or not,[94] and it has sometimes minimized the importance of content neutrality by focusing on the absence of viewpoint discrimination.[95]

I will argue below that, although the Court never addressed what it was actually doing when it upheld these content-based laws without invoking the cardinal rule, a close reading of the decisions reveals that the Justices found a whole variety of ways to satisfy themselves that these laws did not encroach on values protected by the First Amendment. These examples show that when a regulation seeks to address harm that is not produced by the dissemination of an idea, but rather the government offers a persuasive non-censorial theory of how the speech causes harm, then the Court has tended to be sympathetic and has found a way to make an end run around the cardinal rule. The lack of acknowledgment of this important role of the harm principle in free-speech jurisprudence, however, leaves the matter easily open to inappropriate considerations, in-group biases, and error.

III. THE STRANGE PROVENANCE OF THE CARDINAL RULE

The literature overwhelmingly[96] lays the cardinal rule at the feet of one Supreme Court decision issued in 1972, when Justice Marshall, writing for a seven-member majority of the Court in *Police Department of Chicago v. Mosley*,[97] included the following stentorian sentence in his analysis of the case: "[A]bove all else, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its

93. *See* Hill v. Colorado, 530 U.S. 703, 725 (2000) (finding restrictions on protests outside health facilities content neutral); City of Renton v. Playtime Theatres, Inc., 475 U.S. 41, 48 (1986) (finding zoning restrictions on adult theaters content neutral).

94. *See* Austin v. Mich. Chamber of Commerce, 494 U.S. 652 (1990), *overruled by* Citizens United v. FEC, 558 U.S. 310, 363 (2010); Rowan v. U.S. Post Office Dep't, 397 U.S. 728 (1970).

95. CBS v. Democratic Nat'l Comm., 412 U.S. 94, 130 (1973). Viewpoint discrimination is a subset of content-based regulation, and the most worrisome, but the fact that a law is not discriminatory against viewpoint does not mean that it is content neutral.

96. *See, e.g.*, Chemerinsky, *supra* note 10, at 50; Steven J. Heyman, *Spheres of Autonomy: Reforming the Content Neutrality Doctrine in First Amendment Jurisprudence*, 10 WM. & MARY BILL RTS. J. 647, 650 (2002); Stephan, *supra* note 8, at 203; Stone, *supra* note 7, at 196.

97. Police Dep't of Chi. v. Mosley, 408 U.S. 92 (1972). The majority opinion was joined by Chief Justice Burger and Justices Douglas, Brennan, Stewart, White, and Powell. Two Justices (Rehnquist and Blackmun) concurred in the result without opinion.

subject matter, or its content."[98]

The *Mosley* statement did at least two significant things. First, it depicted the prohibition on content-based rules as absolute. The phrasing leaves no room for judgment about the legitimacy of government reasons or the nature of the contribution of the speech being restricted: "no power," it declares. Although technically enforced by the application of strict scrutiny,[99] the cardinal rule in practice has meant that once the Court characterizes a statute as content based, the statute falls.[100]

Even more consequentially, the other significant feature of the quoted language from *Mosley* is its expansion of the meaning of content-based regulation. Power to restrict is denied with respect to the expression's "message, its ideas, its subject matter, or its content."[101] The inclusion of "subject matter" and "content" broadened dramatically the reach of the First Amendment's barrier to regulation.[102] Together, these two facets of the *Mosley*-turned-cardinal-rule have contributed to the degradation of the constitutional analysis of free speech. In order to see how, it is necessary to take a whirlwind tour through the pre-*Mosley* approach to free-speech protection.

A. A REVISIONIST ACCOUNT OF THE REGULATION OF SPEECH BEFORE *MOSLEY*

In the fifty years prior to *Mosley*, the key distinction in First Amendment jurisprudence is better understood, not as a line between regulations that targeted content and those that did not—and within the former group a dyad of high-value and low-value speech—but rather as a line between regulations that suppressed ideas or messages and those that

98. *Id.* at 95.

99. For all content-based regulations of protected expression, Justice Anthony Kennedy has maintained that strict scrutiny is inappropriate; in his view, such regulations should be per se unconstitutional. *See* Republican Party of Minn. v. White, 536 U.S. 765, 792–93 (2002) (Kennedy, J., concurring); Simon & Schuster, Inc. v. Members of N.Y. Crime Victims Bd., 502 U.S. 105, 124 (1991) (Kennedy, J., concurring). The Court has nominally adhered to the strict-scrutiny standard, but has found it virtually impossible to meet.

100. *See* sources cited *supra* note 10.

101. *Mosley*, 408 U.S. at 95.

102. *See* Stephan, *supra* note 8, at 204 ("[T]he notion that the Constitution with equal force forbids distinctions based only on the subject matter of expression, or on any aspect of its content, was new."). The Supreme Court recently expanded that principle even further in last term's case, *Reed v. Town of Gilbert*, 135 S. Ct. 2218, 2232 (2015) (finding sign code that singles out specific subject-matter categories of street signs for differential treatment is content based and subject to strict scrutiny, which it fails).

did not. While not saying so explicitly, the cases suggest this dichotomy as the guiding principle of analysis for the issue of speech regulation.[103] This analysis better explains the Court's evolving decisions and better situates them within a defensible theoretical framework than does the conventional account. And, not without consequence, it also brings free-speech analysis back into the fold of more venerable methods of constitutional jurisprudence.

Through this lens, it becomes possible to see beyond mere hostility to free speech or war hysteria in the early opinions, which largely upheld convictions for sedition and incitement.[104] Instead, these regrettable decisions can be seen as recognizable incipient efforts to develop a distinction between censorial government acts—those that sanction ideas—and non-censorial acts—those that punish speech for a different purpose having to do with validly preventing harm. Consider, for example, the important dialogue among early free speech thinkers regarding the punishment of incitement to violence. This issue lies exactly at the tipping point between the censorial and non-censorial interests of government: the government seeks to punish someone for advocating the violation of law, usually for political purposes, and claims that the harm the state seeks to prevent is violence and disorder (non-censorial), not political dissent (a censorial theory). There is important nuance.

The most casual student of the First Amendment is familiar with the series of cases upholding convictions under the Espionage Act of 1917 and the Sedition Act of 1918.[105] Those cases are widely understood to reflect simply a wrong approach to the suppression of speech, applying too lenient a standard of review.[106] This reading goes a long way toward explaining

103. Stephan, *supra* note 8, at 215–19, 233–36 (making a convincing argument that, before *Mosley*, the Court was concerned only with viewpoint discrimination).

104. *See* David M. Rabban, *The First Amendment in Its Forgotten Years*, 90 YALE L.J. 514, 558 (1981) (documenting the long tradition of hostility to free-speech claims before the famous cases starting in 1919, and suggesting that the Supreme Court merely followed that tradition during and after World War I).

105. *See* GEOFFREY R. STONE ET AL., CONSTITUTIONAL LAW 1038–39 (7th ed. 2013) (discussing series of cases).

106. *See* CHEMERINSKY, *supra* note 54, at 1020 (describing the Court's approach as "mild and ineffectual"); HARRY KALVEN JR., A WORTHY TRADITION: FREEDOM OF SPEECH IN AMERICA 147 (Jamie Kalven ed. 1988) (decisions of this time are "dismal evidence" of the Court's willingness to adopt "the mood of society"); STONE, *supra* note 52, at 548 (acknowledging the Court's "undue pressures to stifle dissent in the worst of times"); Zechariah Chafee, Jr., *Freedom of Speech in War Time*, 32 HARV. L. REV. 932, 968–69 (1919) (discussing how the early cases using a bad tendency standard are "wholly inconsistent with freedom of speech" because they punish ideas); Geoffrey R. Stone, *The Origins of the "Bad Tendency" Test: Free Speech in Wartime*, 2002 SUP. CT. REV. 411, 441 ("What occurred in the federal courts in 1917-1918 was . . . the consequence of hysteria.").

the later advent of the cardinal rule as a reactionary response. But it oversimplifies.

The discussions by the major participants in the decades-long dialogue about punishing incitement demonstrate an inchoate distinction between censorial and non-censorial theories of harm. Justice Holmes, for example, famously wrote for the Court in *Schenck v. United States*,[107] that "[t]he most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic."[108] This confirms the early recognition of the distinction. The punishment of that shouter would be permitted because, although perhaps espousing an idea or message of "fire," the punishment would rest on the provocation of instant harmful conduct—panicked efforts to escape—unmediated by thought. This is a good example of a situation in which the "content" of an expression is important to its harmfulness—there would be no danger, after all, in someone shouting "Bravo" in a crowded theater—yet still, the regulation of it does not abridge the freedom of speech because the theory of harm does not rest on the suppression of an idea. I emphasize that this conclusion does not entail a judgment about the value of the speech, but rather considers the *way* that the spoken word can wreak harm.

The principle of immediacy, so salient in the early discussions of incitement, also reflects a recognition of the censorial/non-censorial distinction. Even the Supreme Court itself, as it upheld a conviction for criminal anarchy, did so only after finding that the words of the statute imply "urging to action," rather than "abstract 'doctrine' or academic discussion."[109] Justice Brandeis elaborated in his concurrences what is significant about the distinction between the two. He urged that the only way to respect "free and fearless reasoning" is to punish only that advocacy that will cause evil so immediately that no discussion or airing of the merits of the idea would be possible.[110] "If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence."[111] Putting it into the terms of my argument, the important thing was to ensure that the harmful aspect of the speech was not its persuasiveness or its power to appeal to the reason of its audience—that

107. Schenck v. United States, 249 U.S. 47 (1919).
108. *Id.* at 52.
109. Gitlow v. New York, 268 U.S. 652, 664–65 (1925).
110. Whitney v. California, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring).
111. *Id.*

kind of expression is counteracted only by more speech (and is protected). Rather, to be punishable, the speech must act to cause harm in a way *not* mediated by reason or deliberation.

The element of time emphasized by Justice Brandeis can thus be understood not as a bow to pragmatism, suggesting that quick violence is somehow more dangerous than slow violence. Rather, the time element illuminates a key component of a qualitative analytical difference between the suppression of an idea and the restriction of an act.[112] The familiar quotations from the incitement cases draw an explicit connection between a theory of harm and the objectives of the First Amendment. They make clear that the repression of an idea is never justified by the harm the idea may cause by being wrong or repugnant, but when the use of words is said to trigger immediate harm with no time for the processing of an idea, then the government is entitled to put forth at least the possibility of a legitimate, non-censorial, police-power justification for restriction. The distinction between these two types of harm, still in its infancy, was not sufficiently refined to protect free speech adequately in the early days,[113] but it contained the seeds of a principle that is fundamental to a robust speech-protection regime.

In short, the Court's error in the incitement cases was not, as the standard trope goes, in failing to require sufficient justification for suppression of ideas as such. That would be an error that could plausibly be addressed by the cardinal rule, which allows of no justifications. Rather, the Court's error lay in not drawing a robust enough line to identify when the suppression of ideas is at stake and when it is not. The cardinal rule is not responsive to that error; indeed, it exacerbates it. As the Court continued to refine its application of the inchoate censorial principle to the incitement offenses, however, it eventually, in 1969, completed the task begun a half century earlier. In *Brandenburg v. Ohio*,[114] the Court

112. Prominent academic Ernst Freund felt the line could be drawn when speech "incite[s] directly" the commission of violence and crime. ERNST FREUND, THE POLICE POWER: PUBLIC POLICY AND CONSTITUTIONAL RIGHTS 510 (1904). *See also* Rabban, *supra* note 104, at 572 (discussing Freund's view). A related view voiced by Justice Holmes was that incitement cases call for "the same reasoning that would justify punishing persuasion to murder." Abrams v. United States, 250 U.S. 616, 627–28 (1919) (Holmes, J., dissenting). The latter had been long understood as so tied up in conduct as not to be speech implicating the First Amendment at all. *See* Frohwerk v. United States, 249 U.S. 204, 206 (1919) ("We venture to believe that neither Hamilton nor Madison, nor any other competent person then or later, ever supposed that to make criminal the counselling of a murder within the jurisdiction of Congress would be an unconstitutional interference with free speech.").

113. *See* MEIKLEJOHN, *supra* note 43, at 39–45 (describing the threats to political ideals posed by the nonrigorous "clear and present danger" test of the early cases).

114. Brandenburg v. Ohio, 395 U.S. 444 (1969).

developed more stringent requirements of imminence, causation, and intent sufficient to make meaningful a distinction recognized in 1917 by Judge Learned Hand, between what he called "keys of persuasion"—or ideas—and "triggers of action"—or incitement.[115] Thus, it was not until 1969 that the offense of incitement as enforced by the Court actually transformed from a censorial offense to a non-censorial one. Incitement became, for the first time in 1969, comparable to the items that the Court had recognized as unprotected.

This interpretation of the changing debate within the Court in the early twentieth century, concerning the limits of incitement as an offense, sets the stage for a later significant contribution to the understanding of what kinds of verbal offenses are punishable. The renowned early-twentieth-century judicial philosopher, Zechariah Chafee, laid the foundation for what came to be understood as the "exceptions" to the rule against content-based regulation. In 1941, Chafee observed:

> [T]he law . . . punishes a few classes of words like obscenity, profanity, and gross libels upon individuals, because the very utterance of such words is considered to inflict a present injury upon listeners, readers, or those defamed, or else to render highly probable an immediate breach of the peace. *This is a very different matter from punishing words because they express ideas which are thought to cause a future danger to the State*.[116]

His statement recognizes the distinction that I argue was latent in the incitement cases. The very utterance of the words Chafee identified caused immediate injury, not because of the implications of an offensive idea, but by some unmediated effect of the words having been spoken. Thus, such words could be prohibited without offending the First Amendment, "a very different matter," he insisted, from punishing a subversive idea.[117] Chafee's treatise went on to discuss *why* he thought these classes of speech were capable of punishment. He was emphatic that the reason was not the historical one that these kinds of crimes existed at common law. Rather, suggested Chafee, these kinds of speech were regulable because they "do not form an essential part of any exposition of ideas, [and] have a very slight social value as a step toward truth, which is clearly outweighed by

115. *See id.* at 447 (requiring that any prosecution of incitement require imminent harm, a likelihood of producing illegal action, and an intent to cause imminent illegality). *See also* Masses Publ'g Co. v. Patten, 244 F. 535, 540 (S.D.N.Y. 1917), *rev'd*, 246 F. 24 (2d Cir. 1917).

116. CHAFEE, *supra* note 36, at 149 (emphasis added).

117. *Id.*

the social interests in order, morality, the training of the young, and the peace of mind of those who hear and see."[118]

Although there is language in the quoted statement suggesting a balancing, subsequent scholars have been too quick, I maintain, to interpret this regime described by Chafee as simply a weighing of interests to determine which categories of speech will be vulnerable to regulation without First Amendment interference. Although the language about weighing the social interests has been quoted relentlessly by the Supreme Court and scholars,[119] it occurs in Chafee's text just before a less famous passage that is the more helpful to understanding the operative principle of speech protection. The passage goes on:

> Words of this type offer little opportunity for the usual process of counter-argument. The harm is done as soon as they are communicated, or is liable to follow almost immediately in the form of retaliatory violence. The only sound explanation of the punishment of obscenity and profanity is that the words are criminal, not because of the ideas they communicate, but like acts because of their immediate consequences to the five senses. . . . When A urges B to kill C and tells him how he can do it, this has nothing to do with the attainment and dissemination of truth. . . .[120]

What Chafee's analysis reveals is that restrictions on speech that were not aimed at suppressing ideas were considered structurally *different in kind* from those that did seek to address alleged harms caused by an idea or message. Indeed, Chafee felt that the laws of the former type were "too well-recognized to question their constitutionality,"[121] while the latter were unconstitutional. Notice he did not say that obscenity could be regulated because it is of low value, but rather because it operates on senses separate from the intellect, analogously to punishable acts of a nonverbal sort.[122] The important element is that there is a response unmediated by thought.

The following year, in *Chaplinsky v. New Hampshire*,[123] the Supreme

118. *Id.* at 150.

119. *See* New York v. Ferber, 458 U.S. 747, 763–64 (1982) (exceptions apply when "the evil to be restricted . . . overwhelmingly outweighs the expressive interests"); Geoffrey R. Stone, *Free Speech in the Twenty-First Century*, 36 PEPP. L. REV. 273, 283 (2009) (identifying such speech as of "low value").

120. *Id.* at 150, 152. He says also, "[i]t is altogether different from sedition." *Id.* at 150. This supports my claim to the extent that sedition, through a lackluster reasonableness or bad tendency test, punished dissident ideas, and the doctrine had not yet evolved to require immediacy of the harm that might support a non-censorial theory for such a prosecution.

121. *Id.* at 149.

122. For further discussion of obscenity see *infra* Part III.B.

123. Chaplinsky v. New Hampshire, 315 U.S. 568 (1942).

Court invoked Chafee when it recognized the principle that some kinds of words can be outright prohibited. The Court offered a brief rationale based on Chafee's account, using some—but significantly not all—of Chafee's words. According to the Court:

> There are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem. These include the lewd and obscene, the profane, the libelous, and the insulting or 'fighting' words—those which, by their very utterance inflict injury or tend to incite an immediate breach of the peace. It has been well observed that such utterances are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.[124]

What was missing from the Court's version was Chafee's explanation that the list of words that had been regulated was not a list of *exceptions* to the First Amendment, but rather a list of *examples* of situations in which the theory of harm in support of regulation did not involve the suppression of an idea, and did not, therefore, implicate the First Amendment by abridging speech. By omitting this significant part of the analysis, the short and under-theorized statement from the Court, quoted above in its entirety, eventually came to be viewed as a list of categories of speech that were "unprotected" by the First Amendment because of their low value. This transformation of the list overlooks its historical derivation and theoretical justification, and the immanent harm principle at work in the identification of the categories.[125] Neither the Court nor Chafee had used the term "exceptions" to First Amendment, nor had either referred to a list of types of speech that were simply "unprotected" by the First Amendment.

My claim, then, is that the important line for resolving whether restrictions on the content of speech receive highest scrutiny and the heaviest presumption of invalidity is that between censorial and non-censorial theories of harm offered in support of restrictions on speech. The analysis turns on the government's claim of what harm the expression causes and how it does so. Censorial restrictions are those that rest on the possibility that expression of a dangerous idea will cause social harm

124. *Id.* at 571–72 (footnotes omitted) (citing ZECHARIAH CHAFEE, FREE SPEECH IN THE UNITED STATES (1941)). The final sentence is a close paraphrase of Chafee's sentence. *See infra* text accompanying note 120.

125. *See* TRIBE, *supra* note 68, at 583 & n.21 (stating categories subsume consideration of harm).

through the processing, influence, and spread of that idea or message.[126] In these cases, the government seeks to suppress the idea or message for the purpose of avoiding that social harm. Non-censorial restrictions, on the other hand, allege that harm will result from expression by virtue of some kind of nonrational impact that the expression can be expected to have by its utterance, not dependent on any idea or message that it may also contain. These non-censorial theories of harm can involve harm caused by the mere statement of words,[127] by an effect that results without thought,[128] or by an inextricable relationship between the expression and prohibited conduct.[129] They are undeniably harms based on the content of words, but not on any idea that the words impart.[130]

B. THE *CHAPLINSKY* CATEGORIES AS EMBODYING A NON-CENSORIAL THEORY OF HARM

A closer look at some of the items on the Court's list of kinds of speech subject to regulation reveals that these items share a qualitative characteristic that is distinct from merely an assessment that they tend to have low value. The *Chaplinsky* list comprises "the lewd and obscene, the profane, the libelous, and the insulting or 'fighting' words."[131] Most analysts have described the items on the list simply as low-value speech.[132] But the list is both over- and under-inclusive as to this characterization of

126. An example of this would be the banning of *Lady Chatterley's Lover* from libraries because it appeared to glorify the idea of marital infidelity.

127. Libel is an example of this theory. By the mere stating of a false and incriminating statement, it is said that the libeler instantly harms the victim, whether or not the statement is believed.

128. An example is an argument that a child watching violence will be more aggressive and act accordingly.

129. An example of this is a theory that a film must be banned in order to prevent the filmmakers from performing illegal conduct depicted in the film, such as a "snuff" film.

130. An idea that seeks to capture something similar to this distinction has appeared recently in Goldberg, *supra* note 40. The distinction drawn there focuses on the kinds of harm that speech can cause, distinguishing between those "unique to speech" and those that "flow from conduct." *Id.* This may provide an indirect way to identify harms targeted on what I am calling a censorial theory.

131. Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942). Significantly, the 1942 *Chaplinsky* list did not include incitement. This was true even though the Court had been at that time consistently upholding convictions for incitement, and so perhaps one could have said that it belonged on the list of offenses not precluded by the First Amendment. It is worth noting that, by the time the Supreme Court intoned the *Chaplinsky* list again in 2010, however, incitement had silently made its way onto the list of those categories of speech that the Court said could be regulated without "any Constitutional problem." United States v. Stevens, 559 U.S. 460, 468–69 (citing *Chaplinsky*, 315 U.S. at 571–72). This migration was due to the Court's having, over time, adopted a non-censorial theory of harm to undergird the offense of incitement, distinguishing it as a qualitative matter from a government effort to punish dissent, as in *Brandenburg v. Ohio*, 395 U.S. 444 (1969). *See supra* notes 114–15 and accompanying text.

132. *See* Lakier, *supra* note 86.

it. There are items on the list, such as fighting words or obscenity, that can have political import, and thus potential value to democratic self-government. And there are, of course, kinds of speech that are not on the list that may be of low value indeed, such as vitriolic hate speech or plans to make an atomic bomb. An attempt to theorize the *Chaplinsky* list, therefore, is in order.

In 1957, the Court in *Roth v. United States*,[133] confronted directly the question whether obscenity is protected expression, despite its presence already on the *Chaplinsky* list. The Court found support in history for regulating some category of expression called obscenity, but struggled to reach a definition that would reconcile that history with any sort of First Amendment theory. The real difficulty, although not articulated this way, lay in distinguishing a government theory of harm that restricted material because of its tendency to offend for its unorthodox message or idea (censorial), from that narrower category of material that could honestly be said to cause legitimate social harm—a "deleterious effect upon human conduct," as Justice Harlan wrote.[134] Only the latter, with its non-censorial justification, would be dubbed obscene, the Court insisted.

The Court settled on a definitional linchpin that would provide the dividing line between sexual material that may be banned and that which may not be. That linchpin was "prurience." Prurience, the Court emphasized, had a nonintellectual focus: "[i]tching; . . . uneasy with desire or longing; . . . having . . . lascivious longings . . . ."[135] Prurience was hardly, then or now, a household word, so why would that be the standard of choice for identifying obscenity? The importance of the Court's selection of "prurience" as a determinative characteristic, in my view, is that "prurience" is a word suggesting the incitement of a physical response. While acknowledging that sex could be a legitimate matter of public concern in the intellectual sense,[136] the Court sought to draw a line between the treatment of sex to promulgate an idea and one "appealing to the prurient interest."[137] It upheld jury instructions setting the standard as "whether it would arouse sexual desires or sexual impure thoughts" in the relevant audience.[138] The adoption of this notion of prurience as the test

133. Roth v. United States, 354 U.S. 476 (1957).
134. *Id.* at 497 (Harlan, J., concurring in part and dissenting in part).
135. *Id.* at 487 & n.20.
136. *Id.* at 487–88 (quoting Thornhill v. Alabama, 310 U.S. 88, 101–02 (1940)).
137. *Id.*
138. *Id.* at 490 (quoting the trial court's jury instructions).

suggests that the Court was concerned with the triggering of sexual arousal by erotic materials. That is important because it could (at least in theory) put obscenity onto the non-censorial side of the line, as expression that operates subrationally and not through the mediating process of thought.[139] The social harm, if any, wrought by obscenity would be a kind of behavioral response to the expression rather than the dissemination of a message.[140]

Of course, the critical line I am suggesting the Court tacitly adopted, between the trigger of a nonrational response (non-censorial) and the spread of an ostensibly immoral message (censorial), is frustratingly porous, but it is theoretically coherent. The coherence was bolstered, to some extent, by the Court's later refinement of the obscenity test in *Miller v. California*.[141] There, the Court devised a requirement that any punishable material, in addition to its appeal to prurient interests, also "lacks serious literary, artistic, political, or scientific value."[142] If the conventional account is correct,[143] that the regulation of obscenity depends only on a conclusion that prurient material, as a category, has low social value, then why add an element requiring a showing that the prurient work at issue in each case also lacks serious social value? On this account, the added *Miller* element is a redundancy, serving perhaps only to make sure that an individual case comports with the categorical presumption that prurient sexual speech is in fact of low value. This is an odd quantitative check on categorical balancing, ensuring that low-value speech is really of low value.

The non-censorial understanding of the obscenity test, however, makes better sense of the *Miller* element of redeeming social value. Recall that the non-censorial approach requires that the government articulate a theory of harm caused by an effect of speech not involving the processing of a thought or idea. If a work is "prurient" in the sense of inciting a

139. Frederick Schauer has made a valiant effort at this justification for the obscenity exception. *See* FREDERICK SCHAUER, THE LAW OF OBSCENITY 35 (1976). Without pressing the point, I seek only to demonstrate that, for a Court determined to permit regulation of obscenity, the approach it developed sought to distinguish such regulation from the regulation of content, and thus preserved the distinction I am pressing here.

140. *See* Andrew Koppelman, *Does Obscenity Cause Moral Harm?*, 105 COLUM. L. REV. 1635, 1637 (2005) (arguing that obscenity alone, of the unprotected categories, does not rest on a likelihood of harm).

141. Miller v. California, 413 U.S. 15 (1973).

142. *Id.* at 24.

143. *See* Geoffrey R. Stone, *Sex, Violence, and the First Amendment*, 74 U. CHI. L. REV. 1857, 1864–65 (2007) (listing a "noncognitive impact" as one of the factors contributing to the low value of obscene speech).

physical response, then the non-censorial element is satisfied. But it is possible that a prurient work could also, separately, seek to convey thoughts and ideas whose censorship would offend First Amendment values *at the same time* that it was evoking its physical response. Such work would be *both* prurient *and* of redeeming social value. *Miller*'s grant of protection to such work (due to its social value) makes sense under the harm theory because it ensures that the government's theory of harm is only the *non-censorial* one of banning prurience, and not the *censorial* one of banning provocative ideas. Thus, instead of being a strange redundancy, "redeeming social value" plays an essential role in speech protection by limiting the kinds of speech that can be restricted to those that rest solely on the non-censorial theory of harm. Once redeeming social value exists, a work would not qualify for a non-censorial theory of harm. Thus, this feature of the obscenity test protects not a quantitative judgment about how valuable a work is, but rather a qualitative judgment about whether the government is in fact offering a non-censorial theory of harm.

Libel, too, can be perceived through this lens. Libel, the stating of a false claim damaging to a person's reputation, has been said to cause harm, not through the message or information conveyed, but by its "very utterance."[144] Long actionable at common law, libel was "regarded as a form of personal assault," the remedy for which could be provided under the state's police powers.[145] Viewed in this concededly archaic way, libel follows the pattern of the other *Chaplinsky* categories, justified by a non-censorial theory of harm.[146]

A better understanding is needed for why libel can now be punished. While some have explained libel's presence on the unprotected list as due to its low value,[147] the Supreme Court has since rejected any idea that statements are of low value merely because they are false.[148] Of course, the Supreme Court famously narrowed the category of unprotected libel in *New York Times v. Sullivan*, in which it recognized that even false statements about public officials require some buffer against punishment.[149] By requiring proof of actual malice for recovery by a public

144. Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942).

145. TRIBE, *supra* note 68, at 631.

146. *See Chaplinsky*, 315 U.S. at 572.

147. STONE, *supra* note 105, at 1134–41 (discussing libel as one of the low-value categories of speech in *Chaplinsky*, but suggesting that it may deserve some protection in support of the protection of high-value speech).

148. *See* United States v. Alvarez, 132 S. Ct. 2537, 2546 (2012).

149. N.Y. Times Co. v. Sullivan, 376 U.S. 254, 277 (1964).

official, the *New York Times* Court made clear that knowing falsity was still punishable. That is, the First Amendment establishes "a right to speak defamatory truth," but not defamatory falsity, even about public officials.[150]

This modernization of the treatment of libel under the First Amendment makes sense under the harm principle. How does falsity work harm? The utterance of some falsehoods works harm by bypassing the rational thinking process through which society accumulates knowledge, and hijacking decisionmaking processes, not through persuasion or ideas but through deceit. When the government offers a remedy for certain falsehoods, it seeks not to shape debate, but to improve the public's ability to rely on the integrity of public debate.[151] This is a non-censorial theory of harm, justifying, in theory, the regulation of falsehoods. The Court has insisted that governments not tread too closely to the line, in order to protect against consequences such as risks of chilling speech and government entanglement in scrutinizing speech,[152] to provide "breathing space" for public debate.[153] But at the end of the day, libel and other prohibitions on harmful falsehoods can be understood as resting on a foundation of a non-censorial theory of harm. That theory is that prohibitions on certain false speech are the government's way of erecting an infrastructure in which public discourse can take place safely.

The "fighting words" category, too, easily conforms to the censorial/non-censorial framework. The Court has defined such words as those that are "likely to provoke the average person to retaliation, and thereby cause a breach of the peace."[154] The ostensible harm that governments have sought to prevent by regulating such words has been

150. TRIBE, *supra* note 68, at 635. Justice White offered a non-censorial theory of his own:

> The central meaning of *New York Times*, and for me the First Amendment as it relates to libel laws, is that seditious libel—criticism of government and public officials—falls beyond the police power of the State. In a democratic society such as ours, the citizen has the privilege of criticizing his government and its officials.

Gertz v. Robert Welch, Inc., 418 U.S. 323, 382–88 (1974) (White, J., dissenting) (internal citations omitted).

151. The government's ability to regulate false commercial advertising rests on a similarly non-censorial theory. *See* Friedman v. Rogers, 440 U.S. 1, 8–10 (1979) (discussing that "[s]ociety . . . has a strong interest in the free flow of commercial information" to enable informed consumer choices, which justifies restrictions on false, deceptive, and misleading commercial speech).

152. *See Alvarez*, 132 S. Ct. at 2547 ("Permitting the government [ban on false speech without more] . . . would endorse government authority to compile a list of subjects about which false statements are punishable. That governmental power has no clear limiting principle.").

153. *N.Y. Times*, 376 U.S. at 271–72.

154. Street v. New York, 394 U.S. 576, 592 (1969) (citing Chaplinsky v. New Hampshire, 315 U.S. 568, 574 (1942)).

violence caused by an immediate response to a personal attack. By narrowing this definition over time, the Court has made even clearer that this kind of harm cannot legitimately support regulation if it is based on the offensiveness of a word or idea; rather, to be regulated, such a verbal assault must be highly personal and likely to be responded to by an immediate violent response.[155] That is, the government must offer a non-censorial theory of harm in order to punish fighting words. By contemporary standards of what a person can be expected to tolerate, there is little, if any, speech that could meet this high standard of unmediated harm.

Understood as intoning non-censorial theories of harm, the categories in the *Chaplinsky* list are not properly viewed either as "exceptions" to the rule against content regulation or as examples of speech that is "unprotected" by the First Amendment. The list is better understood as a set of illustrative instances in which the First Amendment does not prevent regulation because there is no abridgement of speech, no censorship of a message or idea.[156] This is indeed, roughly, how the Court characterized its list in 1942. With the eventual evolution of a rule that prohibited regulation of content regardless of the theory of harm on which it rested, however, came a corresponding pressure to recognize exceptions to the absolute rule. This pressure led to a recharacterization of the *Chaplinsky* list, which happened gradually over the ensuing decades.

Such was the landscape onto which the Supreme Court entered with its *Mosley* decision in 1972. The cases had, for half a century, evinced a qualitative distinction between government actions that punished ideas or messages because of their impact as ideas and those that punished speech for reasons other than the impact of its ideas or messages. There was no hint of the cardinal rule.

IV. THE "ABSOLUTE" PROTECTION AGAINST CONTENT RESTRICTION

The pivotal case of *Police Department of Chicago v. Mosley*[157] seems inconsistent with this revisionist account of prior history. A close reading

155. *See* Cohen v. California, 403 U.S. 15, 16, 18–20 (1971) (disallowing the punishment of the words "fuck the draft" because offensiveness is an inappropriate harm to support regulation).

156. *See* TRIBE, *supra* note 68, at 605 (describing *Chaplinsky* as a case "in which the Court singled out certain categories of speech as not representing 'speech' within the meaning of the first amendment").

157. Police Dep't of Chi. v. Mosley, 408 U.S. 92 (1972).

of the opinion, however, suggests that perhaps it was not so much a departure from that history as one might think. A discussion of the opinion will help make the point.

Mosley involved an ordinance passed by the City of Chicago, prohibiting picketing near schools, with an exception for picketers having labor disputes with a school. Plaintiff-Respondent Mosley had a different kind of dispute with a school: he was protesting a particular high school's under enrollment of African-American students, and he challenged the anti-picketing ordinance on several constitutional grounds. In the process of striking down the ordinance, the Court uttered the momentous statement: "[A]bove all else, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content."[158]

This statement—voicing what has since become the cardinal rule—invited overreading because of the tone of absolute prohibition that it appears, in isolation, to adopt. Looking at the statement in context, however, one can see that the statement was not written to be read for all it could be, nor was the broad reading even justified as an accurate statement of then-existing law. Moreover, there is no reason to believe that Justice Marshall, or the Court for which he wrote, intended the dictum to impose an absolute prohibition of all government regulation of content, for which it has since been cited as authority. To the contrary, there is every reason to believe that he meant to call for meaningful scrutiny of any such restriction, in search of legitimate, non-censorial reasons on the part of the government.

It would be surprising if Justice Marshall, of all people—the avid proponent of a so-called sliding scale for equal protection[159]—adopted such an absolute posture with respect to the closely related right of free speech. Indeed, in *Mosley*, he actually stated that he viewed this case the same way that he viewed equal protection, discussing the special treatment that the Chicago picketing ordinance gave to labor: "[T]he crucial question is whether there is an appropriate governmental interest suitably furthered by the differential treatment."[160]

158. *Id.* at 95.

159. *See* San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 109 (1973) (Marshall, J., dissenting) (advocating an approach that adjusts the care with which the Court will review a claim of discrimination in light of the constitutional significance of the interests affected and the invidiousness of the classification).

160. *Id.* (first citing Reed v. Reed, 404 U.S. 71, 75–77 (1971); then citing Weber v. Aetna Casualty & Surety Co., 406 U.S. 164 (1972); and then citing Dunn v. Blumstein, 405 U.S. 330, 335

The three cases Justice Marshall cited—*Reed v. Reed*, *Weber v. Aetna Casualty & Surety Co.*, and *Dunn v. Blumstein*—were all Equal Protection cases, including one decision, *Blumstein*, that he had authored just three months earlier. In that opinion, he explicitly acknowledged the Court's obligation to look to "the character of the classification in question; the individual interests affected by the classification; and the governmental interests asserted in support of the classification,"[161] in determining whether a state restriction violated the Constitution. This approach did not, he had emphasized, "have the precision of mathematical formulas. The key words emphasize a matter of degree. . . ."[162] His citation in *Mosley* to *Blumstein*, therefore, signals a continued commitment to the sliding scale approach.[163]

The majority opinion in *Mosley* reflected what was a characteristic concern of Justice Marshall's, that the ordinance might be read to discriminate among people without a legitimate reason for the differential treatment.[164] Indeed, the brief of the would-be picketer, Respondent Mosley, to the Court put it that way explicitly.[165] It was clear, however vigorously the city argued that the labor exemption was a reasonable differentiation,[166] that the Court was worried about the city taking sides. A "government may not grant the use of a forum to people whose views it

(1972)).

161. *Dunn*, 405 U.S. at 335.

162. *Id.* at 342–43.

163. The full articulation of Justice Marshall's approach, which some have called the sliding scale, came the following year in his dissent in *San Antonio Indep. Sch. Dist. v. Rodriguez*, 411 U.S. 1, 70–133 (1973) (Marshall, J., dissenting).

164. *See* Rebecca L. Brown, *Deep and Wide: Justice Marshall's Contributions to Constitutional Law*, 52 HOW. L.J. 637, 650 (2009) ("The sliding scale approach to equal protection scrutiny is the most obvious manifestation of Justice Marshall's understanding of equality as an obligation on government to produce public-regarding reasons for its acts.").

165. The brief explained as follows: "Had Respondent . . . sought to picket claiming that the Board of Education discriminates against the hiring of black teachers, the 150-foot ban would not apply. Yet, because Respondent picketed asserting that black students were discriminated against he is prohibited from expressing these ideas by means of a sign on the public sidewalk adjacent to the school." Brief of Respondent at 19–20, Police Dep't of Chi. v. Mosley, 408 U.S. 92 (1972) (No. 70-87).

166. *See* Brief for the Petitioner at 27–28, *Mosley*, 408 U.S. 92 (No. 70-87). The doctrinal pedigree of the case is complicated. At the Supreme Court, Mosley made two distinct arguments for why the ordinance was unconstitutional: first, it was overbroad under the First Amendment; second, it created a classification that was invalid under the Equal Protection Clause. The Court did not address the overbreadth argument, nor did it directly address the Equal Protection Clause challenge in equal protection terms (such as whether there was a rational basis for the classification between labor and nonlabor picketers), but rather read an equality component into the First Amendment, which effectively gave the highest scrutiny to the classification because of its content-based nature. *See Mosley*, 408 U.S. at 101.

finds acceptable, but deny use to those wishing to express less favored or more controversial views."[167] But there was nothing about this nondiscrimination principle that necessitated the broad cardinal rule. Quite the opposite, it invited the more traditional and narrow approach presumptively invalidating what we now call "viewpoint discrimination," which is the most suspect of all possible content-based regulation. And even for viewpoint discrimination, the *Mosley* opinion eschewed a *per se* rule:

> [T]here may be sufficient regulatory interests justifying selective exclusions or distinctions among pickets. Conflicting demands on the same place may compel the State to make choices among potential users and uses. And the State may have a legitimate interest in prohibiting some picketing to protect public order. But these justifications for selective exclusions from a public forum must be carefully scrutinized.[168]

The city was unsuccessful in persuading the Court that its distinction was justified by the unique factual and legal characteristics of labor picketing.[169] That conclusion is a far cry, however, from a holding that regulation of content can never be justified.

It is the height of irony that Justice Marshall's opinion in *Mosley*, due to a dictum neither needed by the analysis nor—read in isolation—consistent with the discussion that surrounded it, was seized upon and conscripted as a foundation for an edifice that has come to stand for the exact opposite of Justice Marshall's deep-seated commitment to a judgment-based protection of equality and liberty under the Constitution.

The cardinal rule's birth story, of course, has no bearing on its robustness as a doctrinal principle today. There is no question that it has been reiterated, followed, and reinvested with precedential authority by subsequent cases. But the story highlights one very important feature of the cardinal rule: it became law without ever having been the focus of serious engagement by the Supreme Court on the question whether it is indeed the right rule of interpretation of the First Amendment's freedom of speech guarantee.

167. *Mosley*, 408 U.S. at 96 (speaking of both the First Amendment and Equal Protection Clause).

168. *Id.* at 98–99. Note that Justice Marshall made no effort to make the distinction, now so foundational, between content-based and content-neutral restrictions.

169. The city argued that labor picketing tended to be less disruptive to schools than other forms of protest, including protests against bussing. While "[l]abor picketing is now usually token picketing," there had been some 6,000 reported incidents of violent civil rights demonstrations at public high schools in the prior year, 1969. Brief for the Petitioner, *supra* note 166, at 29–30. The city also hinted that regulating labor picketing could raise risks of "federal pre-emption." *Id.* at 28–30.

A. THE EXPANDED UNDERSTANDING OF CONTENT-BASED RESTRICTION

Before *Mosley*, a sense that ideas may not be suppressed was in evidence,[170] but not in the sense that the Court recognized it in *Mosley* and since. Early commentators on this issue appear to have assumed that restricting "content" meant government preference for or against a particular point of view.[171] *Mosley* is not inconsistent with this understanding.[172]

The principle, of course, is correct, and it was not a novel proposition.[173] But the problem with the *Mosley* language is that it went way beyond viewpoint discrimination to include even viewpoint-neutral regulation of "subject matter" and "content" on the list of impermissible targets of government restriction, without any limiting principle to reflect the theoretical foundation of the proposition, equality of ideas.[174]

The influential 1978 treatise of Laurence Tribe helps put the problem into perspective.[175] This treatise, published just six years after *Mosley*, offered an incisive synthesis of the then-current law governing expression. It identified the only two ways, labeled Track One and Track Two, in which government might abridge speech. Track One, where laws are highly

170. *See* Cohen v. California, 403 U.S. 15, 19 (1971) (Cohen could not be prosecuted for wearing offensive words on his jacket if the punishment is for "the substantive message it conveys."). The terms "content based" and "content neutral" do not appear until after *Mosley*. *See* TRIBE, *supra* note 68, at 584–85 (equating "content-based abridgments" with laws "targeted at ideas or information").

171. *See* McDonald, *supra* note 10, at 1355–56 (describing history of content regulation).

172. *Mosley*, 408 U.S. at 96 ("There is an 'equality of status in the field of ideas,' and government must afford all points of view an equal opportunity to be heard." (footnote omitted)). The question of whether to consider a statute viewpoint neutral when it facially applies to both sides, but there is a much greater likelihood that one side will benefit from the provision, is a complex one that has divided the Court. I have always thought the early case of *Adderley v. Florida*, 385 U.S. 39 (1966), involving protests outside a jailhouse, was an example of this. The Court found the prohibition on protests constitutional because "[t]here [was] not a shred of evidence . . . that . . . the sheriff objected to what was being sung or said by the demonstrators. . . ." *Adderley*, 385 U.S. at 47. It seems likely, however, that protestors in front of a jail would tend to share a point of view. The Court addressed a similar issue in the context of a ban on approaching patients entering a health facility. The Court divided over whether this was content based or content neutral, as allegedly it was only abortion protestors who were likely to want to commit the prohibited acts. *See* Hill v. Colorado, 530 U.S. 703, 719 (2000) (finding statute content neutral). The Court appeared to reaffirm that understanding last term. *See* McCullen v. Coakley, 134 S. Ct. 2518, 2539–41 (2014) (concluding that buffer zone around abortion clinics was content and viewpoint neutral, but nevertheless invalidating it on narrow tailoring grounds).

173. *See* Stephan, *supra* note 8, at 217–18 (demonstrating that, until at least 1965, the Court's primary commitment was to prevent viewpoint discrimination).

174. *Id.* at 204 ("[T]he notion that the Constitution with equal force forbids distinctions based only on the subject matter of expression, or on any aspect of its content, was new [with *Mosley*]".).

175. *See* TRIBE, *supra* note 68.

suspect, applies when government "aim[s] at ideas or information," either because of the specific viewpoint expressed or because of "the effects produced by awareness" of the expression, known as communicative impact.[176] These methods of abridging speech would now be called "content-based." Track Two offers a more deferential approach when government does not aim at ideas or information, but restricts the flow of expression in a manner not based on the content of the expression, but on its "noncommunicative impact," such as a ban on all handbills to reduce litter, now called content neutral.[177]

Much of this analysis is still applicable today. But there is one highly significant respect in which it is dated: it fails to include an issue that has become salient only recently, following the unacknowledged expansion of the *Mosley* rule. What is missing from Tribe's account is the situation in which, unlike Track One, a law does *not* take aim at ideas or communicative impact, but still does not qualify for Track Two either, because the law is not neutral as to content. This is the problem at the heart of this Article: a harm that is produced by the content of expression, but not from its message. The problem arises in the issue of violent video games, crush videos, and virtual child pornography, all of which have come before the Court in recent years.[178] In all of those cases, the government seeks to remedy a harm that is the direct result of the content of the expression, but has nothing to do with the suppression of an idea. This is not a situation in which the law is merely neutral with respect to viewpoint. It is also uninterested in any thought or idea contained in the expression. The possibility that that kind of regulation could violate the First Amendment was simply not contemplated as a problem covered by the *Mosley* rule, and not even addressed by the leading analyst of the law following *Mosley*.[179] Indeed, on a related point, Tribe recognized that where the "message is irrelevant to the regulation," a law "is not a content regulation at all."[180] The idea that the Supreme Court would later mechanically apply the full

176. *Id.* at 580 & n.9 (first citing John Hart Ely, *Flag Desecration: A Case Study in the Roles of Categorization and Balancing in First Amendment Analysis*, 88 HARV. L. REV. 1482 (1975); and then citing T.M. Scanlon, *A Theory of Freedom of Expression*, 1 PHIL. & PUB. AFF. 204 (1972)).

177. *Id.* at 580.

178. *See* Brown v. Entm't Merchs. Ass'n, 131 S. Ct. 2729 (2011); United States v. Stevens, 559 U.S. 460 (2010); Ashcroft v. Free Speech Coal., 535 U.S. 234 (2002).

179. Tribe did address the problem of a *facially neutral* law that had a *content-based* motivation, and would consign those laws to Track One. An example would be the firing of a teacher for incompetence when there is evidence that the dismissal was motivated by the teacher's criticism of the school board. TRIBE, *supra* note 68, at 592 (discussing Pickering v. Bd. of Educ., 391 U.S. 563 (1968)). But Tribe did not contemplate the converse, where a *content-based* law has a *content-neutral* motive.

180. *Id.* at 606 ("The message is irrelevant to the regulation.").

force of the absolute rule against content regulation to such laws has no basis in the contemporaneous understanding of the cardinal rule or its animating commitments.

The Tribe discussion underscores a principal point in First Amendment analysis: the strong presumption against validity of a law applies whenever the government is acting to disfavor an idea. This principle is consistent with all First Amendment theory, and with all case law up to and including a modest reading of *Mosley*. When that concern is absent, however, there should be a more sympathetic consideration of what the government seeks to accomplish.

B. REVEALING RETREATS FROM THE ABSOLUTE RULE

In indirect ways, when intuition has overcome empty formalism, the Court has demonstrated sympathy for exactly this understanding. In *New York v. Ferber*, the Court added a certain kind of child pornography to the list of speech not protected by the First Amendment in the course of addressing a state law banning sale or possession of depictions of child sex abuse.[181] While recognizing in passing that this law was a content-based regulation,[182] it found the expression unprotected, in part because the concern underlying the rule against content-based regulation "bears no connection to the issue of whether a child has been physically or psychologically harmed in the production of the work."[183] That is, the theory of harm was entirely focused on the acts committed in making the expression, not on the offensiveness or disfavor of any ideas in the expression itself. The Court acknowledged that the case presented no question "of censoring a particular literary theme or portrayal of sexual activity."[184] In other words, the censorial concerns at the heart of the First Amendment were not present, and that persuaded the Court that this expression should be an "exception" to the cardinal rule.[185] This is the only such "exception" the Court has recognized since 1942.

While not creating any other exceptions as such, the Court has

181. New York v. Ferber, 458 U.S. 747, 758, 764 (1982).

182. *Id.* at 763–64.

183. *Id.* at 761.

184. *Id.* at 763.

185. Looking back on this case from the vantage point of 2010, however, the Court appeared to cabin it, not by its theory of harm, but by its association with a "long-established category of unprotected speech," that is integral to commission of a crime. United States v. Stevens, 559 U.S. 460, 471 (2010).

demonstrated many times, never explicitly, that it is willing to uphold content-based regulations when it discerns that the government's theory of harm does not involve the suppression of ideas. A classic example is *City of Renton v. Playtime Theatres*, decided fourteen years after *Mosley*.[186] That case involved a city ordinance that imposed special zoning restrictions on theaters exhibiting adult films. The ordinance qualifies easily under *Mosley*'s definition of a content-based restriction, as a particular theater's amenability to the zoning requirements was determined by the nature (content) of the films it showed. To the chagrin of a cacophony of scholarly critics,[187] the Court determined, however, that the ordinance was not based on content, and thus not subject to the cardinal rule. It upheld the ordinance on the ground that the ordinance met the low-level standard of reasonableness reserved for content-neutral laws.[188] There is widespread agreement that this was a capricious analysis. But its significance here is not that it was wrong, or result oriented, or biased. The case is significant because it reveals the deep-seated instinct of the Justices to resist the cardinal rule itself. This becomes clear in the rationale that the Court gave for finding the rule to be content neutral. The Court looked at the reasons advanced by the city in support of its ordinance. (Notice that this examination of reasons would not even take place if the cardinal rule were in full effect.)

The Court found that the city's justification for the ordinance was the prevention of what it termed "secondary effects": "to prevent crime, . . . maintain property values, and generally 'protect[t] and preserv[e] the quality of . . . urban life,' not to suppress the expression of unpopular views."[189] The Court was quite right to sense that the justifications offered for the restriction were indeed relevant to whether the regulation should be struck down.[190] However, these justifications have nothing whatever to do with whether the regulation is based on content of expression.[191] But, reasoning backwards, because the Court was satisfied that the city did not have censorial interests underlying its content-based

186. City of Renton v. Playtime Theaters, Inc., 475 U.S. 41, (1986).

187. *See* Marcy Strauss, *From Witness to Riches: The Constitutionality of Restricting Witness Speech*, 38 ARIZ. L. REV. 291, 314–15 (1996) (discussing criticism).

188. *Renton*, 475 U.S. at 47–49.

189. *Id.* at 47–48 (alteration in original).

190. Once finding the ordinance to be content neutral, the Court did pay attention to such factors as the degree and kind of burden, existence of alternative outlets for the expression, and the like—none of which is considered when a regulation is determined to be content based.

191. *See id.* at 56–57 (Brennan, J., dissenting) (arguing that the regulation was content based, not content neutral).

ordinance, it concluded that the law should be upheld in the interest of valid police-power objectives.[192] The cardinal rule precluded the Court from doing this analysis expressly—by acknowledging that the justifications offered were sufficient to justify the non-censorial theory of harm—and so it developed the evasive tactic of finding the law to be content neutral rather than address the inadequacy of the cardinal rule itself.

The Court was right to intuit that the theory of harm alleged by the city in support of its zoning rule was not censorial in its nature. Thus, barring a showing of pretext, an automatic invalidation, or even imposition of strict scrutiny, was indeed uncalled for in the case. The "secondary effects" analysis was an effort to capture this principle, inartfully. But the technique of using those instincts to conclude that the law was not based on content at all is pernicious because it drives underground the critical judgment at the heart of all constitutional adjudication. In the free-speech context, that judgment requires consideration of whether the government is seeking to suppress an idea or message, and if not, whether the government's non-censorial theory of harm is justified by good police-power reasons. The *Renton* case shows that if a draconian doctrinal rule purports to prevent the intuitively correct constitutional analysis, the intuition will emerge in another form, just like squeezing a balloon and watching the air bulge out in another place. That is hardly the stuff of enlightened constitutional discourse.

Further support for the Court's intuitive discomfort with the cardinal rule, but unwillingness to reconsider it, lies in the treatment of cases involving restrictions on speech based on their subject matter or topic. Recall that these restrictions are explicitly included within the definition of content-based restrictions put forth in *Mosley*, and thus doomed to invalidation. Thus, laws regulating lewd speech, nudity, or political advertising would fall into the category of content regulation that is based on subject matter. Some scholars believe that all such laws should be treated with the full force of the cardinal rule.[193] But rather than approach the Court's uneven treatment of such cases as evidence that the Court is sometimes right and sometimes wrong,[194] I suggest these cases can be

192. *Id.* at 54. *Cf.* Ward v. Rock Against Racism, 491 U.S. 781, 791 (1989) (finding "[t]he principal inquiry in determining content neutrality" to be whether the regulation was adopted "because of disagreement with [its] message").

193. *See* Stone, *supra* note 64, at 100–07.

194. *See id.* at 83 ("[T]he Court has encountered considerable difficulty in attempting to make

understood better as evidence of the same unarticulated instinct coming through to refine the straitjacket of the cardinal rule in a manner similar to what happened in *Renton*.

When the Court has sensed that a subject-matter restriction poses a risk of censorship of an idea, then it has invoked the cardinal rule.[195] When the Court has sensed that a subject-matter restriction is not problematic in that way, the Justices have upheld it.[196] These cases present strong evidence that the content-based nature of a restriction is simply not the dispositive characteristic that gives rise to concern about constitutionality. But that has not been acknowledged.

Even more revealing than the Court's reluctance to strike down content-based laws that it found to be non-censorial is the Court's symmetrical willingness to strike down a law even though content neutral, yet resting on a censorial theory of harm. In *Bartnicki v. Vopper*, the Court ruled it would violate the First Amendment to apply a criminal punishment to an innocent publisher's publication of an illegally intercepted telephone call,[197] even though the statute calling for such punishment was content neutral.[198]

The government offered two theories of harm to support punishment of publication: (1) to deter the initial illegal interception; and (2) to protect the privacy of the victims of the illegal wiretap.[199] The first theory is a non-censorial theory, but the Court rejected it because there was no legislative or empirical evidence to support the deterrence prediction.[200]

sense of these cases."). Sometimes the Court senses a non-censorial theory when a statute vests in private hands a power to make distinctions based on subject matter of speech, and thus does not risk government censorship. *See, e.g.*, CBS v. Democratic Nat'l Comm., 412 U.S. 94, 130–32 (1973) (upholding a law allowing broadcasters to distinguish between commercial and public-issue speech); Rowan v. U.S. Post Office Dep't, 397 U.S. 728, 731, 739–40 (1970) (upholding statute allowing homeowners to block mailings that they found "offensive because of their lewd and salacious character").

195. *See, e.g.*, Erznoznik v. City of Jacksonville, 422 U.S. 205, 209, 211–12 (1975) (invalidating prohibition on drive-in movie theater showing of nudity, as content based).

196. *See* Young v. Am. Mini Theatres, Inc., 427 U.S 50, 70 (1976) (upholding zoning restriction for sexually explicit movie theaters, finding that the regulation operates "without violating the government's paramount obligation of neutrality"); *Rowan*, 397 U.S. at 731, 740 (upholding statute allowing homeowners to block delivery of mail that is offensive because of its "lewd and salacious character"); Lehman v. City of Shaker Heights, 418 U.S. 298, 304 (1974) (upholding rule allowing city buses to display commercial, but not political, messages).

197. *See* Bartnicki v. Vopper, 532 U.S. 514, 517, 534–35 (2001).

198. *Id.* at 526.

199. *See id.* at 529.

200. *Id.* at 530–31 & n.17.

With regard to the second theory of harm, the Court treated it as if it were a censorial theory of harm. That is, the Court understood this second justification as a claim that the publication of the wiretapped conversation was punishable *because of the damaging effects of what was said* in the conversation (rather than how the recording was procured).[201] While the Court could have plausibly understood the government's claim differently,[202] it manifestly chose the censorial interpretation when it described the law as imposing "sanctions on the publication of truthful information of public concern."[203] The truthfulness and public concern the Court associated with the recording at issue are qualities about the nature and impact of the message (censorial), not the means by which it was obtained (non-censorial). Once down that path, the Court quite predictably found that the statute could not survive First Amendment scrutiny.[204] This strange example shows that the content neutrality of the law is less important to the Court's instinctual judgments than is the way that the government describes how its predicted harm will occur as a result of the regulated expression.

C. RECENT RETURNS TO ABSOLUTISM

In recent years, there have been cases involving restrictions that are clearly content based, but rather than restricting by subject matter, they target the effect that certain types of expression may have on an audience. I am not addressing what early scholars called "communicative impact," defined as harm caused by choices that an audience makes in response to speech.[205] That idea, of restricting speech because it may lead people to make bad decisions, is a theory of harm that has been roundly condemned because of the inappropriately paternalistic premises it adopts.[206] Thus, the

201. *See id.* at 533–34.

202. In a world where legal arguments were framed around the distinction between censorial and non-censorial theories of harm, the government could have pressed the importance of the source, rather than the content, of the recording in its justifications. This should have passed muster as a non-censorial theory of harm.

203. *Id.* at 534.

204. *Id.* at 535.

205. *See* Strauss, *supra* note 187, at 320 (arguing that government may not justify a measure restricting speech by invoking harmful consequences that are caused by the persuasiveness of the speech).

206. *See* Boos v. Berry, 485 U.S. 312, 321 (1988) (plurality opinion) (finding ban on embassy protests that would bring foreign government into disrepute in order to protect foreign government's dignity an unconstitutional content-based restriction because theory of harm rested on communicative impact of protest on listeners).

Court has struck down, for example, restrictions on advertising alcohol prices that sought to reduce consumption,[207] restrictions on display of "for sale" signs on houses to avoid white flight,[208] and restrictions on tobacco advertising to reduce consumption of cigarettes.[209] In my scheme, these laws fall on the censorial side of the line and are rightly suspect. The other type of impact, however, involves an effect that does not pass through the process of decisionmaking of the audience, but rather (allegedly) operates directly on the audience in harmful ways. A recent case in the Supreme Court supplies an example of this type of issue.

In *Brown v. Entertainment Merchants Association*, the State of California sought to justify a restriction on the sale of violent video games to minors.[210] "Violent video games" were defined as those in which the player of the game is interactively engaged in "killing, maiming, dismembering, or sexually assaulting" a virtual "human being."[211] The Court identified this as a restriction on the content of protected speech, and invalidated it. Not only did the Court reject the state's argument that such speech[212] should be ruled unprotected for minors, but it chastised the state for urging an approach that was "unprecedented and mistaken."[213] Although the Court characterized the restricted games as "speech about violence,"[214] the state's theory of harm was not that the speech was "about" violence in the sense used in the subject-matter-restriction cases. Rather, the state argued that the act of participating in these extreme violent acts

207. *See* 44 Liquormart, Inc. v. Rhode Island, 517 U.S. 484, 503 (1996) ("The First Amendment directs us to be especially skeptical of regulations that seek to keep people in the dark for what the government perceives to be their own good.").

208. *See* Linmark Assocs., Inc. v. Township of Willingboro, 431 U.S. 85, 96–97 (1977).

209. *See* Lorillard Tobacco Co. v. Reilly, 533 U.S. 525, 555–56, 561 (2001). *See also* Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748, 770 (1976) ("It is precisely this kind of choice, between the dangers of suppressing information, and the dangers of its misuse if it is freely available, that the First Amendment makes for us.").

210. Brown v. Entm't Merchs. Ass'n, 131 S. Ct. 2729, 2732–33 (2011).

211. *Id.* at 2732. The statute contained a test tracking the *Miller* obscenity test to ensure that the covered games "appeal[ed] to a deviant or morbid interest of minors" that "is patently offensive to prevailing standards in the community as to what is suitable for minors," and that they "lack[ed] serious literary, artistic, political, or scientific value for minors." *Id.* at 2744 (citing Cal. Civ. Code § 1746(d)(1)(A) (West 2009)).

212. The concept of speech has expanded to include virtually everything that is verbal or visual, and so whether a game was speech at all was not a question that detained the court long. *See id.* at 2733 ("[W]e have long recognized that it is difficult to distinguish politics from entertainment, and dangerous to try."). No one seriously challenges this expansion of the idea of speech, but perhaps there should be an inquiry about whether doctrinal rules adopted when speech was more narrowly defined should be reexamined.

213. *Id.* at 2735–37.

214. *Id.* at 2735.

worked an unseen effect on the minors, causing them to act out in the real world aggressively, thus working harm on the minors themselves.[215] While there was legitimate dispute about the empirical validity of this claim, it is clear that this is a non-censorial theory of harm and should not be subjected to the rigors of the cardinal rule. Analytically, it is indistinguishable from the kinds of harm that have been recognized as supporting "exceptions" to the cardinal rule, yet because there is no tradition in society of recognizing this particular kind of harm—indeed doing the empirical work to support such a theory of harm is only recently even imaginable and still elusive—the Court simply dumped the law into the content-based category and was essentially done with its analysis. No serious engagement with the state's justifications was needed once it intoned the cardinal rule.[216]

The same fate met the government's argument in *Stevens*, discussed above,[217] which sought to defend a prohibition on the sale or possession of depictions of extreme animal cruelty.[218] The federal government asked the Court to deem such depictions as unprotected, using the "categorical balancing" that has been offered to describe the principle at work in the list of unprotected categories.[219] Because this speech lacked expressive value, in the government's view, the First Amendment should not bar its regulation. As previously noted, the government's argument was rejected as "startling and dangerous."[220] But what the Court did not consider is the government's theory of harm. The Solicitor General argued that the violent and unlawful animal cruelty depicted in the videos was itself, in many cases, committed for the sole purpose of creating the video, and that the government needed to be able to prosecute the sale and possession in order to detect and curb the underlying unlawful conduct.[221] This is a non-censorial theory of harm. Rather than arguing about whether or not depictions of animal cruelty are protected by the First Amendment, which even the dissent engaged in, it would have been much more fruitful to inquire whether there was any implication of a thought or idea or

215. *See id.* at 2738–39. This kind of harm can also be understood as a harm to others caused by aggressive minors. Schauer, *supra* note 2, at 98–100.

216. Justice Alito's opinion pointed out the irony of using tradition-based limits in analyzing new technologies and new kinds of harm, but joined the judgment on vagueness grounds. *See Brown*, 131 S. Ct. at 2746 (Alito, J., concurring).

217. *See supra* notes 90–91 and accompanying text.

218. United States v. Stevens, 559 U.S. 460, 481–82 (2010).

219. *Stevens*, 559 U.S. at 469–70. The approach defended in this Article is not at all the one that the government proffered in *Stevens*, which advocated balancing.

220. *Id.*

221. *See id.*

government suppression of a disfavored message in the proffered theory of harm. There was not. Thus, there was no reason why the cardinal rule, which the Court quoted from *Mosley*,[222] should have been involved.[223] The Court distinguished its earlier analogous analysis allowing the regulation of child pornography in *New York v. Ferber*,[224] by seeking to link the statute in *Ferber*, but not that in *Stevens*, to long-established categories of unprotected speech (speech incidental to criminal conduct).[225] There was no discussion of why tradition should be the test. The theories of harm offered in the two cases were indistinguishable. But the *Ferber* law had to do with expression involving sex, which has traditionally been open to some regulation, while the *Stevens* law had to do with expression involving animal cruelty, which reflects more modern sensibilities about social harm. The pairing of these two cases alone, I submit, demonstrates that the theoretical inquiries and commitments that should be driving the protection of speech in our society are in need of examination.

V. THE HARM PRINCIPLE IN ACTION

The analysis I suggest is emphatically not a balance.[226] Rather, it identifies a qualitative line to distinguish those content-based restrictions that should be subject to rigorous scrutiny from those that should not. The line involves the kind of harm that the government is alleging.

The regulation of hate speech presents a useful theoretical test for the harm principle. Legitimacy of such a restriction would depend on the theory of harm that the government offered. If the government sought to prohibit speech because it was offensive and caused distress to individuals, its restriction would rest on a censorial theory of harm and would have to be subjected to strict scrutiny. This is tantamount to the proscription of bad or offensive ideas.

As a matter of political theory, however, it is possible to construct an understanding of an appropriate social order in which certain kinds of hate speech cause harm irrespective of their wounding effects. Jeremy Waldron

222. *Id.* at 467–68 (quoting statement of cardinal rule in Ashcroft v. ACLU, 535 U.S. 564, 573 (2002)).

223. *See id.* at 481–82. The Court actually resolved the case on the ground of overbreadth, but it addressed the substantive question in response to the government's primary and unsuccessful claim that the material was unprotected by the First Amendment.

224. *See supra* note 181 and accompanying text.

225. *See Stevens*, 559 U.S. at 471–72.

226. *See* Ely, *supra* note 176, at 1506 (concluding that "balancing tests are simply not the stuff on which reassurance can confidently be built").

has made just such a case.[227] He argues that it is a primary obligation on just government to supply a public culture in which all persons are assured of their status as citizens in good standing.[228] As Waldron eloquently elaborates, the provision of dignity-based assurances is a public good that can be undermined by private citizens.[229] The harm that government would seek to prevent by prohibiting certain kinds of hate speech by private speakers, accordingly, is the harm to its own ability to meet its responsibility to its citizens.[230] This is a non-censorial theory of harm. This theory of harm would also circumscribe the kinds of speech that could be restricted while still maintaining the non-censorial character of the law. While Waldron is not speaking to a distinction between censorial and non-censorial theories of harm as such, his distinction captures the critical difference for the theory of free speech that this Article puts forward. His theory is assuredly not based on the goal of protecting people from being offended, which is a goal not permitted to government.[231] Waldron is confident that the line can be maintained between preventing offense and according the infrastructure necessary to assure dignity to all citizens.[232] At an analytical level, these two kinds of harm are distinct.

The contrast between censorial and non-censorial harms can also be illustrated by the 1980s debate about the restriction of violent pornography. An ordinance drafted by the noted feminist scholar, Catharine MacKinnon, prohibited pornography, defined as the "sexually explicit subordination of women," which included showing women ostensibly enjoying pain, rape, mutilation, and humiliation.[233] The Seventh Circuit Court of Appeals, which found the ordinance unconstitutional, did so because it viewed the effect of the ordinance as censorial and discriminatory with regard to viewpoint: behavior that glorified the subjugation and harming of women was forbidden, while depiction of other behaviors was not, and Judge Easterbrook viewed this as an effort to suppress the spread of an unpopular idea, or "thought control," as he called it.[234] Supporters of the ordinance,

227. *See* JEREMY WALDRON, THE HARM IN HATE SPEECH (2012).

228. *See id.* at 82–87.

229. *See id.* at 92–93.

230. *See id.* at 100.

231. *See* Strauss, *supra* note 187 and accompanying text.

232. *See* WALDRON, *supra* note 227, at 106.

233. Catharine A. MacKinnon, *Pornography, Civil Rights, and Speech*, 20 HARV. C.R.-C.L. L. REV. 1, 22 (1985).

234. Am. Booksellers Ass'n v. Hudnut, 771 F. 2d 323, 328 (7th Cir. 1985), *aff'd without opinion*, 475 U.S. 1001 (1986).

however, sought in part to offer a theory of the restriction as non-censorial, in that the banned material was not banned for its ideas—any person was free to preach the subjugation or humiliation of women in other forms of expression—but for the effect that exposure to such sexual materials works implicitly on the attitudes and behaviors of those who view them, allegedly increasing a propensity in viewers to exhibit violence and discrimination towards women.[235] This theory of harm, if borne out by empirical data, would supply a non-censorial basis for government to restrict some speech demonstrated to have that effect.

Is this just a rhetorical game? Not at all. It reflects, rather, a forthright engagement with a commitment to making judgments about social harm and the role of government in a democracy. As illustrated by Waldron's project, the understanding of how government appropriately achieves a well-ordered society, to use Rawls's words,[236]—or ordered liberty, to quote Justice Harlan[237]—is not a simple matter suited to mechanical doctrinal answers. There must be a judgment about how it is that words or expression may undermine social good in order to adjudge whether the law poses a threat to free thought. With that process will come changing understandings of social goods, and of the means by which they can be undermined.

Consider sexual harassment, for example. There was a time when sexually demeaning and subordinating work environments and employment actions were part of the employment landscape. Title VII of the Civil Rights Act, as interpreted by the Supreme Court, recognized a new kind of harm in such harassment—a harm caused, sometimes, by verbal expression.[238] Under the cardinal rule, anti-harassment regulations are

235. The court characterized the argument this way: "It does not persuade people so much as change them. It works by socializing, by establishing the expected and the permissible. In this view pornography is not an idea; pornography is the injury." *Hudnut*, 771 F.2d at 328. The court rejected this theory on the ground that it did not adequately separate out the injury from the "mental intermediation" involved in the alleged effects of pornography. *Id.* at 329. That is, in my terms, it was too close to a censorial theory of harm. Apparently the proponents of the legislation also pressed a more censorial argument, and did not rely solely on the non-censorial one. *See* MacKinnon, *supra* note 233, at 21–22 (suggesting that the suppression of an idea is tolerable to prevent harm to women).

236. *See* JOHN RAWLS, POLITICAL LIBERALISM 35 (1993) (quoted in WALDRON, *supra* note 227, at 82).

237. Palko v. Connecticut, 302 U.S. 319, 325 (1937). *See* Poe v. Ullman, 367 U.S. 497, 542 (1961) (Harlan, J., dissenting) (describing "the balance which our Nation . . . has struck between . . . liberty and the demands of organized society").

238. *See* 42 U.S.C. § 2000e-2(a)(1) (2012) (prohibiting employment discrimination on basis of, inter alia, sex); Meritor Sav. Bank, FSB v. Vinson, 477 U.S. 57, 60, 62 (1986) (holding claim of "hostile environment" type of sexual harassment is actionable under Title VII).

constitutionally vulnerable, as several commentators have argued.[239] After all, the argument goes, liability under the statute can depend on the content of words said or posters displayed in a workplace.[240] In the world of the cardinal rule, that is all we need to find such rules unconstitutional.[241] That simplistic response, however, takes no account of the radical changes in societal understandings of equality and dignity or the changing pragmatics of work life and public demand for protection from new harms. It deprives courts of the opportunity to judge whether government is indeed addressing a plausible harm that happens to be caused by speech, or rather is threatening freedom of thought and expression in the way that decades of free-speech jurisprudence have considered to be abridgement.

The obvious concern here is pretext. In its initial judgment about whether a theory offered by the government is indeed non-censorial, a court must determine the authenticity of the claim of harm supporting that theory. There are at least two responses to the pretext problem. One could impose a wooden rule that categorically disables government from addressing any harms caused by any expression, without consideration of whether First Amendment values are implicated or social harms are in need of legislative redress. That is the approach currently in effect under the cardinal rule, and embraced by many scholars.

Alternatively, one could put the government's feet to the fire and insist on a factual showing by the government to demonstrate that its restriction is indeed a non-censorial one. That would mean, for example,

239. Eugene Volokh and others have made this argument explicitly. *See* Kingsley R. Browne, *Title VII as Censorship: Hostile-Environment Harassment and the First Amendment*, 52 OHIO ST. L.J. 481, 548 (1991); Eugene Volokh, *Freedom of Speech and Workplace Harassment*, 39 UCLA L. REV. 1791, 1793–94 (1992). *See also* DeAngelis v. El Paso Mun. Police Officers Ass'n, 51 F.3d 591, 596–97 (5th Cir. 1995) ("[W]hen Title VII is applied to sexual harassment claims founded solely on verbal insults, pictorial or literary matter, the statute imposes content-based, viewpoint-discriminatory restrictions on speech."); Kingsley R. Browne, *Zero Tolerance for the First Amendment: Title VII's Regulation of Employee Speech*, 27 OHIO N.U. L. REV. 563, 578–79 (2001); Andrea Meryl Kirshenbaum, *Hostile Environment Sexual Harassment Law and the First Amendment: Can the Two Peacefully Coexist?*, 12 TEX. J. WOMEN & L. 67, 95 (2002).

240. Justice White saw the handwriting on the wall in 1992:

> Under the broad principle the Court uses to decide the present case, hostile work environment claims based on sexual harassment should fail First Amendment review; because a general ban on harassment in the workplace would cover the problem of sexual harassment, any attempt to proscribe the subcategory of sexually harassing expression would violate the First Amendment.

R.A.V. v. City of St. Paul, 505 U.S. 377, 409–10 (1992) (White, J., dissenting).

241. *See generally* Frederick Schauer, *The Speech-ing of Sexual Harassment*, *in* DIRECTIONS IN SEXUAL HARASSMENT LAW (Catharine A. MacKinnon & Reva B. Siegel eds., Yale Univ. Press 2004) (discussing how speech values interact with sexual harassment law).

offering data to show that violent pornography or sexual harassment in the workplace injures women, or that violent interactive video games lead to aggression in children, or that blocking publication of leaked classified information is needed to deter leaks. The objective would be to demonstrate empirically that the claimed harm is real and not a pretext for stifling disfavored messages. An empirical showing of this kind would entitle the law to more relaxed treatment, such as the doctrinal rules applicable to content-neutral regulations, in which the importance of the government interest is taken into account. The government would still be responsible for satisfying a court as to tailoring and leaving alternative avenues of expression open, but would not be subject to the virtually insuperable obstacle posed by the cardinal rule.[242]

The Court has been extremely wary of entertaining any such claims about harm.[243] Some healthy wariness probably should remain, but it would be a very auspicious jurisprudential shift if the Court would begin to entertain the data as potentially influential in its judgment. Empirical methods are expanding beyond anything imagined back in the days of *Chaplinsky* in 1942, when harm was simply presumed for certain categories of harm. It is time for the Court both to unleash the potential of social science research and to harness it. Government should bear the burden to demonstrate consequences of harmful expression that it wishes to restrict. At the same time, the Court should discipline its response to this kind of showing and put an end to using seat-of-the-pants intuitions to resolve difficult questions of ordered liberty. The important values at stake in these judgments are not well served by a jurisprudence of unexamined, tradition-bound assumptions that mechanically allow governments, for example, to protect children from obscenity but not violence.

This approach reconciles First Amendment law with other kinds of constitutional analysis, where a qualitative evaluation is made to trigger an inquiry, and then state justifications are considered. Ordered liberty describes the balance that the nation has struck between the rights of the

242. *See, e.g.*, Brown v. Entm't Merchs. Ass'n, 131 S. Ct. 2729, 2736, 2742 (2011) (striking down state law restricting sale of violent video games to minors and rejecting evidence of an effect by comparing such games to *Grimm's Fairy Tales*); Ashcroft v. Free Speech Coal., 535 U.S. 234, 253, 259–60 (2002) (striking down Child Pornography Prevention Act of 1996, prohibiting sexually explicit virtual images of children, and rejecting government's claim that viewing such materials increases the sexual exploitation of children, citing no data).

243. *Cf.* Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 641–42 (1994) (stating that content-neutral regulations are subject to intermediate scrutiny); United States v. O'Brien, 391 U.S. 367, 377 (1968) (stating that law restricting expressive conduct survives if it has an important interest unrelated to suppression of the message and burdens no more expression than necessary).

individual and the need for order.[244] Liberty analysis under the Due Process Clause begins by assessing qualitatively whether a constitutional interest is implicated, and then proceeds to assess the sufficiency of the government's reasons for restricting it.[245] In equality cases, the first inquiry addresses the nature and degree of an inequality, including whether similarly situated persons are being treated unequally, and then shifts to an examination of the government's claimed reasons for creating that inequality.[246] The trend in Supreme Court jurisprudence is toward unifying the constitutional inquiries and avoiding the rigid categories that once characterized constitutional analysis.[247] This process has been our method of analysis to navigate between the liberties protected by the Constitution and the power of our elected governments to assist in the pursuit of happiness by legislating in the common good.

The role of government in stepping in to prevent and remedy harm has deep historical roots in our Anglo-American jurisprudence. The republican traditions that left their mark on the development of American law and history have always relied on the aspiration to develop a well-ordered society, which has contemplated a thoughtful reconciliation of private rights and public welfare.[248] Indeed, at one time, any attempt to subsume public welfare in private rights was considered a "perversion" of the republican ideal "and a reversion to that dark age when governance was captured by private lords and manors."[249] Thus, there was a strong view that "[civil] liberty requires the supremacy of the law."[250]

While the republican tradition does not dictate twenty-first century constitutional doctrine, it provides a reminder that rights, critically important to the equality and liberty that define our contemporary constitutionalism, do not come without cost in kind. One need not go all the way in arguing that the Constitution *requires* states to protect against

244. *See* Poe v. Ullman, 367 U.S. 497, 542 (1961) (Harlan, J., dissenting).

245. *See, e.g.*, Lawrence v. Texas, 539 U.S. 558, 578–79 (2003).

246. *See, e.g.*, United States v. Windsor, 133 S. Ct. 2675, 2695–96 (2013) (holding government's justifications for treating gay married couples differently to be insufficient).

247. *See* Suzanne B. Goldberg, *Equality Without Tiers*, 77 S. CAL. L. REV. 481, 491 (2004) (arguing for single standard).

248. *See* WILLIAM J. NOVAK, THE PEOPLE'S WELFARE: LAW AND REGULATION IN NINETEENTH-CENTURY AMERICA 26 (1996).

249. *Id.* at 9.

250. *Id.* at 35 (quoting FRANCIS LIEBER, ON CIVIL LIBERTY AND SELF-GOVERNMENT 273 (3d ed. 1891)). *See also* Harry N. Scheiber, *Public Rights and the Rule of Law in American Legal History*, 72 CAL. L. REV. 217, 221 (1984) (describing vigorous historic role of state police power in promoting common good).

certain privately inflicted harms[251] to acknowledge still that the Supreme Court should be careful in using the Constitution to *preclude* it. In giving meaning and scope to constitutional rights, the discussion and engagement at the heart of ordered liberty are a necessary part of legitimate judicial review of state and federal law.[252]

The cardinal rule operates at cross-purposes with ordered liberty in two ways. First, it encourages evasion (efforts to label laws as content neutral when they are not), which can result in under-protection of speech. And second, it undercuts the ability of government to protect against harm, without a correlative benefit to rights. The fetishized distinction between content-based and content-neutral laws does not embody the right technique to protect precious freedoms under the First Amendment.

Interpreting the First Amendment instead as suggesting a line between censorial and non-censorial theories of harm does respect the important philosophical justifications of the First Amendment. Thus, the analysis should not be rejected on the ground that it undervalues the importance of free speech to a democracy. A short discussion of the principal theories of free speech illustrates how the censorial/non-censorial analysis actually contributes to the most salient philosophical justifications for the First Amendment.

A major foundation of free-speech theory is the importance of the societal search for truth. Justice Holmes explicitly recognized this important value, as have many others.[253] Mill identified the particular evil of suppression as the robbing of the human race of the means of judging and accessing truth.[254] Although this "marketplace of ideas" justification for protecting speech has its critics,[255] most agree that it is better than any

251. *See* Susan Bandes, *The Negative Constitution: A Critique*, 88 MICH. L. REV. 2271, 2346–47 (1990).

252. *See* Rebecca L. Brown, *The Logic of Majority Rule*, 9 U. PA. J. CONST. L. 23, 39–41 (2006) (arguing that political equality obliges government to provide reasons).

253. *See* Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting). *See also, e.g.*, Bose Corp. v. Consumers Union of U.S., Inc., 466 U.S. 485, 503–04 (1984) ("The First Amendment presupposes that the freedom to speak one's mind is not only an aspect of individual liberty—and thus a good unto itself—but also is essential to the common quest for truth and the vitality of society as a whole.").

254. MILL, *supra* note 1, at 117. *See also* Kent Greenawalt, *Free Speech Justifications*, 89 COLUM. L. REV. 119, 135–36 (1989) (defending freedom of speech as a means of discovering the truth); Harry H. Wellington, *On Freedom of Expression*, 88 YALE L.J. 1105, 1130 (1979) (defending free speech as a way of reaching the truth).

255. *See* C. EDWIN BAKER, HUMAN LIBERTY AND FREEDOM OF SPEECH 4 (1989) (noting that truth is elusive); TRIBE, *supra* note 68, at 576–77 (criticizing the marketplace of ideas concept as insufficient to protect free speech, especially given wealth disparities).

alternative, especially censorship.[256] The harm-based approach that I am pressing here protects against degradation of this truth value by focusing on the qualitative nature of the harm. Any harm purportedly caused by the spread of an idea is presumptively not a legitimate basis for regulation precisely because of the theories underlying the marketplace of ideas. Harm resulting from an impact that bypasses the rational process of communication, however, is not part of the marketplace of ideas. The person who falsely shouts "fire" in a crowded theater is not disseminating an idea, and the loss of that speech to society will not occasion an obstacle to the societal search for truth.

The marketplace of ideas and several other important justifications of free speech rest on the proposition that protecting speech has salutary consequences for society. It contributes to self-government, promotes transparency of those in power, and produces overall better social results, including tolerance.[257] A rule of content regulation that forbids the restriction of speech based on censorial theories of social harm is consistent with these instrumental justifications for the First Amendment, as it leaves intact the heavy presumption against suppression of messages or ideas. At the same time, however, it also respects the core intuition behind these instrumental theories of social utility by permitting the government to regulate when harm is the result of expression that is not espousing an idea or a message.

Other justifications for protecting speech are more "constitutive," as Ronald Dworkin called them.[258] These theories value freedom of speech because it protects the dignity and moral agency of the people. The constitutive value is implicated both for potential speakers and for potential audiences. Dworkin explained that government denies the moral responsibility of its people when it withholds from them the opportunity to "hear opinions that might persuade them to dangerous or offensive convictions."[259] Being subject to influence from a wide range of ideas, good and bad, is critically important. T.M. Scanlon clarified that these interests of the audience are not served, however, when expression

256. CHEMERINSKY, *supra* note 54, at 956–57.

257. *See* BOLLINGER, *supra* note 70, at 9–10; MEIKLEJOHN, *supra* note 43, at 27; Alexander Meiklejohn, *The First Amendment Is an Absolute*, 1961 SUP. CT. REV. 245, 255; David A. Strauss, *Why Be Tolerant?*, 53 U. CHI. L. REV. 1485, 1491–92 (1986).

258. *See* RONALD DWORKIN, FREEDOM'S LAW: THE MORAL READING OF THE AMERICAN CONSTITUTION 200–01 (1996) (arguing that free speech both allows individuals the dignity of making up their own minds, and conveys a responsibility to share those convictions).

259. *Id.* at 200.

"influences us in ways that are unrelated to relevant reasons, or in ways that bypass our ability to consider these reasons."[260] This is the kind of operation upon us that expression can have, and the suppression of which I am calling non-censorial.

Scanlon offered subliminal advertising as an example. "What is objectionable about subliminal advertising, if it works, is that it causes us to act—to buy popcorn, say, or to read Dostoevsky—by making us think we have a good reason for so acting, even though we probably have no such reason."[261] He went on to affirm an important truth relevant to the assessment of governmental restrictions on speech. "The central audience interest in expression," he wrote, "is the interest in having a good environment for the formation of one's beliefs and desires."[262] Government can contribute to that environment by protecting against some kinds of speech harms. Indeed, the autonomy interests of an audience could be enhanced by regulation of speech that causes a harmful impact that may not be obvious on the surface. Scanlon's view speaks favorably to the idea of a non-censorial First Amendment theory.

Another salient aspect of the constitutive justification for freedom of speech involves the interest of a speaker in expressing one's own convictions to others. To silence speakers is to deny the worth of their participation in society, and this transgresses the government's obligation to its people.[263] But this autonomy interest is not implicated by a rule against subliminal speech, or a rule against shouting "fire" in a crowded theater. The autonomy involved in verbal outbursts, which cause harm by being voiced, but not on the basis of any message communicated, is not different from other autonomy interests in conduct that can be regulated in the balance of ordered liberty: the harm principle should treat those acts the same because they do not involve what Dworkin calls the branding of a person's convictions as unworthy. By definition, the non-censorial theories of harm that justify police-power regulation do not silence any idea on the ground that it is offensive or unworthy of acceptance.

There is an additional autonomy-promoting benefit flowing from the recognition of a qualitative distinction between censorial theories of harm and non-censorial ones. The cardinal rule, with its absolute bar to content regulation, requires that some speech be branded as less worthy than other

260. Scanlon, *Freedom of Expression*, *supra* note 65, at 525.
261. *Id.*
262. *Id.* at 527.
263. DWORKIN, *supra* note 258, at 200.

speech. It is no accident that the so-called "two-level" theory of speech, branding all speech with a high-value or low-value label, took hold just as the cardinal rule was gaining acceptance. No sensible scheme could bar all content regulation. Laws against fraud, perjury, and conspiracy, to name just a few, require the punishment of speech based on what is said. And, as we have seen, the Supreme Court has also endorsed the punishment of certain other words characterized as obscene, libelous, fighting words, and the like. There must be some escape valve to permit such regulation, and the Court settled on branding certain kinds of speech as of "low value." This is potentially more a threat to a speaker's autonomy than a qualitative analysis that looks to the process by which the words cause harm.

Free speech enthusiasts should worry about a scheme that requires a judicial assessment of the value of speech.[264] Yet such a valuation is inevitable if the default rule is an absolute bar to the regulation of speech based on content. Above all, the approach involving theory of harm gives meaning to the interests of society at large in "living in a society that enjoys the cultural, political, and technological benefits of free expression," while still respecting "the interest we all have in not bearing the costs . . . that expression can entail."[265] These latter are the interests addressed by the police power.

CONCLUSION

I have long believed that the key to constitutional validity is a robust and incisive assessment of government power and justification. This includes a burden on government to offer a theory of how any expression that it wishes to curb contributes to a cognizable social harm, and to offer justifications supporting that theory once the theory is found to be non-censorial. It has always bothered me that those who prize the First Amendment have insisted on treating it as different in kind from the rest of our Constitution, with an exceptionalism that renders the usual process of adjudication, and the usual analysis of ordered liberty, inapplicable to the freedom of speech.

The main argument for excepting the First Amendment from the general approach to constitutionalism is the fear that, when it comes to

264. *Cf.* Scanlon, *Freedom of Expression*, *supra* note 65, at 534 (discussing the particular dangers when government is asked to balance interests in cases in which their political issues are involved: "governments are notoriously partisan and unreliable").

265. *See id.* at 155.

speech, there is a special danger in allowing a court to assess the sufficiency of government reasons in the heat of an actual case where a threat to the public good is alleged, sometimes compellingly. Better, the Court has thought, to tie itself to the mast with an absolute rule and a predetermined list of specific exceptions. This way, the categorical balancing approach of current doctrine will ensure that there is no slippage of commitment to free-speech principles.

The problem is that the world is not wooden. A rule that sometimes requires unjust results, as any absolute rule does, is going to result in slippage of its own. Any court worth its salt is going to come across regulations that should not be subject to the cardinal rule, and with the court's hands tied, it is likely to evade and dodge, in ways that are neither transparent nor subject to challenge. A rule that better effectuates foundational commitments like the harm principle has a much better chance of contributing to a coherent jurisprudence.

The Court's specific acts of infidelity to the cardinal rule over the years provide insight into what the right rule ought to be. In the end, the Court is not troubled by government regulations that restrict speech to avoid harms that are caused by the speech in ways untethered to the idea or message. In such cases, the cause of the harm might as well be action or conduct not involving speech at all. There is no need to contort the idea of speech to call it "conduct."[266] Rather, just acknowledge that the validity of suppressing speech depends on how it is claimed to harm the common good.

While there can be close cases, it is an assessment not of value but of mechanics, and thus should not be a locus of worrisome discretion. Claims of harm should be supported by persuasive data. Once a law is determined to fall on the non-censorial side of the line as a matter of theory, then it must still be justified and tested for scope and possible pretext, but without the heavy presumption of invalidity that comes with the suppression of ideas. Still, any restriction of speech, even when an idea is not implicated, can limit the amount of expression and thus there should be a showing that the law will, in fact, address a real, legitimate, police-power interest. In this age of sophisticated empirical methods, the standards should be exacting, but not insuperable.

It is not an accident that the emergence of cases involving

266. *See* TRIBE, *supra* note 68, at 598–99 (distinction between speech and conduct is unhelpful and "has no real content").

(undeclared) non-censorial theories of harm is increasing. The more we learn about human development and influences, and the more we evolve in our understandings of what liberty and equality require of a society, the greater the potential to identify harms that were not historically imagined. Just as we now regulate lead in paint, having discovered that it harms development of children, efforts have also been made to regulate certain kinds of misogynistic, racist, child-pornographic and violent expression that are alleged to cause harm in those exposed to it. This is due not only to new sensibilities to causation, but also to a new appreciation of what people may or should be protected against. A constitutional doctrine that is forever rooted in the common-law social and sociological assumptions, impervious to new discoveries, values, and heterogeneity, is doomed to fail, as *Lochner* failed.

Moreover, the development of new communication media brings with it new possibilities for the causation of harm that would never have been imaginable without advanced technology. The World Wide Web, with its ability to transmit speech globally in an instant, is an obvious example, raising new issues about the potential of communication to cause harm. The point is that such questions are hard, not easy. They are too important to be answered glibly by reference to a priori proclamations from prior ages, untethered to any theory of the freedom of speech in a democracy.

Our greatest hope for liberty lies in an expectation that courts will honestly confront the competing values at stake in a constitutional challenge, whether it be privacy versus national security, guns versus safety, or boundless expression versus a well-ordered society. That is the promise of ordered liberty. The cardinal rule against content regulation drives underground the engagement with the important tradeoffs that face a democracy, inviting subterfuge and compromise. In preparation for the next inevitable wave of dire threats to free speech, it is critical to separate out the kinds of restrictions that actually pose a danger to free speech from those that do not. Lessons from history teach us that over-protection of liberty, to the detriment of a progressive society, is not a safe path. In order to strike the right balance, the First Amendment will have to take account of the harm principle and the common good, and work out a new deal for content regulation.