

Private Speech, Public Purpose: The Role of Governmental Motive in First Amendment Doctrine

Elena Kagan†

INTRODUCTION

In one of the most frequently quoted passages of one of the most frequently cited First Amendment decisions, the Supreme Court declared that “the purpose of Congress . . . is not a basis for declaring [] legislation unconstitutional.”¹ Noting several hazards of attempting to ascertain legislative motive, the Court in *United States v O’Brien*² eschewed this endeavor in First Amendment cases, as well as in other constitutional adjudication. It was no task of the judiciary to discover or condemn “illicit legislative motive” relating to the freedom of speech; the question for courts was only whether a challenged statute, by its terms or in its application, had an “unconstitutional effect” on First Amendment freedoms.³

In keeping with this approach, most descriptive analyses of First Amendment law, as well as most normative discussions of the doctrine, have considered the permissibility of governmental regulation of speech by focusing on the effects of a given regulation. This focus on effects comes in two standard varieties. In one, the critical inquiry relates to the effect of a regulation on the speaker’s ability to communicate a desired message. In the other,

† Professor of Law, The University of Chicago (on leave 1995-96 to serve as Associate Counsel to the President). I am grateful to Mary Becker, David Currie, Richard Epstein, Richard Fallon, Stephen Gilles, Dan Kahan, Larry Lessig, Michael McConnell, Mark Ramseyer, Anne-Marie Slaughter, Geoffrey Stone, David Strauss, Cass Sunstein, and Laurence Tribe for extremely helpful comments. The Class of 1964 Fund and the Russell J. Parsons Faculty Research Fund provided financial support.

¹ *United States v O’Brien*, 391 US 367, 383 (1968).

² 391 US 367 (1968).

³ *Id.* at 383, 385. In the *O’Brien* inquiry, the nature of the governmental interest asserted played an important role. See *id.* at 380-82; text accompanying notes 237-38. But the *O’Brien* Court cared not at all—or at least professed to care not at all—whether the asserted governmental interest matched, or even resembled, the actual interest underlying the enactment or enforcement of the legislation. *O’Brien*, 391 US at 383-85.

the critical inquiry relates to the effect of a regulation on the listener's ability to obtain information. In either case, however, what matters is the consequence of the regulation.

This Article shifts the focus from consequences to sources; I argue, notwithstanding the Court's protestations in *O'Brien*, that First Amendment law, as developed by the Supreme Court over the past several decades, has as its primary, though unstated, object the discovery of improper governmental motives. The doctrine comprises a series of tools to flush out illicit motives and to invalidate actions infected with them. Or, to put the point another way, the application of First Amendment law is best understood and most readily explained as a kind of motive-hunting.

This claim stands in need of much explanation, for as *O'Brien* indicates, even the attentive observer rarely catches a glimpse of the Court inquiring directly into governmental purpose. But assume for a moment that discovering impermissible motive stood as the Court's overriding object in the consideration of cases involving the First Amendment. Would the Court then charge itself with the task of dissecting and analyzing in each case the reasons animating the action of a governmental official or body? Not likely, for obvious reasons relating to the ease of legislatures' offering pretextual motives and the difficulty of courts' discovering the real ones. Would, then, the Court have to surrender its concern with motive? Not at all. The Court could construct and use objective tests to serve as proxies for a direct inquiry into motive. It could develop rules that operated, like certain burden-shifting mechanisms or presumptions, to counter the difficulties involved in determining motive and to enable the judiciary to make, if only indirectly, that determination.

The most important components of First Amendment doctrine—indeed, the very structure of that doctrine—serve precisely this function. If a court were to attempt to devise easily manageable rules for ferreting out impermissible governmental motives in the First Amendment context, it first would create a distinction between speech regulations that are content neutral and those that are content based. It then would develop a series of exceptions to that distinction in order to handle unusually suspicious kinds of content-neutral regulations and unusually trustworthy kinds of content-based restrictions. (This effort might give rise, for example, to the doctrine of so-called low-value speech.) It would add a division of great import between generally applicable regulations, only incidentally affecting speech, and regulations specifically targeted at expressive activity. If, in other words, a

court wished to construct a set of rules to determine impermissible motive in the First Amendment area, it might well devise the complex set of distinctions and categories currently governing First Amendment law. And conversely, if a court could determine governmental purpose directly, these rules, principles, and categories might all be unnecessary.

Courts, of course, rarely construct law in so deliberate a fashion; at least, the current Supreme Court—fractured, clerk-driven, and uninterested in theoretical issues as it is—rarely does so. The self-conscious rationalization and unification of bodies of law is not something to expect from the modern judiciary. So I do not mean to stake a claim that individual Justices, much less the Court as a whole, have set out intentionally to create a doctrinal structure that detects illicit motive by indirect means. The story I tell about purpose in the law does not depend on any assertion about the purpose of the Court. What I provide is simply a reading—I think the best reading—of the Court's First Amendment cases. I contend not that the Court self-consciously constructed First Amendment doctrine to ferret out improper motive, but that for whatever uncertain, complex, and unknowable reasons, the doctrine reads as if it had been so constructed.

I do not wish to overstate the case here, though perhaps I already have done so; I am not about to craft (yet another) all-encompassing—which almost necessarily means reductionist—theory of the First Amendment. First, what follows is primarily a descriptive theory; although I discuss its normative underpinnings, I make no claim that a sensible system of free speech should be concerned exclusively with governmental motivation. Second, even seen as descriptive, the theory has limits. Some aspects of First Amendment law resist explanation in terms of motive; other aspects, though explicable in terms of motive, are explicable as well by other means; and sometimes, the concern with governmental motive is itself intertwined with other apprehensions. And yet, all these qualifications notwithstanding, the concern with governmental motive remains a hugely important—indeed, the most important—explanatory factor in First Amendment law. If it does not account for the whole world of First Amendment doctrine, it accounts (and accounts alone) for a good part of it.

Section I of this Article introduces the discussion by using a recent case—*R.A.V. v City of St. Paul*⁴—to explore how a concern

⁴ 505 US 377 (1992).

with impermissible motive underlies First Amendment doctrine. Section II moves backward to address the prior questions of what motives count as impermissible under the First Amendment and how such motives differ from legitimate reasons for restricting expression. Section III demonstrates how a wide range of First Amendment rules—indeed, the essential structure of the doctrine—are best and most easily understood as devices to detect the presence of illicit motive. Section IV concludes with some thoughts on the normative underpinnings of the Court's unstated, perhaps unrecognized, but still real decision to treat the question of motive as the preeminent inquiry under the First Amendment.

I. AN EXAMPLE: THE PUZZLE OF *R.A.V. v CITY OF ST. PAUL*

Consider first the recent, important, and hotly debated Supreme Court decision of *R.A.V. v City of St. Paul*. The decision, invalidating a so-called hate speech ordinance, raises many questions about what counts, or should count, as the core concern of the First Amendment. An exploration of some of these questions shows in dramatic form the importance of governmental motive in the Court's First Amendment analysis.

R.A.V. arose from St. Paul's decision to charge a juvenile under its Bias-Motivated Crime Ordinance for allegedly burning a cross on the property of an African-American family. The ordinance declared it a misdemeanor for any person to "place[] on public or private property a symbol . . . which one knows or has reasonable grounds to know arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender"⁵ In an effort to avoid constitutional problems, the Minnesota Supreme Court interpreted this statute narrowly to apply only to "fighting words" based on race, color, and so forth.⁶ Courts long have considered fighting words to be unprotected expression—so valueless and so harmful that the government may ban them entirely without abridging the First Amendment.⁷ The question thus raised by the state court's decision was whether St. Paul constitutionally could prohibit some, but not all, un-

⁵ St. Paul, Minn, Legis Code § 292.02 (1990).

⁶ See *In re R.A.V.*, 464 NW2d 507, 510-11 (Minn 1991). In *Chaplinsky v New Hampshire*, the United States Supreme Court defined "fighting words" as words "which by their very utterance inflict injury or tend to incite an immediate breach of the peace." 315 US 568, 572 (1942).

⁷ See *Chaplinsky*, 315 US at 571-72.

protected speech—more specifically, fighting words based on race and the other listed categories, but no others.⁸

A majority of the Court, speaking through Justice Scalia, held that St. Paul could not take this action because it violated the principle of content neutrality. No matter that a city may ban all fighting words; it may not (as, the majority held, St. Paul did) ban only fighting words that address a particular subject or express a particular viewpoint. Although the category of fighting words is unprotected—although it has, “in and of itself, [no] claim upon the First Amendment”—the government does not have free rein to regulate selectively within the category.⁹ Even wholly proscribable categories of speech are not “entirely invisible to the Constitution, so that they may be made the vehicles for content discrimination.”¹⁰ To sustain such discrimination within categories of speech, just because the categories as a whole are proscribable, would be to adopt “a simplistic, all-or-nothing-at-all approach to First Amendment protection.”¹¹

Three explanations for the Court’s decision offer themselves, the first two relating to different effects of the St. Paul ordinance, the last relating to its purpose. First, the Court might have held as it did because the St. Paul ordinance too greatly interfered with the opportunity of speakers to communicate their desired messages. Second, the Court might have reached its decision because the ordinance harmed the ability of the public—that is, the audience—to become exposed to a desirable range and balance of opinion. Third, the Court might have invalidated the ordinance because regardless how (or whether) it affected either speaker or audience, it stemmed from an improper purpose on the part of the government. Which of these three possibilities best explains the *R.A.V.* holding?

Not the first—not, that is, a perspective focusing on the speaker’s opportunity to engage in expression. As all of the Justices agreed, St. Paul could have enacted a statute banning all fighting words—a statute, in other words, imposing a more expansive restriction on speech than did the ordinance in ques-

⁸ Four Justices believed there was no need to reach this question because the Minnesota Supreme Court had failed in its effort to limit the ordinance to fighting words, and the ordinance thus remained overbroad. See *R.A.V.*, 505 US at 413-14 (White concurring). The majority of the Court, however, declined to consider this argument, *R.A.V.*, 505 US at 381, and the dispute in the case focused on the question set out in the text.

⁹ *R.A.V.*, 505 US at 386.

¹⁰ *Id.* at 383-84.

¹¹ *Id.* at 384.

tion.¹² If St. Paul could have passed this broad limitation, silencing both the speakers affected by the actual ordinance and a great many others, then the flaw in the ordinance must have arisen from something other than its simple curtailment of expression. Consider here the views of Justice White, who wrote that “[i]t is inconsistent to hold that the government may proscribe an entire category of speech . . . but that the government may not [proscribe] a subset of that category”¹³ If expressive opportunities were the only constitutional interest, Justice White would be correct that the greater restriction includes the lesser. If he erred—if the greater does not, or does not always, include the lesser—it must be because of another interest.

Perhaps, then, the interest protected in *R.A.V.* is the interest of listeners in a balanced debate on public issues. The argument, initially plausible, goes as follows. The St. Paul ordinance on its face restricted speech on the basis of subject matter; fighting words based on race, but not on other topics, fell within its coverage. More, and more nefariously, the ordinance discriminated in its operation on the basis of viewpoint; the law effectively barred only the fighting words that racists (and not that opponents of racism) would wish to use.¹⁴ The ordinance, while not restricting a great deal of speech, thus restricted speech in a way that skewed public debate on an issue by limiting the expressive opportunities of one side only. The reason the St. Paul ordinance posed a greater constitutional difficulty than a ban on all fighting words related to this skewing effect; the ordinance ensured that listeners would confront a distorted debate, thus interfering with “the thinking process of the community.”¹⁵

¹² Id at 383-84; id at 401 (White concurring); id at 415-16 (Blackmun concurring); id at 417-18 (Stevens concurring).

¹³ Id at 401 (White concurring).

¹⁴ The question whether the St. Paul ordinance, in operation, discriminated on the basis of viewpoint divided the Justices, as it has divided commentators. Contrast the majority opinion, *R.A.V.*, 505 US at 391-92 (holding that ordinance was viewpoint discriminatory), with the concurring opinion of Justice Stevens, id at 434-35 (arguing that ordinance was not viewpoint discriminatory). Contrast also Cass R. Sunstein, *On Analogical Reasoning*, 106 Harv L Rev 741, 762-63 & n 78 (1993) (no viewpoint discrimination), with Elena Kagan, *The Changing Faces of First Amendment Neutrality: R.A.V. v St. Paul, Rust v Sullivan, and the Problem of Content-Based Underinclusion*, 1992 S Ct Rev 29, 69-71 (viewpoint discrimination). It is not necessary to resolve this issue here. If the ordinance, in application, did not amount to viewpoint discrimination, then the rationale based on skewing effects becomes much weaker. To best present the claim that the effects of the ordinance justify the *R.A.V.* decision, I assume the ordinance was viewpoint discriminatory.

¹⁵ Alexander Meiklejohn, *Political Freedom: The Constitutional Powers of the People*

But on close inspection, this argument appears contrived. Even if the St. Paul statute distinguished between viewpoints on racial issues, several considerations, detailed here in ascending order of importance, suggest that the statute would not dangerously have distorted public debate.

First, the Court repeatedly has suggested that the expressive content of fighting words is, in Justice White's words, "worthless" or "*de minimis*."¹⁶ If this understanding is accepted, a concern with the distortion of public debate in a case like *R.A.V.* looks awkward, even wholly misplaced. Assuming that fighting words have no expressive value—that they are not a part of public debate because not in form or function true communication—then the restriction of some fighting words, even if all on one side, cannot easily be thought to distort discussion. True, a law of this kind subtracts from one side only, but it subtracts a thing valued at zero and thus cannot change the essential equation.

I do not mean to claim that the distortion argument has no meaning in a sphere of unprotected expression. For one thing, any restriction on racist (but only racist) fighting words inevitably will chill racist (but only racist) speech outside the fighting words category; the chilling effect of such a regulation thus will cause some distortion in the realm of protected expression. More important, though courts often claim that fighting words and other unprotected speech have no expressive content or function, these claims ought not to be taken at face value. As Justice Scalia noted, "sometimes [fighting words] are quite expressive indeed."¹⁷ Claims to the contrary serve as shorthand for a complex calculation that the harms of such speech outweigh their contribution to the sphere of expression. Thus, the subtraction of fighting words from one side of a debate is the subtraction not of a void, but of something quite tangible. And yet, in such a case, the concern with skewing the deliberative process continues to ring oddly, as it might if a law prevented one side of a debate from throwing brickbats at the other (an activity that also might be "expressive indeed"). A law of this kind would be unconstitu-

27 (Harper 1960) (emphasis omitted). Geoffrey Stone made this argument with respect to content-based regulation generally in *Content Regulation and the First Amendment*, 25 Wm & Mary L Rev 189, 198 (1983). I used the argument with respect to *R.A.V.* in Kagan, 1992 S Ct Rev at 63-64 (cited in note 14). As will become clear, I now find the argument unpersuasive, both in its application to *R.A.V.* and more broadly.

¹⁶ *R.A.V.*, 505 US at 400 (White concurring).

¹⁷ *R.A.V.*, 505 US at 385.

tional, but there is something peculiar in saying that this is because the law harms the thinking process of the community.

Second, even putting aside this objection, the skewing effect in a case like *R.A.V.* is very modest. Racists can continue to communicate their message in many ways; we need have no worry that the St. Paul ordinance will excise the idea of racism from public discourse, or indeed that the ordinance will noticeably cut into the idea's incidence. Of course, a flat rule, excluding case-specific inquiry of this kind, may have benefits. Judges, some will say, cannot reliably determine whether a given viewpoint-based law works only a modest distortion: the matter is one of degree and difficult to measure; perhaps, for example, the particular means restricted, though apparently modest, constitute the most effective way of delivering the message.¹⁸ But this insight, and the preference for rule-based approaches that goes with it, cannot explain the *R.A.V.* decision. Within a sphere of unprotected speech (such as fighting words or obscenity), the most accurate generalization is that viewpoint distinctions will not significantly distort public discourse. Were skewing effects all that mattered, the *R.A.V.* Court thus would have established a bounded exception to the usual rule against viewpoint discrimination, applying in spheres of unprotected expression.

Finally, the notion of a skewing effect, as an explanation of *R.A.V.* or any other case, rests on a set of problematic foundations. The argument assumes that "distortion" of the realm of ideas arises from—and only from—direct governmental restrictions on the content of speech. But distortion of public discourse might arise also (or instead) from the many rules of property and other law that, without focusing or intending to focus on any particular speech, determine who has access to expressive opportunities.¹⁹ If there is an "overabundance" of an idea in the absence of direct governmental action—which there well might be when compared with some ideal state of public debate—then action disfavoring that idea might "unskew," rather than skew, public discourse.²⁰ Suppose, for example, that racists control a

¹⁸ See Stone, 25 *Wm & Mary L Rev* at 225-27 (cited in note 15).

¹⁹ See Cass R. Sunstein, *Democracy and the Problem of Free Speech* 178-79 (Free Press 1993); Owen M. Fiss, *Why the State?*, 100 *Harv L Rev* 781, 786-87 (1987).

²⁰ See Sunstein, *Democracy and the Problem of Free Speech* at 178-79 (cited in note 19); Fiss, 100 *Harv L Rev* at 786-87 (cited in note 19). The formulation in the text suggests that there is such a thing as an ideal speech environment and may suggest, further, that we can describe its appearance. For doubts as to whether it is possible to provide an account of an optimal speech market, see David A. Strauss, *Rights and the System of*

disproportionate share of the available means of communication; then, a law like St. Paul's might provide a corrective.

A court well might—as the *R.A.V.* Court did—refuse the government the power to provide this corrective, but to do so, the court must discard a rationale focused purely on effects and adopt a rationale focused on motive. In denying the government a power of this kind, a court effectively determines that the “appropriate” distribution of speech is the distribution existing prior to direct governmental action. This determination, as I have noted, cannot be based on the view that the “preregulatory” distribution represents some platonic ideal of public discourse. It must be based on the view that whatever the existing state of affairs, direct restrictions, such as the St. Paul law, probably would worsen matters. And this thinking—the use of a presumption that governmental regulation will exacerbate, rather than ameliorate, distortion—is most naturally viewed as arising from a concern with the motives that underlie the regulation.²¹ The worry in a case like *R.A.V.* is not with skewing effects per se; the fear of skewing effects depends upon, and becomes meaningless without, the fear that impermissible considerations—call them for now “censorial” or “ideological” considerations—intruded on the decision to restrict expression.

The *R.A.V.* Court made this concern about illegitimate, censorial motives unusually evident in its opinion, all but proclaiming that sources, not consequences, forced the decision. The First Amendment, the majority stated, “prevents government from proscribing speech . . . because of disapproval of the ideas expressed.”²² And again: “The government may not regulate [speech] based on hostility—or favoritism—towards the underlying message expressed.”²³ The Court maintained that the structure of the ordinance—the subject-matter distinctions apparent on its face, the viewpoint distinctions apparent in operation—suggested illicit motive: “the nature of the content discrimination,” in the Court’s view, posed a “realistic possibility that of-

Freedom of Expression, 1993 U Chi Legal F 197, 205-07. The argument here does not rest on an ability to define fully an ideal state of debate; it rests only on the claim that the distribution of speech prior to direct governmental regulation need not, and usually will not, constitute such an ideal, or even something close to it.

²¹ The presumption also might be thought to arise from a view of governmental incapacity to promote a healthier or more balanced speech market. I reject this alternative explanation in Section III.A, arguing that even if this general incapacity existed (which I doubt), it would provide insufficient reason to adopt the presumption.

²² *R.A.V.*, 505 US at 382.

²³ *Id* at 386.

ficial suppression of ideas [was] afoot.”²⁴ And going beyond the structure of the law, the Court found that “comments and concessions” made by St. Paul in the case “elevat[e] the possibility to a certainty” that St. Paul was “seeking to handicap the expression of particular ideas” because of hostility toward them.²⁵

Indeed, half hidden beneath a swirl of doctrinal formulations, the crux of the dispute between the majority and the concurring opinions concerned the proper understanding of St. Paul’s motive in enacting its hate-speech law.²⁶ The majority understood this motive as purely censorial—a simple desire to blot out ideas of which the government or a majority of its citizens disapproved. The concurring Justices saw something different: an effort by the government, divorced from mere hostility toward ideas, to counter a severe and objectively ascertainable harm caused by (one form of) an idea’s expression.²⁷ In part, this different understanding of motive emerged from a different view of the structure of the ordinance: in arguing that the ordinance did not discriminate on the basis of viewpoint, Justice Stevens suggested that the Court need not fear illicit purpose.²⁸ In part, the divergent interpretations of St. Paul’s purpose reflected varying levels of sensitivity to the harms such speech causes.²⁹ In any event, the dispute was clear. “[T]his case does not concern the official suppression of ideas,” said Justice White, but only a reasonable response to “pressing public concerns.”³⁰ And Justice Stevens agreed that the ordinance had its basis not in “censorship,” but in “legitimate, reasonable, and neutral justifications.”³¹

²⁴ Id at 390.

²⁵ Id at 394. The Court noted especially a statement in St. Paul’s Supreme Court brief that the purpose of the law was to show that the prohibited speech “is not condoned by the majority.” Id at 392-93 (citations omitted).

²⁶ For a similar point, see Akhil Reed Amar, *The Case of the Missing Amendments: R.A.V. v. City of St. Paul*, 106 Harv L Rev 124, 146-51 (1992).

²⁷ The line between regulation based on hostility and regulation based on harm may be exceedingly fine. I discuss this distinction further in Section II.C.

²⁸ See *R.A.V.*, 505 US at 435 (Stevens concurring).

²⁹ Compare *R.A.V.*, 505 US at 393 (“St. Paul has not singled out an especially offensive mode of expression—it has not, for example, selected for prohibition only those fighting words that communicate ideas in a threatening (as opposed to a merely obnoxious) manner.”), with id at 408 (White concurring) (“A prohibition on fighting words . . . is a ban on a class of speech that conveys an overriding message of personal injury and imminent violence, . . . a message that is at its ugliest when directed against groups that have long been targets of discrimination.”).

³⁰ *R.A.V.*, 505 US at 411, 407 (White concurring).

³¹ Id at 434, 416 (Stevens concurring).

The *R.A.V.* decision thus serves as a stark example of the importance of governmental motive in the Court's First Amendment analysis. Here, a debate about motive occurred in the open, revealing how a desire to punish impermissible purpose may explain and animate the Court's elaboration of doctrine. In the usual case, no such discussion occurs, but still the motive inquiry retains its power. The concern with filtering out illicit motive, though in these cases hidden, determines the content of the categories and rules that constitute First Amendment doctrine.

II. THE CONCEPT OF IMPERMISSIBLE MOTIVE

What exactly does it mean to say that an effort to filter out impermissible motive animates and explains First Amendment doctrine? In part, the question will have its answer only after I show how particular categories and rules of First Amendment law reflect a concern with governmental motive. But before taking on that task, I must discuss in a more general way the nature of the concern with motive and the kinds of legal tools suited to address it. In this Section, I first compare an approach to the First Amendment focused on motive with two approaches focused on effects. I next attempt to define and delimit what motives count as improper under the First Amendment. Here, I describe the concept of impermissible motive operative in the doctrine, while deferring to Section IV a discussion of why this concept might have become central. Finally, I examine methods a legal system can use to address the question of impermissible motive, given the difficulties of proof (and, some might say, the problems of coherence) such an inquiry raises. In much of this Section's discussion, the reader will hear echoes of *R.A.V.*, as the concerns and strategies of the Court in that case assume a more general shape and structure.

A. Three Perspectives on the First Amendment

In recent scholarship, a trend has developed to distinguish between two approaches to the First Amendment, which are sometimes complementary but often conflicting.³² According to

³² See Strauss, 1993 U Chi Legal F at 199-202 (cited in note 20) (contrasting an approach based on speakers' rights to a structural approach); Fiss, 100 Harv L Rev at 785-86 (cited in note 19) (contrasting an autonomy principle to a public debate principle); Robert Post, *Meiklejohn's Mistake: Individual Autonomy and the Reform of Public Discourse*, 64 U Colo L Rev 1109, 1132-33, 1136-37 (1993) (contrasting an autonomy theory to a collectivist theory).

this scholarly scheme, one conception of the First Amendment focuses on expanding the expressive opportunities open to speakers, whereas another focuses on improving the sphere of discourse encountered by the public "audience." To these two conceptions, which turn on different effects of speech regulation, I here juxtapose a third, which turns on the regulation's reasons.

The first approach—call it the "speaker-based" model—understands the primary value of the First Amendment to reside in its conferral of expressive opportunities on would-be communicators. A system of free expression, in allowing individuals to communicate their views, enhances their "autonomy" or "self-respect" or "self-development" or other (equally amorphous but desirable) human quality.³³ Under this theory, any limitation of expressive opportunities constitutes a harm because it interferes with some speaker's ability to communicate to others and with the benefit that speaker thereby derives. Moreover, the greater the limitation on speech, the greater the harm; under this theory, a broad restriction always poses greater constitutional concerns (because it interferes with more expressive activity) than a narrow one. Quantity, in other words, is of the essence; as one proponent of this model has stated, First Amendment doctrine should concern itself with how much a law "reduces the sum total of information or opinion disseminated."³⁴

By contrast, the second approach to the First Amendment—call it the "audience-based" model—focuses on the quality of the expressive arena. A system of free expression, under this theory, has value because it enables the public—the audience for the speech—to arrive at truth and make wise decisions, especially about matters of public import.³⁵ In order best to fulfill this function, a system of free expression should promote not speech alone, but speech of a certain kind and mixture. Rich public debate is the goal; the concern is the expressive realm as a whole, rather than each opportunity for expression. Under this theory, restrictions on speech pose more or less danger depending not on

³³ For versions of this approach, see C. Edwin Baker, *Human Liberty and Freedom of Speech* 47-69 (Oxford 1989); Martin H. Redish, *The Value of Free Speech*, 130 U Pa L Rev 591 (1982); David A.J. Richards, *Free Speech and Obscenity Law: Toward a Moral Theory of the First Amendment*, 123 U Pa L Rev 45, 59-70 (1974).

³⁴ Martin H. Redish, *The Content Distinction in First Amendment Analysis*, 34 Stan L Rev 113, 128 (1981).

³⁵ See Meiklejohn, *Political Freedom* at 19-28 (cited in note 15); Sunstein, *Democracy and the Problem of Free Speech* at 53-77 (cited in note 19); Fiss, 100 Harv L Rev at 787-94 (cited in note 19).

the sum total of speech prevented, but on the extent to which the restrictions distort or impoverish the realm of discourse. Further, some restrictions on speech are preferable to none, given that some enhance public discussion—for example, by preventing a few voices from drowning out others.³⁶ The purpose of the First Amendment—the goal against which regulation must be measured—is the protection of what Alexander Meiklejohn called the public’s “thinking process” from injury or “mutilation.”³⁷

The differences between these two approaches are captured in another of Meiklejohn’s sayings. “What is essential,” he wrote in support of the audience-based model, “is not that everyone shall speak, but that everything worth saying shall be said.”³⁸ Place to one side Meiklejohn’s view of the essential, and what remains in this aphorism is the core divergence between the models: one focuses on the effects of regulation on who speaks, the other on the effects of regulation on what is spoken. But this statement of the models’ disparity reveals also their likeness: both make critical an action’s consequences.

The third approach to the First Amendment—call it the “government-based” or “motive-based” model—claims that what is essential is not the consequences of a regulation but the reasons that underlie it.³⁹ The point of attention is neither the

³⁶ It is possible to contend that direct speech regulation never will serve the goal of rich public debate, or that it will so rarely serve that goal as to allow courts to assume that it never will do so. In some sense, the classic “marketplace of ideas” theory takes this position: the goal is to have a realm of discourse that leads to truth, but the means is a laissez-faire system. For this view to make sense, however, there must be a reason to think that the absence of regulation will lead to better results—here, a truth-producing market—than the allowance of regulation designed (or purportedly designed) to achieve this object. The most powerful such reason has to do with the government’s other motives for curtailing speech. (I reject in Section III.A an alternative reason, relating to the sheer incapacity of government to improve the speech market.) In this sense, any argument that advances the quality of debate as a goal, but assumes that an absence of regulation will best achieve this object, should be seen as an argument about governmental motive. The point here resembles the one made in considering *R.A.V.*: there is little reason to think a speech restriction necessarily will skew, rather than balance, public debate in the absence of a concern about governmental motive.

³⁷ Meiklejohn, *Political Freedom* at 27 (cited in note 15) (emphasis omitted).

³⁸ *Id.* at 26.

³⁹ Of the three approaches, the government-based approach is least represented in the literature. Frederick Schauer is the principal proponent of an approach of this kind; he has emphasized the danger of illicit governmental motive as part of a normative defense of providing heightened protection for expression. See Frederick Schauer, *Free speech: a philosophical enquiry* 80-86 (Cambridge 1982). Geoffrey Stone and Cass Sunstein both have included considerations of motive in broader analyses of First Amendment law. See Stone, 25 *Wm & Mary L Rev* at 227-33 (cited in note 15); Sunstein, *Democracy and the Problem of Free Speech* at 154-59 (cited in note 19). Finally, advocates of

speaker nor the audience, but the governmental actor standing in the way of the communicative process. Under this model, an action may violate the First Amendment because its basis is illegitimate, regardless of the effects of the action on either the sum of expressive opportunities or the condition of public discourse. Conversely, an action may comport with the First Amendment because legitimate reasons underlie it, again regardless of its range of consequences. The critical inquiry concerns what lies behind, rather than what proceeds from, an exercise of governmental power.⁴⁰

The divide between the model based on motive and the models based on effects can be overstated.⁴¹ One reason for First Amendment law to worry about governmental motive is itself consequential in nature; it refers to the predictable tendency of improperly motivated actions to have certain untoward effects.⁴² To say this is not to collapse the distinction I have offered. First, the government- or motive-based model may emerge as well from nonconsequential, deontological considerations, relating to the stance or attitude we expect the government to adopt in relation to its citizens. Second, the government-based approach—even if in an ultimate sense inspired by a concern for consequen-

so-called listener-autonomy theories of the First Amendment, such as David Strauss and (at one time) Thomas Scanlon, may be engaged in a form of motive analysis, in that they appear to contest the legitimacy not of speech regulations themselves, but of certain (autonomy-infringing) justifications for them. See David A. Strauss, *Persuasion, Autonomy, and Freedom of Expression*, 91 Colum L Rev 334, 353-60 (1991); Thomas Scanlon, *A Theory of Freedom of Expression*, 1 Phil & Pub Aff 204, 209 (1972). The authors of such theories, however, do not speak in terms of motive and might well contest my characterization. See Strauss, 1993 U Chi Legal F at 201 (cited in note 20) (asserting that listener autonomy theories fall under what I have called an "audience-based" approach).

⁴⁰ As should be obvious by now, I make no distinction between such terms as "purpose," "intent," "motive," "basis," and "reason." The Court has used these terms interchangeably, both in First Amendment jurisprudence and elsewhere; in *O'Brien*, for example, the Court treated the terms "motive" and "purpose" as synonymous. See 391 US at 383. See also David A. Strauss, *Discriminatory Intent and the Taming of Brown*, 56 U Chi L Rev 935, 951 (1989) (noting the interchangeable use of these terms in equal protection law). Moreover, attempts by scholars to distinguish among these terms have proved unhelpful. See John Hart Ely, *Legislative and Administrative Motivation in Constitutional Law*, 79 Yale L J 1205, 1217-21 (1970) (criticizing such efforts).

⁴¹ So too can the divide between the two models based on effects. For example, protection of audience-based interests demands that the government accord some substantial rights to individual speakers. Conversely, protection of speaker-based interests may demand some attention to the condition of public discourse, to prevent opportunities for expression from becoming purely formal and ineffective. A set of complicated relations exists among all these models, as well as among the concerns that underlie them.

⁴² For further discussion of the points made in this paragraph, see text accompanying notes 257-78.

ces—very often will lead to different doctrinal rules, producing different results, than an approach that focuses on effects, whether on the speaker or the audience. Still, this analytic relationship between a motive-based approach and effects-based approaches should not be disregarded; I do not want to suggest that these approaches exist hermetically sealed from each other.

Further, the motive-based model and the effects-based models can operate in confluence with each other, except in their starkest forms—each contributing something to First Amendment doctrine. A body of law predominantly concerned with effects (whether on the speaker or audience) can make some place for considerations of governmental purpose; so too, but conversely, for a body of law predominantly concerned with motive. I have no doubt that current doctrine responds, in some manner and at some times, to all the concerns I have mentioned. The government-based approach does not wholly exclude the others.

The delineation of the three approaches, however, remains important. The approaches often will point in divergent directions, prescribing both different rules of law and different outcomes. And the pattern of decisions where such conflicts take place says much about the concerns that drive the law of free speech. To prepare the way for showing that among the potential concerns, illicit motive takes pride of place, I turn now to the meaning of illicit motive in First Amendment analysis.

B. Defining Impermissible Motive

Assuming for now that First Amendment law constitutes an attempt to flush out impermissible motives, what motives count as impermissible? The Court has not fully addressed, much less resolved, this question. Despite the proscription in *O'Brien*, the Court sometimes has probed the government's reasons for restricting expression;⁴³ too, the Court has articulated several statements of First Amendment principle that sound in terms of

⁴³ The Court most recently inquired into legislative motive in *R.A.V. and Turner Broadcasting System, Inc. v FCC*, 114 S Ct 2445, 2461-62 (1994). A much earlier example of such an examination appears in *Grosjean v American Press Co.*, 297 US 233, 250 (1936). In cases involving executive action, the Court routinely speaks in terms of motive. For example, in addressing a First Amendment challenge brought by a discharged employee of the government, the Court will ask whether the government fired the employee because it disapproved of her expression. See, for example, *Connick v Myers*, 461 US 138, 143-46 (1983). The Court apparently sees the examination of motive in such cases as different in kind from—and less problematic than—the examination of the motives underlying legislation.

motivation.⁴⁴ But the Court, as the edict in *O'Brien* shows, usually has hesitated to discuss the issue of illicit motive in any detail or with any directness. The effort to define the concept of illicit purpose operative in First Amendment law thus must depend on a fair amount of extrapolation, as well as on a “reading backwards” from the doctrines discussed later in this Article.

The definition of illicit motive that this effort reveals is in certain respects imprecise and conceptually puzzling. It is not necessary, given my purpose, to untangle all the complexities this definition raises; what matters for this Article is that the doctrine emerges from an understanding of illicit motive, however inexact or enigmatic. This Section, then, provides only a sketch of the definitional issues. I first lay out the concept of impermissible motive evident in the law, in part describing it in terms of a neutrality principle. I then consider both a limitation on and an objection to the definition I have offered, and I finally compare that definition with some alternatives.

1. A definition.

Consider the following snapshot of impermissible motives for speech restrictions. First, the government may not restrict expressive activities because it disagrees with or disapproves of the ideas espoused by the speaker; it may not act on the basis of a view of what is a true (or false) belief or a right (or wrong) opinion.⁴⁵ Or, to say this in a slightly different way, the government cannot count as a harm, which it has a legitimate interest in preventing, that ideas it considers faulty or abhorrent enter the public dialogue and challenge the official understanding of acceptability or correctness. Second, though relatedly, the government may not restrict speech because the ideas espoused threaten officials’ own self-interest—more particularly, their tenure in office.⁴⁶ The government, to use the same construction as above, cannot count as a harm, which it has a legitimate interest in preventing, that speech may promote the removal of

⁴⁴ See notes 45, 48-49.

⁴⁵ See *City Council v Taxpayers for Vincent*, 466 US 789, 804 (1984) (asking whether a law “was designed to suppress certain ideas that the City finds distasteful”); *Consolidated Edison Co. v Public Service Commission*, 447 US 530, 536 (1980), quoting *Niemotko v Maryland*, 340 US 268, 282 (1951) (Frankfurter concurring) (asking whether speech was barred “merely because public officials disapprove the speaker’s views”); Stone, 25 Wm & Mary L Rev at 227-28 (cited in note 15).

⁴⁶ See Stone, 25 Wm & Mary L Rev at 228 (cited in note 15); Sunstein, *Democracy and the Problem of Free Speech* at 155 (cited in note 19).

incumbent officeholders through the political process. Third, and as a corollary to these proscriptions, the government may not privilege either ideas it favors or ideas advancing its self-interest—for example, by exempting certain ideas from a general prohibition.⁴⁷ Justice Scalia summarized these tenets in *R.A.V.*: “The government may not regulate [speech] based on hostility—or favoritism—towards the underlying message expressed.”⁴⁸

To this statement of illicit motive, one further gloss must be added: the government may not limit speech because other citizens deem the ideas offered to be wrong or offensive⁴⁹—or for that matter, because they see the ideas as threatening to incumbent officials. This ban echoes those just stated, except for the identity of the party (above the government, now the public) that disapproves the ideas; the theory is that this substitution of party name should make no constitutional difference. Some of course may argue that restrictions based on public dislike boast a democratic legitimacy separating them from restrictions based on governmental hostility. But this distinction falters on the difficulty of disentangling the actions of officials from the desires of constituents. When the government acts, its reasons for doing so usually reflect the views of some part of the public. Distinguishing between public and governmental hostility thus seems hopeless as a practical matter. Further, the distinction shatters on the Court’s longstanding view that the First Amendment protects

⁴⁷ See John Hart Ely, *Flag Desecration: A Case Study in the Roles of Categorization and Balancing in First Amendment Analysis*, 88 Harv L Rev 1482, 1506-07 (1975); Stone, 25 Wm & Mary L Rev at 227-28 (cited in note 15).

⁴⁸ 505 US at 386. See also *Young v American Mini Theatres, Inc.*, 427 US 50, 67 (1976) (“[R]egulation of communication may not be affected by sympathy or hostility for the point of view being expressed by the communicator.”).

⁴⁹ See *Texas v Johnson*, 491 US 397, 414 (1989) (“If there is a bedrock principle underlying the First Amendment, it is that the Government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable.”); *Hustler Magazine, Inc. v Falwell*, 485 US 46, 55 (1988), quoting *FCC v Pacifica Foundation*, 438 US 726, 745-46 (1978) (“The fact that society may find speech offensive is not a sufficient reason for suppressing it. Indeed, if it is the speaker’s opinion that gives offense, that consequence is a reason for according it constitutional protection.”). See also Stone, 25 Wm & Mary L Rev at 214-16 (cited in note 15); Sunstein, *Democracy and the Problem of Free Speech* at 155-56 (cited in note 19). The notion of offense is, of course, a tricky one. There is a fine line between offense at the content of ideas, to which I refer, and offense at the means by which those ideas are expressed. There is also a fine line between mere offense and emotional injury, in that a certain kind and degree of the former (the “offense” felt, for example, by the concentration camp survivors in Skokie) may constitute what society recognizes, or would wish to recognize, as the latter. Finally there is a complex relationship between offense at ideas (or any other sort of hostility toward ideas) and the entire range of harms those ideas cause. See text accompanying notes 60-66.

no less against majority oppression than against runaway government.⁵⁰ In keeping with this general view, the Court's conception of illicit motive must apply not only to officials but also to the public acting through them. Just as in equal protection law the government may not discriminate among persons on the basis of majoritarian biases,⁵¹ so too in First Amendment law the government may not so distinguish among messages. The key principle with respect to motive is that the government may not limit speech on grounds of mere disapproval, no matter whose or how widely shared.

This principle leaves untouched many reasons to restrict expression; in this Article, I call these reasons "harm-based" and contrast them to the "ideological" reasons I have just discussed. The distinction raises difficult issues, which I address below, but to understand first its essential nature, a return to *R.A.V.* may prove helpful. Consider some different explanations for the St. Paul ordinance. First, the city may have enacted the statute to express its own or its citizenry's hatred of the ideas of racism, sexism, and so forth.⁵² Alternatively, the city could have enacted the statute to prevent harms that it thought the covered speech posed to the community. Perhaps the city feared that the speech would cause some persons to suffer psychic trauma or other emotional harm; or that the speech would spark bloody public riots, because of strong popular resistance to it;⁵³ or that the speech

⁵⁰ See cases cited in note 49. See also *Kingsley Pictures Corp v Regents*, 360 US 684, 689 (1959) (stating that the First Amendment's "guarantee is not confined to the expression of ideas that are conventional or shared by a majority"). This view has its ancestry in the Framers' fear of majority factions. As James Madison wrote: "In our Governments the real power lies in the majority of the Community, and the invasion of private rights is chiefly to be apprehended, not from acts of Government contrary to the sense of its constituents, but from acts in which the Government is the mere instrument of the major number of the constituents." Letter from James Madison to Thomas Jefferson (Oct 17, 1788), in Robert A. Rutland, et al, eds, 11 *The Papers of James Madison* 298 (Virginia 1977). But see Akhil Reed Amar, *The Bill of Rights as a Constitution*, 100 Yale L J 1131, 1147-52 (1991) (suggesting that the First Amendment protects against government self-dealing, rather than majority tyranny). Amar's narrow understanding of the Amendment suggests, among other things, that no decision made by a properly selected jury—in, for example, a defamation case—would violate free speech principles.

⁵¹ See, for example, *City of Cleburne v Cleburne Living Center, Inc.*, 473 US 432, 448 (1985); *Palmore v Sidoti*, 466 US 429, 433 (1984).

⁵² Justice Scalia believed the city intended the law to serve just this function and accordingly savaged the city's motives. Asserting that the city enacted the ordinance to convey the majority's disapproval of an idea, Scalia wrote that "[t]he point of the First Amendment is that majority preferences must be expressed in some fashion other than silencing speech" *R.A.V.*, 505 US at 392.

⁵³ This reason for restricting speech in one sense depends on popular hostility to ideas, which I have deemed an illegitimate reason for speech regulation: were it not for

would persuade listeners to engage in acts of race-based violence. The concept of impermissible motive I have described applies to the first of these explanations, but to none of the others. The first violates the principle that the government may not restrict speech on the basis of its own or the majority's view of what ideas are right or wrong, praiseworthy or shameful. The others do not violate this principle because they relate not to the message as message, but to the consequence of its expression; they stem not from ideological hostility, but from a perception of material harm.⁵⁴ In short and critically, they relate to harms that the government has a legitimate interest in preventing and obviously could act to prevent if not caused by expression.

Whenever hostility toward ideas as such (or the other impermissible factors of sympathy or self-interest) has played some part in effecting a restriction on speech, the restriction is irretrievably tainted; what has entered into the action commands its invalidation.⁵⁵ In contrast, when such factors have played no role—when the government has restricted ideas only as and when they bear harmful consequences—the government's purposes support sustaining the action. The critical inquiry is whether the government would have imposed the restriction in the absence of impermissible factors, solely on the basis of a neutral and legitimate evaluation of harm. Or to put the question in another way, it is whether the government would have treated (or did treat) identically ideas with which it disagreed, ideas with which it agreed, and ideas to which it was indifferent, to the extent those ideas caused the same harms.⁵⁶ This inquiry tests

popular hostility, the government would have no fear of riot, and thus would have no reason to restrict the expression. Nonetheless, there is a difference between restricting speech because of public hostility alone and restricting speech because this hostility will lead to bloodshed. In the former case, the government acts only to advance the majority's version of truth; in the latter, the government acts to avert violence, even if it is the majority's desire to impose its will that makes this action necessary.

⁵⁴ The Court recently drew this kind of distinction in explicit terms, asking whether "the Government seeks not to advance a legitimate regulatory goal, but to suppress unpopular ideas or information or manipulate the public debate." *Turner Broadcasting System, Inc. v FCC*, 114 S Ct 2445, 2458 (1994).

⁵⁵ For a similar understanding of the consequence of finding an impermissible motive, see Stone, 25 Wm & Mary L Rev at 229-30 (cited in note 15). Note that it should make no difference whether the impermissible motive has played a role on a conscious or unconscious level. If, as I argue, the injury is differential treatment based on prohibited considerations, the injury is not affected by the level of consciousness at which the considerations operated. See Strauss, 56 U Chi L Rev at 960 (cited in note 40).

⁵⁶ This test resembles the test proposed by many commentators to determine discriminatory intent in the equal protection context. See, for example, Paul Brest, *The Supreme Court, 1975 Term—Foreword: In Defense of the Antidiscrimination Principle*, 90 Harv L

whether the government regulated, even in part, on the basis of ideas as ideas, rather than on the basis of material harms.

This test is a measure of the neutrality or impartiality that the First Amendment often is said to command.⁵⁷ The First Amendment allows distinctions among speech on many bases. What it does not allow is classifications built on hostility or sympathy to ideas. The neutrality principle thus mirrors the doctrine of impermissible purpose. The government may classify speech to achieve legitimate governmental objects, such as the prevention of illegality or violence. But the government may not rest a classification, even in part, on the ground that some messages are worthier than others. Differences of this kind with respect to ideas must count as legally irrelevant. To say that “[t]here is an ‘equality of status in the field of ideas,’”⁵⁸ is to say that the government cannot regulate speech for such impermissible reasons.

2. A limitation, an objection, and a comparison.

One question about the principle just articulated relates to its scope of operation: does the principle apply only when the government acts in its traditional role as regulator of private speech, or does it also apply (in either pure or diluted form) when the government performs the increasingly important functions of speaker, employer, and educator? In this Article, I discuss the issue of governmental motive only in relation to restrictions on private speech; except for a few words, I leave for another time the question how the understanding of improper motive I have described translates (or does not) into contexts in which the government itself performs speech functions. My thumbnail view is that the principle has greater relevance in these contexts than

Rev 1, 6-8 (1976); Ely, 79 Yale L J at 1266-68 (cited in note 40). David Strauss calls this the “reversing the groups” test because it asks “if the government would have made the same decision even if the races of those affected had been reversed.” Strauss, 56 U Chi L Rev at 957 (cited in note 40). Similarly, the test in the First Amendment context might be called the “reversing the ideas” test, because it asks whether the government would have made the same decision if different ideas were affected. Strauss criticizes the test in equal protection law on the ground that it forces courts to make speculative, counterfactual determinations. To the extent the test asks courts to consider the question of “reversal” directly, this criticism is valid; my argument, offered in Section III, is that First Amendment doctrine relieves courts of this impossible task by providing rules that capture the gist of this inquiry in a concrete and easily administrable fashion.

⁵⁷ See, for example, *Carey v Brown*, 447 US 455, 462-63 (1980); *Police Department of Chicago v Mosley*, 408 US 92, 96 (1972).

⁵⁸ *Mosley*, 408 US at 96, quoting Meiklejohn, *Political Freedom* at 27 (cited in note 15).

might be thought, though less than when the government restricts private speech,⁵⁹ I also believe that the concept of illicit purpose should apply in these contexts even more strongly than it does, thus narrowing (though not eliminating) the importance for First Amendment analysis of the particular role the government is playing. But because I cannot defend these views in this Article, the key point here is one of limitation: the concept of impermissible motive I have described refers to the government in its capacity as regulator of private expression.

A second and, for my purposes, more important question concerns the coherence of the distinction I have drawn between motives based on harm and motives based on ideology—a distinction that might be viewed as possessing rhetorical appeal, but collapsing on deep reflection.⁶⁰ What is it, after all, to hate a message if not, and other than, to think the message causes injury? Perhaps opposition to speech on what I have termed “ideological” grounds—sheer hostility toward a message—does not exist as a real-world phenomenon. Perhaps such opposition always stems from, and thus reduces to, a conviction that the idea causes harms that the government has a legitimate interest in preventing.⁶¹ If this is so, the distinction I have drawn might be said to rest only on the level of generality chosen to frame the critical question. Query 1: Why did officials restrict the expression? Answer 1: Because they disliked its message. Query 2: Why did the officials dislike its message? Answer 2: Because they believed the message caused material harm of a serious nature. If the distinction I have drawn depends on failing to ask the second question, then it seems a foundation too weak to support First Amendment doctrine.

This challenge is strong and the issue complex, but some examples indicate that the two kinds of motives, though closely interwoven, retain distinct characters. Assume that racist speech—or, to see the point from another perspective, assume that flag-burning—poses dangers: such speech may spark a riot,

⁵⁹ The law in this area is largely a mess, resisting any coherent understanding. If motive-based theory does not wholly explain the doctrine, neither does any other. See Kagan, 1992 S Ct Rev at 40-45 (cited in note 14).

⁶⁰ I assume that the distinction between motives based on harm and motives based on self-interest is not so mysterious. It seems clear that self-interest can counsel a speech restriction that an evaluation of harm (at least of harm that can be counted as harm) would not.

⁶¹ See Elena Kagan, *Regulation of Hate Speech and Pornography After R.A.V.*, 60 U Chi L Rev 873, 880 (1993).

induce a violation of law, or cause emotional injury. Not everyone will measure or respond to these potential harms in the same manner. Persons will differ both in assessing the magnitude (indeed, the existence) of danger and in deciding what amount of danger will justify a restriction. And these divergent judgments about the harm the speech causes and the need to limit it rest in part on what I have said cannot count in the equation: the desire of persons, conscious or not, to suppress ideas that challenge (just because they challenge) and to privilege ideas that ratify (just because they ratify) their own belief systems.

So too we might explain other instances, past and present, of deciding when neutral interests counsel a restriction on speech. Consider the core cases of our free speech tradition, involving the questions whether speech opposing World War I or supporting communism threatened resistance to law or overthrow of the government.⁶² Or consider the string of cases in the 1960s raising the issue whether civil rights protests would cause public riots.⁶³ Or consider this past year's debate as to whether right-wing talk radio provokes crimes of violence.⁶⁴

As examples of this kind suggest, hostility toward speech (or its opposite) may affect the decision to regulate speech, separate from and independent of neutrally conceived harms.⁶⁵ Such hos-

⁶² See, for example, *Schenck v United States*, 249 US 47 (1919) (World War I); *Dennis v United States*, 341 US 494 (1951) (communism).

⁶³ The Court held unconstitutional in these cases the actions of Southern law enforcement officers in dispersing (shall we say "prematurely") civil rights demonstrations on the ground that they would provoke a riot or other hostile audience response. See *Edwards v South Carolina*, 372 US 229, 237-38 (1963); *Cox v Louisiana*, 379 US 536, 550-51 (1965). Professors Sunstein and Fiss have interpreted these cases to require affirmative police protection of any speaker whose words arouse a threatening response. See Cass R. Sunstein, *Free Speech Now*, 59 U Chi L Rev 255, 273-74 (1992); Fiss, 100 Harv L Rev at 786 (cited in note 19). I think the decisions have a narrower meaning, consistent with the theory I have proposed. The decisions established not a duty to provide police protection for all speakers, but rather a duty to provide as much police protection for speakers whose ideas officials hate as for speakers whose ideas the officials approve.

⁶⁴ Ronald Dworkin has suggested another example to make a similar point. He asks why the feminist movement has focused so much attention on pornography when (by his estimation) "popular forms of [mass] culture—the view of women presented in soap operas and commercials, for example—are much greater obstacles to [] equality than the dirty films watched by a small minority." Ronald Dworkin, *Women and Pornography*, NY Rev Books 36 (Oct 21, 1993). He concludes that pornography, though less harmful than these other forms of culture, is more detestable—that the rawness with which it expresses the idea of sexual subordination causes it to be "deeply offensive in itself, whether or not it causes any other injustice or harm." *Id.* Dworkin, of course, may be wrong about the relative harms caused by these two forms of speech. But his example suggests the potential for purely "ideological" motives to influence regulatory proposals and the estimations of harm on which they are built.

⁶⁵ See also Schauer, *Free speech* at 82 (cited in note 39) (noting "in people a desire for

tility no doubt may derive from the fact of harm and have no significance of its own. But so too hostility toward ideas may exist apart and freestanding, or even impel the judgment of harm. Most often, perhaps, the two kinds of motives become hopelessly entangled, as one influences the other which in turn influences the first in a kind of endless feedback loop. But the complexity of this relationship—the way the different motives interact with each other, on both a conscious and an unconscious level—should not obscure the role that ideological factors may play. Hostility against speech (or sympathy toward it) may lead the government or public to overassess (or underassess) the harm speech causes. Likewise, hostility against speech (or sympathy toward it) may lead the government or public to tolerate a lesser (or greater) degree of the harm than it otherwise would. In either case, hostility (or sympathy) is doing some of the work in the decision to impose a limit on speech. The desire to suppress for its own sake—the tendency to count challenge or opposition itself as harm—is impermissibly entering into the calculation.⁶⁶

The distinction between harm-based and ideological motives I have offered here differs from a distinction several other commentators have drawn relating to governmental purpose. They have argued that the great divide is between laws based on the communicative effect of speech and those based on other effects—or more narrowly, between laws based on the “persuasive” effect of speech and those based on other effects. When phrased in terms of communicative effect, the argument runs as follows. The government may not restrict speech for any reason having to do with either the messages embedded in the speech or the consequences of those messages; the government may impose restrictions for reasons relating only to aspects of the speech independent of and extraneous to the message, such as the speech’s decibel level.⁶⁷ When phrased, alternatively, in terms of “persuasive”

unanimity, an urge to suppress that with which they may disagree even if there seems no harm to that expression”).

⁶⁶ The complex relationship between harm and ideology has a familiar analogue in equal protection law. Discrimination on the basis of race, gender, and so forth often has a basis in reason—in accurate generalizations about the different characteristics, behaviors, and needs of members of particular groups. But such discrimination also often has a basis in fear, loathing, and prejudice. Hatred of this kind in part may emerge from actual difference, in part may exist as something independent, in part may construct and influence the perception of difference. The entanglement of hostility and harm-based reasons in First Amendment law is in many ways similar.

⁶⁷ The concept of communicative effects has received its fullest explication in the work of John Hart Ely and Laurence Tribe. See John Hart Ely, *Democracy and Distrust*:

effect, the argument has a narrower cast. It now posits that the government may not restrict speech for any reason having to do with either the message itself or the ability of the message to persuade listeners to take some action.⁶⁸ So whereas my conception of motive countenances (to use but a few examples) reasons relating to the capacity of speech to cause psychic trauma, trigger a hostile audience response, or persuade an audience to violate a law, the communicative effects theory views all of these reasons as impermissible, and the persuasive effects theory rules out the final reason, though not the two others.

These alternative theories are deficient in two respects. First, they conflate motives that I have just argued differ from each other, albeit in a complex, shifting, and elusive manner. Second, and more important for my purposes, they fail as descriptive theories of what constitutes illicit motive in First Amendment law. Courts in fact allow the government to restrict speech for reasons concededly related to its communicative (including persuasive) effects. True, the government usually must meet a heightened standard when it justifies a law on these grounds.⁶⁹ But if the motives identified by these theories were impermissible, in the way I use the term, a court would have to invalidate in all circumstances restrictions concededly based on them. A reason that is impermissible cannot count as a reason because it refers to a thing that cannot count as a harm. Reasons related to communicative or persuasive effects are not of this kind: the

A Theory of Judicial Review 110-11 (Harvard 1980) (asking whether "the evil the state is seeking to avert . . . arises from something other than a fear of how people will react to what the speaker is saying"); Laurence Tribe, *American Constitutional Law* § 12-2 at 789-90 (Foundation 2d ed 1988) ("[I]f the constitutional guarantee is not to be trivialized, it must mean that government cannot justify restrictions on free expression by reference to the adverse consequences of allowing certain ideas or information to enter the realm of discussion and awareness.").

⁶⁸ See Scanlon, 1 Phil & Pub Aff at 212-13 (cited in note 39); Strauss, 91 Colum L Rev at 334 (cited in note 39) ("[T]he government may not justify a measure restricting speech by invoking harmful consequences that are caused by the persuasiveness of the speech."). A still narrower version of this theory might posit that the government is forbidden from restricting speech on the ground that it will persuade people to adopt wrong or false opinions (rather than persuade people to take actions causing harm). If phrased in this way, the principle becomes compatible (indeed, is largely synonymous) with my description of impermissible motive.

⁶⁹ The heightened standard actually arises from the content-based terms of a law rather than from its underlying justification. Of course, the distinction between content-based and content-neutral laws may serve as an easily administrable device to test for impermissible motive. See Section III.A. But as I explain in the text, the structure of First Amendment law belies the view that all reasons relating to the communicative or persuasive effects of speech are impermissible.

government may count such effects—for example, violations of law arising from advocacy—as cognizable harms and may move to prevent them upon showing a real need to do so.⁷⁰ By contrast, what I have termed “ideological” reasons are indeed off-limits. The government may not count a challenge to governmental officials or official opinion as a harm and may not restrict speech to defeat such a challenge, even if the restriction is essential for achieving this purpose. The line between licit and illicit reasons thus lies not where these alternative theories have placed it, but between harm-based and ideological motivations.⁷¹

But if I am correct that the central prohibition of the First Amendment relates to ideological motive, then the practical import of the Amendment would seem nonexistent. Even assuming there is such a thing as a governmental motive in the sense I have used the term, how would a court ever discover the motive that I have said is off-limits? Officials will not admit (often, will not themselves know) that a regulation of speech stems from hostility or self-interest. They will invoke in each case a plausible interest, divorced from ideological disapproval, to restrict the affected expression. Then, perhaps, even a speech-protective court will have to approve the government’s action. Or will it? The next Section focuses on this question. It first discusses briefly the difficulty of making a direct inquiry into governmental motive, as well as the very coherence of this project. It then addresses, more fully, the possibility of ascertaining motive through indirect means, by using a set of rules directed to the face of legislation that will demarcate very roughly the set of governmental actions most likely to have arisen from illicit motive.

⁷⁰ The prevailing standard, emerging from *Brandenburg v Ohio*, allows the government “to forbid or proscribe advocacy of the use of force or of law violation . . . where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action.” 395 US 444, 447 (1969). In cases of standard-fare criminal solicitation, with no political or “public” character, most scholars assume a lesser standard would apply. See, for example, Kent Greenawalt, *Speech, Crime, and the Uses of Language* 110-26 (Oxford 1989). Such cases thus pose even starker counterevidence for any theory of impermissible motive based on communicative or persuasive effect. So too do cases dealing with speech proposing an illegal commercial transaction, which the government also may regulate freely under the First Amendment. See *id* at 270-71.

⁷¹ I do not claim that the notion of communicative effect has no operative meaning in the law. As I discuss in Section III, First Amendment doctrine sometimes uses this notion to aid in the search for the true impermissible—that is ideological—motive.

C. Surmounting Problems of Proof

It has become a commonplace among both judges and scholars that the search for legislative intent—indeed, the very notion of legislative intent—raises grave problems. One set of questions relates to whether there is “a” legislative intent to be found. Consider that each legislator possesses a complex mix of hopes, expectations, beliefs, and attitudes. It is not obvious which of these mental states, or combination of them, constitutes her essential intent in voting for legislation.⁷² Now consider that a legislature has many, perhaps hundreds of members. It is, if anything, less obvious how to combine different individual intents (assuming those exist) into a composite group purpose.⁷³

A second set of questions assumes there is such a thing as legislative intent, but asks whether we can find it (and, if so, how). This is the aspect of the problem on which the *O'Brien* Court focused when it declared that “inquiries into congressional motives . . . are a hazardous matter.”⁷⁴ Often, the Court recognized, evidence of legislative purpose will consist “of what fewer than a handful of Congressmen said about it.”⁷⁵ But “[w]hat motivates one legislator to make a speech about a statute is not necessarily what motivates scores of others to enact it . . .”⁷⁶ Indeed, what motivates one legislator to make a speech about a statute is not necessarily what motivates that very legislator to enact it.⁷⁷ The evidentiary materials available—floor statements,

⁷² See Ronald Dworkin, *Law's Empire* 321-33 (Harvard 1986); *Edwards v Aguillard*, 482 US 578, 637 (1987) (Scalia dissenting).

⁷³ See Dworkin, *Law's Empire* at 320-21 (cited in note 72). The branch of public choice theory growing out of Arrow's impossibility theorem presents an especially strong challenge to the notion that a collective body has an intent. See Kenneth A. Shepsle, *Congress Is a "They," Not an "It": Legislative Intent as Oxymoron*, 12 Intl Rev L & Econ 239, 249-50 (1992); Frank H. Easterbrook, *Statutes' Domains*, 50 U Chi L Rev 533, 537-39 (1983). For a defense of the notion of legislative intent, see Andrei Marmor, *Interpretation and Legal Theory* 159-72 (Clarendon 1992).

⁷⁴ 391 US at 383.

⁷⁵ *Id.* at 384.

⁷⁶ *Id.*

⁷⁷ The Court expressed this point in a slightly different way: it worried that legislation “could be reenacted in its exact form if the same or another legislator made a ‘wiser’ speech about it.” *Id.* Congress, that is, could respond to the judicial invalidation of a statute on grounds of improper purpose by passing the identical statute with a cleaned-up legislative record. This argument, often termed the “futility” concern, usually is treated as an independent reason—distinct from the difficulty of ascertaining legislative motive—to disdain an inquiry into purpose. See Ely, 79 Yale L J at 1214-15 (cited in note 40) (accepting the argument); Paul Brest, *Palmer v. Thompson: An Approach to the Problem of Unconstitutional Legislative Motive*, 1971 S Ct Rev 95, 125-27 (rejecting the argument). The futility concern, however, rests entirely on the problem of ascertainability. If motive

committee reports, and so forth—provide a less than reliable guide to the intent of any individual legislator, let alone to the intent of the collective body.

The conception of impermissible motive I use in this Article does not fall prey to questions regarding the coherence of the notion of collective intent. The issue of motive, as I have framed it, is one of but-for causation: would the restriction on speech have passed—that is, would the outcome of the legislative process have differed—in the absence of ideological considerations?⁷⁸ To answer this question, it is unnecessary to consider the essential intent of any individual, much less of the decision-making body; it is irrelevant whether any such intent exists or can exist as a conceptual matter. The “thing” that a court is attempting to find is only the intrusion of a particular factor in a way that affects the decision-making process. Whatever questions attach to the notion of collective intent do not place in doubt these but-for causes of governmental action.⁷⁹

But this conception of impermissible motive cannot avoid questions relating to the difficulty of finding the relevant object. True, the Court need not determine the collective sense of a decision-making body or even a single legislator’s full state of mind. But the Court must perform a task that might be as hard: determining whether a particular factor played a but-for role in a

could be reliably determined, the Court would not fear futility, for it then could invalidate the reenacted, no less than the original, statute (assuming the motive remained the same). The problem of futility arises only because legislators, at any time, can feign a purpose they do not have.

⁷⁸ Compare Laurence H. Tribe, *The Mystery of Motive, Private and Public: Some Notes Inspired by the Problems of Hate Crime and Animal Sacrifice*, 1993 S Ct Rev 1, 33 n 79 (“[T]he motive inquiry in the *O’Brien* context, for example, need go no deeper than to ask whether the law would have been enacted but for the fact that draft-card burning was being used for protest.”); Brest, 1971 S Ct Rev at 119 (cited in note 77) (“[I]t is inappropriate to ask which of several possible objectives was ‘sole’ or ‘dominant’ in the decisionmaker’s mind: an illicit motive may have been ‘subordinate’ and yet have determined the outcome of the decision.”).

⁷⁹ Moreover, conceptual doubts about legislative intent are irrelevant to my project, which is one of understanding the root sources of current doctrine. The Court has not allowed such doubts to prevent it from inquiring into motive—even into “sole,” “dominant,” or “essential” motive—in a variety of circumstances. For example, in determining the constitutionality of legislative action under the Establishment Clause and Commerce Clause, the Court specifically asks about legislative purpose. See *Lemon v Kurtzman*, 403 US 602, 612 (1971); *Pike v Bruce Church, Inc.*, 397 US 137, 142 (1970). And as the *O’Brien* Court itself admitted, courts routinely explore legislative motive in interpreting statutes. 391 US at 383. Even if there is no such thing as legislative intent, the Court often acts as if there is. So long as this is true, objections to the concept of legislative intent do no damage to the claim that some aspect of doctrine, explicitly or implicitly, attempts to discover the intent of the legislature.

decision-making process. This task, in its most simplified form, involves reckoning how many legislators the impermissible consideration swayed and comparing that number to the margin of victory.⁸⁰ What *O'Brien* said about the hazards of inquiring into motive seems to apply in full to this inquiry. The standard evidence of legislative process provides an insufficient basis to make the requisite head count or even estimate its outcome.

Consider first, to highlight the difficulty of the endeavor, how a court would decide whether improper motive tainted the decision of even one legislator. Few legislators, of course, will admit to a constitutionally illegitimate purpose; the legislator instead will point to some real harm that the speech causes. Such a harm-based pretext usually will be available; cases in which the government tries to curtail speech that cannot plausibly be described as harmful are not common. Further, the legislator herself often will not know whether an illicit reason tainted her consideration of the law at issue, given the complex dynamic between legitimate assessments of harm and illegitimate attitudes toward opinions. To make matters worse, the judge handling the matter will possess her own views of the ideas restricted, which may affect her evaluation of the legislator's motives in the same diffuse and incalculable ways as the legislator's views initially affected her decision. Now consider how these difficulties multiply when a court must face the issues of aggregation involved in determining how illicit motive affected a multimember body. Hence the message of *O'Brien*: direct inquiry into motives for restricting speech very rarely will prove productive.

But the impracticality of this inquiry need not force courts to abandon the goal of invalidating improperly motivated legislation, if they can find another, more feasible way of pursuing that project. If courts cannot determine motive directly, by exploring

⁸⁰ There are many complications in determining whether ideological factors altered a legislative decision that this simplified statement of the problem ignores. Most notably, this statement overlooks the disproportionate influence that some legislators wield because of, among other things, their agenda-setting ability or their strategic importance on other issues. What if, for example, only one legislator harbored impermissible reasons for favoring a statute (all other legislators having legitimate, harm-based reasons), but she was the person responsible for bringing the statute to a vote? Or what if other legislators acceded to her wishes on the statute (having no strong views of their own) to get her vote on another issue? The critical issue is whether impermissible reasons altered the outcome of the decision-making process; in both these cases, they did. But in both cases, it will not suffice to ask whether improper motives directly accounted for the votes necessary to enact the statute. The story about how improper motive affected the outcome is more complicated, raising even greater problems of discovery.

what went into the legislative process, perhaps they can determine motive obliquely, by looking at what came out of it. Suppose that courts could develop rules relating only to the terms of legislation; suppose further that these rules predictably operated to sort out actions that had impermissible motives from those that did not. By using these rules, courts could invalidate laws supported by improper reasons without ever confronting the problems of proof generated by a direct inquiry into motive. The function of the rules in flushing out impermissibly motivated actions might not be articulated or even understood. The rules would operate in an autonomous manner, removed from explicit consideration or discussion of the question of motive. But the two would remain integrally connected: the concern with motive would determine the scope of the rules, and the rules would give effect to the concern with motive.

So it might be in First Amendment law: perhaps the Supreme Court has constructed a set of rules that allows a judge to ferret out impermissible motives at the same time as it obviates any need to ask about this issue. We might think of these rules as proxies for a direct inquiry into motive or as rules of an evidentiary nature. These rules use objective criteria, focusing on what a law includes and excludes, on what classifications it uses, on how it is written. But in making such inquiries, the rules in fact serve as an arbiter of motive. Through use of these objective tests and rules, some rough sorting out takes place: between laws tainted by ideological motives and those not so blemished.

The roughness of this division should not be understated: these tests of governmental purpose necessarily will be imperfect—simultaneously under- and overinclusive. Still, it makes sense to use the rules, rather than to ask directly about motive, as it often makes sense to use rules rather than to rely on their underlying reasons. If courts cannot reliably (or cannot at all) determine whether the reason for the rules (here, improper motive) exists, then the mistakes made without any rules will exceed the mistakes arising from the rules' structure.⁸¹ The decision to use such rules thus follows from the combination of one fundamental principle and one unfortunate fact. The principle is

⁸¹ For general discussion of this rationale for the use of rules, see Frederick Schauer, *The Second-Best First Amendment*, 31 *Wm & Mary L Rev* 1, 9-12 (1989). Just as rules better enable courts to determine legislative motive, the rules may perform the same function for legislators. Odd as it may seem, it may be easier for legislators to follow rules relating to the terms of a law than to follow a command not to consider illicit factors.

that the First Amendment bans restrictions on speech arising from hostility, sympathy, or self-interest. The fact is that courts cannot enforce this ban directly.

If all this is so, the First Amendment rules of which I am speaking, though seemingly substantive in content, resemble in function such procedural mechanisms as presumptions and shifting burdens of proof. Consider how a different body of law responds to the difficulty of proving motive. Under the labor laws, an employer may not discharge an employee because of union activity. A court deciding whether such an act has occurred will shift the burden of proof on the question of motive to the employer once the employee has made a lesser (*prima facie*) showing.⁸² In so doing, the law in effect establishes a rebuttable presumption: the law presumes improper motive from a set of facts merely suggestive of it unless the employer proves its absence. The rules of First Amendment law work in a similar manner. They too operate, though not overtly, to make a rebuttable determination of improper motive on the basis of some set of facts—for example, a content-based classification—suggestive but not dispositive of it. It is in this sense that I have spoken of these rules as evidentiary in nature: they, no less than such procedural mechanisms as presumptions and shifting burdens, serve to ameliorate troublesome problems of proving motive by giving exceptional weight to certain evidentiary materials.

This hypothesis suggests a reinterpretation of *O'Brien*. No longer should that decision be viewed as a broad-scale stricture against invalidating regulations of speech on the basis of improper motive. That understanding of the case has always conflicted with too much in the Court's rhetoric and decisions. *O'Brien* stands for a narrower proposition, relating not to the propriety of inquiring into motive, but to the means by which to conduct this inquiry. To be more precise, *O'Brien* stated not that motive was irrelevant, but only that it could not be proved by traditional methods. In so doing, the decision left open the option of adopting a different mechanism to discover motive. The Court, as the next Section of this Article shows, has chosen this course in its elaboration of First Amendment doctrine.

⁸² See *NLRB v Transportation Management Corp*, 462 US 393, 403 (1983). A similar though less potent procedural mechanism, designed to accomplish the same object, is used in Title VII cases. See *Texas Department of Community Affairs v Burdine*, 450 US 248, 252-53 (1981).

III. THE DOCTRINE OF IMPERMISSIBLE MOTIVE

Let us accept that the First Amendment prohibits restrictions on speech stemming, even in part, from hostility, sympathy, or self-interest. And let us accept that the difficulty of proving this impermissible motive—resulting, most notably, from the government's ability to invoke pretextual reasons—gives rise to a set of rules able to flush out bad motives without directly asking about them. What would these rules look like?

The first rule would draw a sharp divide between content-based and content-neutral restrictions, with a fuzzier line bisecting the world of content-based restrictions into those based on viewpoint and those on subject matter. The second and third rules would specify exceptions to the first: instances in which a restriction, though content-neutral, demands heightened scrutiny because of suspect origin; instances in which a restriction, though content-based, could receive relaxed scrutiny because apparently safe. And the fourth rule would draw another sharp distinction, this time between actions directly addressed to speech and those affecting speech only incidentally.

These rules—the rules that would be devised to flush out illicit purpose—in fact constitute the foundation stones of First Amendment doctrine. Examining their structure reveals that the search for impermissible motive animates the doctrine, as the doctrine implements the search for motive.

A. The Distinction Between Content-Based and Content-Neutral Laws

The distinction between content-based and content-neutral regulations of speech serves as the keystone of First Amendment law.⁸³ Content-neutral restrictions on speech—restrictions that by their terms limit expression without regard to what is said—usually are subject to a fairly loose balancing test. So, for example, in reviewing a law that bans all billboards within city limits, the Court might consider the strength of the state interests asserted (say, in aesthetics and traffic safety), the availability of alternative means to protect those interests, the extent to which the law limits expression, and the existence of alternative avenues of communication. This analysis may well result in a decision that the law accords with the First Amendment. Con-

⁸³ The fullest description and analysis of this distinction remains Stone, 25 *Wm & Mary L Rev* at 189 (cited in note 15).

tent-based restrictions on speech—restrictions that by their terms limit expression on the basis of what is said—usually are subject to far more rigorous scrutiny. This is true even in cases like *R.A.V.* in which the government concededly could restrict the speech affected through a broader law written in content-neutral language. Formulations of the standard used to review content-based action vary, but the Court most often requires the government to show a compelling interest that could not be attained through less restrictive means. Application of this standard usually leads to a law's invalidation.

Somewhat mitigating the starkness of this scheme, a further, hazier distinction operates within the realm of content-based regulation. Here, the Court often differentiates between viewpoint-based restrictions and all other content-based restrictions, including, most notably, restrictions based on subject matter.⁸⁴ So, for example, the Court would treat differently a law prohibiting the use of billboards for all political advertisements and a law prohibiting the use of billboards for political advertisements supporting Democrats. The former might meet constitutional standards; the latter would never succeed in doing so. It is not so much that the Court formally uses two different standards for subject matter and viewpoint regulation; in most contexts, a strict scrutiny standard applies to content-based action of all kinds.⁸⁵ But the Court, when reviewing subject-matter restrictions, either may apply a purportedly strict standard less than strictly or may disdain to recognize the law as content based at all.⁸⁶ By contrast, the Court almost always rigorously reviews and then invalidates regulations based on viewpoint.

This scheme makes no sense under the speaker-based model of the First Amendment.⁸⁷ Recall that this model treats as critical the sum total of expressive opportunities; the more a law

⁸⁴ See generally Geoffrey R. Stone, *Restrictions of Speech Because of its Content: The Peculiar Case of Subject-Matter Restrictions*, 46 U Chi L Rev 81 (1978).

⁸⁵ In some rare contexts—most notably, in nonpublic forums—the Court explicitly adopts different standards for subject-matter and viewpoint regulation. See Kagan, 1992 S Ct Rev at 42-43 (cited in note 14).

⁸⁶ For a case in which the Court applied a toothless version of strict scrutiny, see *Burson v Freeman*, 504 US 191, 198-211 (1992). For a case in which the Court pretended that a subject-matter restriction was content neutral, see *Rowan v United States Post Office Department*, 397 US 728, 737-38 (1970). Of course, in many subject-matter cases, the Court applies a strict scrutiny standard with all the rigor its name implies. See, for example, *Police Department of Chicago v Mosley*, 408 US 92, 98-102 (1972).

⁸⁷ See Redish, 34 Stan L Rev at 128-39 (cited in note 34) (criticizing the distinction on this ground); Stone, 25 Wm & Mary L Rev at 197 (cited in note 15) (acknowledging the point, but approving the distinction).

curtails the ability to speak, the greater its constitutional difficulty. Yet a content-neutral law, no less than a content-based law, can lessen the ability to speak; indeed, a content-neutral law can do so more dramatically. To use my earlier example, a general ban on billboards will reduce speech more than a ban on billboards for political advertisements, which in turn will reduce speech more than a ban on billboards disabling only Democrats. Yet under current law, the Court will subject the first of these ordinances to the most relaxed form of review and the last to the strictest. Consider in the same vein the cases of *Police Department of Chicago v Mosley*⁸⁸ and *Grayned v City of Rockford*.⁸⁹ In the latter, the Court upheld a content-neutral ban on speech in the vicinity of a school; in the former, the Court invalidated a similar ban on the ground that it exempted speech about labor disputes from its general prohibition. Finally, recall the Court's view in *R.A.V.* that although a ban on all fighting words would have passed muster, with the category of fighting words treated as content neutral,⁹⁰ a ban on fighting words limited to a certain subject—worse, to a certain viewpoint—violated constitutional norms. Each of these examples shows that a concern with the extent of expressive opportunities cannot explain the most critical aspect of First Amendment doctrine.

Perhaps, however, a concern with the quality of the speech market—the concern of the audience-based model—may explain the distinction between content-based and content-neutral regulation.⁹¹ The argument, anticipated in my discussion of *R.A.V.*, relies on the “distorting” effect of content-based, and especially viewpoint-based, regulation. The edict “no billboards” on its face handicaps equally all ideas. The edict “no ads for Democrats on billboards,” by contrast, disadvantages certain ideas to the benefit of others. Finally, the edict “no political ads on billboards” falls in between these extremes. Bans of this kind at the least disfavor one subject of discussion compared with others. And they often (at any rate, more often than content-neutral restrictions) oper-

⁸⁸ 408 US 92 (1972).

⁸⁹ 408 US 104 (1972).

⁹⁰ The Court analogized the regulation of fighting words to the regulation of sound trucks, which of course is content neutral. The Court explained: “[F]or purposes of [the First] Amendment, the unprotected features of [fighting] words are, despite their verbal character, essentially a ‘nonspeech’ element of communication.” *R.A.V.*, 505 US at 386. I discuss later the reasons for treating the category of fighting words as neutral with respect to the content of speech. See text accompanying note 182.

⁹¹ For such an argument, see Stone, 25 *Wm & Mary L Rev* at 217-27 (cited in note 15); Sunstein, *Democracy and the Problem of Free Speech* at 170 (cited in note 19).

ate to skew debate among competing ideas on a single subject: consider, for example, if in 1970 the government had banned discussion of the Vietnam War. Perhaps, then, current doctrine responds to the different ways in which viewpoint-based, subject-matter-based, and content-neutral laws distort public discourse and thereby (in Meiklejohn's phrase) mutilate the community's thinking process.

Even assuming, however, that a law disparately affecting ideas necessarily skews the speech market—an assumption I contest shortly—this justification of the divide between content-based and content-neutral regulation suffers from two related weaknesses.⁹² First, a doctrinal structure based on the problem of distortion seemingly would subject to heightened scrutiny whatever content-neutral rules fall much more heavily on one idea than others. Suppose, for example, that only Democrats, and not Republicans, use billboards to advertise; then, the skewing effect of a general ban on billboards would match the skewing effect of a law specifically barring Democrats from this forum. To put the point more generally, content-neutral laws often have content-based effects—and sometimes these are quite dramatic. A jurisprudence concerned with distortion should treat these cases with the utmost seriousness. But current doctrine all but ignores the distorting effects of content-neutral law.⁹³

Second and conversely, a body of doctrine based on the problem of distortion apparently would subject to relaxed review any content-based laws that have only a modest tilting effect. Consider again our viewpoint-based billboard regulation; if neither Democrats nor Republicans use billboards, disallowing such use for one party only will not skew public discourse. Or recall again *R.A.V.* where the ban on racist fighting words could not seriously have distorted the deliberative process. In such cases, the small quantity of speech affected, combined with the ready availability of alternative means to communicate the “handicapped” idea, makes the danger of distortion insignificant. Yet First Amendment doctrine distinguishes not at all between content-based

⁹² I consider below whether a model focusing on motive also suffers from these weaknesses and conclude that it does not—or, at least, that it does not do so to the same extent. See text accompanying note 104.

⁹³ For discussion of the cases, see Stone, 25 *Wm & Mary L Rev* at 218-22 (cited in note 15). For an argument that content-neutral laws with significant content-based effects ought to be treated as if facially content-based because of the extent to which these laws distort debate, see Susan H. Williams, *Content Discrimination and the First Amendment*, 139 *U Pa L Rev* 615, 655-63 (1991).

laws of this kind and content-based laws that wholly excise ideas from public discourse.⁹⁴

One explanation of these oddities refers to the difficulty of deciding when a regulation has a skewing effect that is sufficiently large or small to alter the usual standard of review.⁹⁵ Any such decision necessarily will involve difficult questions relating to what speakers use what forms of speech, as well as to how effectively different forms of speech (both the form restricted and its alternatives) communicate a desired message. Perhaps, given these difficulties, the distinction between content-based and content-neutral regulation functions as an imprecise (both over- and underinclusive) but still sensible mechanism for sorting out consequential from inconsequential skewing effects, in accord with the dictates of the audience-based model.

This explanation, however, is unconvincing. True, there would be hard cases if courts evaluated skewing effects—cases involving tricky issues of measurement and line drawing. But such problems seem no more common or intractable in this adjudicative context than in many others, where no one thinks they preclude evaluative efforts. There also would be many simple cases as measured by an effects-based ruler—cases in which courts confidently could say either that a content-based law would have minor skewing effects or that a content-neutral law would cause major distortion. Recall again *R.A.V.*, where the ordinance would have had utterly insignificant skewing effects. Were courts primarily concerned with distortion they would at least modify the strict distinction between content-based and content-neutral laws to respond to the host of cases in which they could directly evaluate skewing effects.

Indeed, to the extent this conclusion is wrong, it is so because of a fear of improper motive. Suppose, that is, there is some special reason to resist case-by-case line drawing with respect to the skewing effects of a speech restriction. What would this reason be? It likely would relate to the fear that a judge's own biases toward the speech affected would taint her decision as to whether the restriction had a severe or narrow skewing effect.⁹⁶ But if this is the reason for preventing judges from mak-

⁹⁴ For discussion of the case law, see Stone, 25 *Wm & Mary L Rev* at 200-01 (cited in note 15).

⁹⁵ See *id* at 224-27.

⁹⁶ See *id* at 225 (Judges "may be influenced by . . . biases that may undermine their ability to evaluate accurately and impartially the extent to which particular . . . restrictions actually impair the communication of specific, often disfavored, mes-

ing case-specific determinations as to distortion, then the doctrine arises from considerations of motive at least to this extent: that fear of illicit motive constrains and structures the inquiry as to effects. And if this is so, then it is at least true that the doctrine attempts not to create a distortion-free universe, but only to accomplish as much as can be accomplished in this direction consistent with an omnipresent fear of improper motive.

But much more than this can be said. The discussion so far has assumed that the disparate impact of a law on ideas will distort the speech market. If that assumption is false, then the distinction between content-based and content-neutral laws—even if the most sensible way of determining whether a law disparately affects ideas—would not further the interest in balanced discourse.

In fact, this assumption is hard to defend, for as a conceptual matter, the disparate impact of a law on a set of ideas might lead to balance as easily as to distortion.⁹⁷ Remember that each regulation affecting speech acts against a backdrop of countless other regulations affecting speech, sometimes directly, sometimes incidentally. Among these are rules of property and contract, which provide some speakers with access to the most effective means of expression and consign other speakers to the least so. All these regulations, operating together, give shape and content to the realm of discourse, and given the nature of these rules—specifically, the ways they effect inequalities of wealth and access—the speech environment they create stands little chance of nearing the ideal condition. Distortion, skew, tilt—whatever one calls lapses from the ideal—will occur all over. In such a setting, any law with a disparate impact on ideas may succeed in balancing, no less than skewing, the speech market; conversely, any law affecting ideas equally may perpetuate a skewed, no less than a balanced, speech environment. As a logical matter, such laws will do the one thing no less than the other.

The distinction between content-based and content-neutral regulation thus cannot rest on a pure audience-based approach to the First Amendment. It is true that this distinction reflects the likelihood that a law will change the prevailing structure of pub-

sages.”); Ely, *Democracy and Distrust* at 112 (cited in note 67) (Such evaluations “inevitably become involved with the ideological predispositions of those doing the evaluating.”).

⁹⁷ See text accompanying notes 19-20. For fuller discussion, see Owen M. Fiss, *Free Speech and Social Structure*, 71 *Iowa L Rev* 1405, 1410-13 (1986); Sunstein, *Democracy and the Problem of Free Speech* at 177-80 (cited in note 19).

lic comment. But if that structure itself departs from the ideal—if, in Meiklejohn's words, the existing distribution of views itself "mutilat[es]" the "thinking process of the community"⁹⁸—then the presumption against content-based law may not serve to protect this process. In such a world (which is our world), a content distinction has no necessary tendency to impede the goals of the audience-based model.

Indeed, this model of the First Amendment might well command (not merely tolerate) the use of content discrimination in some circumstances. As one proponent of the view has urged, "governmental action . . . based on content . . . might be needed to protect our freedom" by ensuring that "public debate [is] enriched and our capacity for collective self-determination enhanced."⁹⁹ No proposal could be further from current doctrine; the use of the audience-based model seems to counsel discarding the keystone of the law for its opposite.

Few courts or commentators would view with equanimity a reform of this kind; indeed, even proponents of the audience-based model might favor the Court's decision to use as its benchmark the actual, rather than the perfect, distribution of viewpoints. The stated reason might run something as follows: no matter how unhealthy the existing speech market, governmental action directed at the content of speech would cause in most cases further harm. But given all I have said, how is it possible to defend this assertion? The answer to this question will suggest the deepest wellspring of First Amendment doctrine—the concern that drives and indeed defines all others.

One explanation for our choice of benchmark refers to the difficulty—even the incoherence—of defining the ideal realm of discourse.¹⁰⁰ We do not have a full picture of what a well functioning marketplace of ideas would look like. Who would say what in such a system? At what point would an idea become over- or underrepresented? Perhaps we can provide no account of the optimal mix of expression. And if we cannot describe the ideal, perhaps we also cannot decide whether an action would bring us closer to, or take us further from, this state of perfec-

⁹⁸ Meiklejohn, *Political Freedom* at 27 (cited in note 15) (emphasis omitted).

⁹⁹ Fiss, 71 *Iowa L Rev* at 1415 (cited in note 97). See also Sunstein, 59 *U Chi L Rev* at 290-91 (cited in note 63) ("[E]fforts to restructure the marketplace [of ideas] might even be seen as the discharge of the legislature's constitutional duty, a duty that courts are reluctant, for good institutional reasons, fully to enforce.").

¹⁰⁰ For fuller discussion of these issues, see Strauss, 91 *Colum L Rev* at 349 (cited in note 39); Strauss, 1993 *U Chi Legal F* at 202-10 (cited in note 20).

tion. The government, on this view, could not tell whether its decisions respecting ideas would ameliorate or exacerbate distortion. Hence derives the command to leave things alone.

But this account of the presumption against content-based regulation of speech suffers from two notable defects. First, the folly of attempting to evaluate these matters can be overstated. Even without a fully fleshed out conception of the ideal, observers sometimes will be able to make sensible claims about what problems of distortion exist and how to fix them. For example, it is not incoherent (it may even be correct) to suggest that campaign finance restrictions would improve the speech market. Of course not everyone will agree on these matters, but not everyone agrees on any matter respecting the desirability of governmental action. The key point is that it will not always (even if it will often) be impossible to reach a cogent, well supported decision on the effects of regulation on the existing speech market.

Second, and more important, the presumption against content-based action cannot arise from an inability to evaluate skewing effects, even assuming this inability existed. Recall that all governmental action, whether or not directed toward the content of speech, has effects on the speech market. If we command the government to forego content-based regulation, on the ground that we cannot evaluate its effects, we do not command the government "to leave things alone"; we instead command the government to let stand the effects of content-neutral action, which we also cannot evaluate. The theory at issue here provides no reason for preferring the one set of effects to the other. Thus, the difficulty of evaluating distortion, taken alone, cannot justify the distinction between content-based and content-neutral regulation; it could as well explain a presumption against the one as against the other—or against both, or against neither.

A better explanation for measuring distortion by reference to the *ex ante* distribution of views—for assuming that content-based action will skew public debate—relates to fear of impermissible governmental motive. We presume that content-based regulation will exacerbate rather than minimize existing bias because we believe that such regulation is disproportionately linked to suspect motives. Suppose, that is, that the content-based nature of an action provides a special reason to distrust the government's motives; this distrust then provides a reason to conclude, without further evidence, that the action's effects are untoward. On this account—the only one that makes sense—the

view that content-based regulation skews debate reduces to the view that content-based regulation emerges from illicit motives.

The goal of the doctrine, then, must be to identify a set of improper motives, which themselves may give rise to untoward consequences—not to identify a set of untoward consequences defined independent of improper motives.¹⁰¹ Purpose is the crux of the matter—whether, as suggested above, the concern with purpose ultimately has something to do with consequences, or whether, as discussed in Section IV, the concern has other, nonconsequential sources. The critical question is thus whether the distinction between content-based and content-neutral action—more specifically, the distinction among viewpoint-based, other content-based, and content-neutral action—facilitates the effort to flush out improper purposes.

The distinction in fact serves just this function: it separates out, roughly but readily, actions with varying probabilities of arising from illicit motives.¹⁰² Consider again our billboard laws. Ideological reasons are unlikely to taint the content-neutral statute, which prohibits billboards of any content. Because the statute applies to all ideas, a legislator's decision to support it probably will not rest on hostility or sympathy to particular messages. (Such circumstances are equally unlikely to taint a court's decision to uphold the statute.) By contrast, improper purpose probably will infect the viewpoint-based statute, which prohibits the use of billboards to endorse candidates of a single party. Here, a legislator's view as to the merits of particular ideas—the idea restricted and its competitors—will intrude, whether consciously or not, on the decision whether the harms caused by the speech justify the regulation. (Likewise, a court probably will incorporate impermissible considerations in ruling on the statute.) A concern with governmental purpose—unlike a concern with effects on speaker or audience—thus explains the division between restrictions applying to all viewpoints and restrictions applying only to one.

So too the concern with purpose explains the intermediate treatment given to subject-matter restrictions. Hostility or sympathy toward ideas is less likely to taint a subject-matter restriction than a viewpoint restriction precisely because the former, by its terms, applies to a range of ideas. But a subject-matter re-

¹⁰¹ Otherwise put, the concept of skewing effects in the law of the First Amendment means only whatever effects arise from actions based on illicit motive.

¹⁰² See Stone, 25 *Wm & Mary L Rev* at 230-33 (cited in note 15).

striction poses a greater risk of improper purpose than the usual content-neutral law. Consider again a ban on political speech in a given forum (in my running hypothetical, on billboards). Though neutral reasons may support such a ban,¹⁰³ there is heightened danger that the government is acting in part for illicit reasons—because, say, officials know that a disfavored political group disproportionately uses this mode of communication. Such a purpose could infect as well the law banning all billboards, but as the law applies to an ever greater range of ideas, the probability of taint decreases.

Finally, the focus on purpose explains why First Amendment doctrine ignores the severe skewing effects of some content-neutral laws and the slight skewing effects of some viewpoint-based laws. It is true that such effects may offer evidence relating to motive; presumably the greater the skewing effect, the greater the chance that illicit considerations have intruded. But in most cases a law's terms more reliably indicate illicit motive than its effects and thus should control the legal analysis.¹⁰⁴ A content-neutral law, even when it has severe skewing effects, poses only a minor risk of improper motive because the law creates such effects along so many dimensions. The diffuseness of the law outweighs the severity of its impact on any particular idea as evidence of motive. Conversely, a content-based law, even when it has insignificant skewing effects, presents a substantial risk of impermissible motive because the effects occur in so narrow an area. The focused nature of the law outweighs the mildness of its impact on an idea as evidence of motive. This is why courts would treat differently our three billboard laws even if, as could be true, the laws similarly affected the distribution of political views. The terms of the laws indicate, even if the effects of the laws do not, disparate risks of improper motive.

The distinctions among viewpoint-based laws, other content-based laws, and content-neutral laws thus create a set of pre-

¹⁰³ The government offered such legitimate justifications in *Greer v Spock*, 424 US 828 (1976), and *Lehman v City of Shaker Heights*, 418 US 298 (1974), which involved bans on political speech on an army base and in a mass transit system's advertising space. The government in the former case invoked the interest in keeping the military removed, in both appearance and reality, from partisan causes. *Greer*, 424 US at 839. The government in the latter case pointed to administrative problems involved in allocating limited space to political candidates. *Lehman*, 418 US at 304. Both of these motives, in the scheme of this Article, are perfectly permissible.

¹⁰⁴ I here put to one side the concern that bias will infect the measurement of a law's skewing effects. These concerns only strengthen the conclusion that the face of a law indicates more reliably than the effects of the law what purposes underlie it.

sumptive conclusions about when improper motive has tainted a restriction on speech. These distinctions will suggest some “wrong” results, for some viewpoint-based laws arise solely from legitimate reasons and some content-neutral laws arise partly from their opposite. We tolerate the imprecision because the alternative—a direct inquiry into motive—will produce even more frequent errors. If the facial markers we use are not perfect, they are better than what they replace.

We also attempt to mitigate the imprecision by making the outcomes suggested by these facial distinctions presumptive only and requiring courts to consider evidence to the contrary. But this evidence of motive again takes an indirect form: it resides at this stage in the substantiality of the asserted legitimate interest for the restriction and the closeness of the fit between that interest and the terms of the law.

So, for example, the strict scrutiny standard—indeed, each component of it—is best understood as an evidentiary device that allows the government to disprove the implication of improper motive arising from the content-based terms of a law.¹⁰⁵ This is true first of the compelling interest requirement: the stronger the state interest asserted, the more likely it is that the government would act to achieve that interest in the absence of antipathy toward the speech affected. Similar reasoning applies to the demand for close tailoring. If a restriction applies to more speech than necessary to achieve the interest asserted, the suspicion deepens that the government is attempting to quash ideas as ideas rather than to promote a legitimate interest. And if a restriction applies to less speech than implicates the asserted interest, so too the concern grows that the interest asserted is a pretext.¹⁰⁶ But if a restriction fits along both dimensions—if it applies to all and also to only the speech that threatens the assert-

¹⁰⁵ John Hart Ely has made a similar argument on the use of strict scrutiny in equal protection law. See Ely, *Democracy and Distrust* at 145-48 (cited in note 67). There, a compelling interest and a close fit between means and end negate the presumption of illicit motive arising from use of a suspect classification such as race. In the First Amendment context, matters are the same, except that the suspect classification is content.

¹⁰⁶ In effect, the content-based nature of the law has raised suspicions so great that the usual defense of taking “one step at a time,” see *Williamson v Lee Optical Co.*, 348 US 483, 489 (1955), is unacceptable. We view the underinclusive action not as a first step toward achieving a legitimate end, but as confirmation of an illegitimate purpose. See, for example, *Erznoznik v City of Jacksonville*, 422 US 205, 214-15 (1975) (holding that an ordinance prohibiting nudity in drive-in movie theaters could not be justified as a traffic regulation because “a wide variety of other scenes in the customary screen diet . . . would be [no] less distracting”).

ed interest—then there is an assurance that the government has acted for proper reasons. In this way, the strict scrutiny test operates as a measure of governmental motive. The showing that the government must make under that standard does not serve, as on a scale, to outweigh impermissible motive or counter its harms. The showing instead serves an evidentiary function: to disprove (again, of necessity indirectly) the inference of bad motive that arises from the content-based face of a law.¹⁰⁷

A similar mechanism, though operating in reverse, is necessary to assess the constitutionality of content-neutral laws. So long as a content-neutral law has differential effects on particular ideas—even assuming those effects are widely dispersed—it may bear the taint of improper motive. Officials may care so much about suppressing a particular idea affected by a content-neutral law as to disregard or tolerate the law's other consequences. Or, in a slightly different vein, officials may desire a broad-scale entrenchment of status quo positions and enact a law restricting all expression in a certain sphere in order to achieve this object.¹⁰⁸ Such a restriction, in addition to benefiting ideas already accepted, allows the government to emerge as the dominant speaker in the sphere, able to control opinion through speaking itself rather than through regulating the speech of private parties. For these reasons, the presumption that content-neutral laws are untainted by impermissible motive must remain just that—a presumption subject to rebuttal.

The relatively deferential standard governing the constitutionality of content-neutral laws serves this kind of evidentiary function. Just as passing strict scrutiny demonstrates that a content-based law has a legitimate purpose, thus rebutting a presumption of impropriety, so flunking this looser standard demonstrates that a content-neutral law has an illegitimate basis, thus rebutting the opposite presumption. The way this occurs should by now be clear. The less significant the legitimate inter-

¹⁰⁷ Justice Kennedy has expressed a similar view of the function of the strict scrutiny standard in First Amendment law. He has written that “the compelling interest test . . . determine[s] the accuracy of the justification the State gives for its law.” *Freeman*, 504 US at 213 (Kennedy concurring). I take this to be another way of saying that the strict scrutiny standard is a tool for discovering the government's real motive.

¹⁰⁸ Subject-matter restrictions accomplish a narrower de facto favoring of status quo positions: for example, a law prohibiting all discussion of the government's foreign policy will favor whatever views on that policy are currently dominant. In such a case, the danger of impermissible motive is significant because of the restricted scope of the law (and of the consequent skewing effect). Content-neutral laws, which favor status quo positions generally, pose a lesser, but still cognizable, danger.

est supporting the law, the greater the reason to distrust the government's action. Similarly, the looser the fit between the interest asserted and the contours of the law, the greater the cause for suspicion. At a certain point—when the asserted interest is insubstantial or when it does not fit the scope of the challenged regulation—the usual presumption of proper purpose topples; there is reason, then, to think that the law, though content neutral, has been tainted by impermissible purpose.¹⁰⁹

I do not mean to say that review of content-neutral regulations serves only, or even primarily, as a mechanism to discover bad motive. The review of content-neutral laws also functions to ensure adequate expressive opportunities, in keeping with what I have called the speaker-based perspective. If expressive activity has special value to individuals, then the government should have to justify in a special way, by offering unusually weighty countervailing interests, any restriction on expression.¹¹⁰ In this way, one of the effects-based models supplements the purpose-based model in explaining the standard of review for content-neutral legislation.¹¹¹

¹⁰⁹ The Court's "public forum" doctrine can be seen as establishing a kind of backstop to this relatively easy test for illicit motive. Even if a content-neutral law tainted with improper motive manages to pass this test, the public forum doctrine may prevent the law from operating in certain places. Thus, public forums serve as a kind of safe harbor for speech, providing in certain areas an extra level of protection. Of course, public forum doctrine also functions to ensure a minimum level of opportunities for expression, in line with the speaker-based model of the First Amendment. Viewed from this perspective, public forum doctrine insulates a zone of speech against even properly motivated governmental action; only outside that restricted zone does the question become one of motive.

¹¹⁰ This rationale explains the tendency of the Court, in practice, to review content-neutral regulations more strictly when they have a severe effect on expressive opportunities and less strictly when they have a modest effect. See Geoffrey R. Stone, *Content-Neutral Restrictions*, 54 U Chi L Rev 46, 58-59 (1987). Under the speaker-based theory, as the effect of a law on expressive opportunities increases, so too should the government's burden of justification. A motive-based theory might add a further reason for varying the standard of review in this way. As the impairment of expressive opportunities increases relative to the importance of the asserted governmental interest, so too does the suspicion grow that the government is acting for illegitimate reasons.

¹¹¹ The audience-based model of the First Amendment explains less well than its competitors the standard of review applied to content-neutral legislation. If "[w]hat is essential is not that everyone shall speak, but that everything worth saying shall be said," Meiklejohn, *Political Freedom* at 26 (cited in note 15), then some content-neutral restrictions on speech (like some content-based restrictions) will count not as harms, but as positive goods. The "traditional American town meeting," *id* at 24, which Meiklejohn used as the model of public debate, indeed depends on rules of order. Some content-neutral laws, of course, may restrict speech so broadly as to disserve the interests of the audience. But the audience-based theory, taken alone, cannot explain the practice of subjecting not only these but all content-neutral laws to special scrutiny. For that result to follow, it is necessary to refer either to the interests of potential speakers or to the possibility of

But only the purpose-based model can explain the difference in the levels of review applicable to content-based and content-neutral laws. In justifying this distinction, the speaker-based model provides no assistance. The audience-based model becomes useful only to the extent that it reduces to a motive inquiry, concerned not with skewing effects per se, but only with such effects as arise from, and are defined by, illicit purpose. It is the quest for purpose that drives the doctrine—that explains the divergent treatment given to content-based and content-neutral laws. So too that quest explains the creation of exceptions to this doctrinal division: instances in which the courts treat facially content-neutral laws as content based or facially content-based laws as content neutral. The next two Sections deal with these exceptions and the role of motive analysis in explaining them.

B. “Suspect” Content-Neutral Laws

Although most content-neutral laws pose small danger of stemming from improper purpose, some such laws present greater risk. Three kinds of content-neutral laws, in particular, raise the same specter of improper purpose as the typical content-based governmental action. This suspect trio of facially content-neutral laws consists of (1) laws conferring standardless discretion on administrative officials; (2) laws turning on the communicative effect of speech; and (3) laws attempting to “equalize” the speech market. Because these laws, though content neutral, sound the alarm of illicit purpose, First Amendment doctrine treats them just as it does content-based actions. Here too, the standard applied operates as a presumption of improper motive, adopted in response to difficulties of proof.

1. Laws conferring discretion.

Consider this law: “It shall be unlawful for any person to maintain and operate . . . any radio device, mechanical device, or loud speaker . . . which is so placed and operated that the sounds coming therefrom can be heard to the annoyance or inconvenience of travelers upon any street . . . [except if] the same be done under permission obtained from the Chief of Police.”¹¹²

In *Saia v New York*, the Court held this ordinance unconstitutional on its face because it prescribed “no standards . . . for

improper purpose on the part of legislators or judges.

¹¹² *Saia v New York*, 334 US 558, 558-59 n 1 (1948).

the exercise of [the police chief's] discretion."¹¹³ The Court likewise has invalidated laws broadly empowering officials to grant (or deny) permission to distribute leaflets, hold parades, or erect newsracks on public property—all, again, on the ground that the laws failed to cabin appropriately administrative discretion.¹¹⁴ In none of these cases did the Court wait to find that the administrator actually had abused her discretion. In several of the cases, the administrator had no opportunity to engage in abuse because the speaker never applied for the requisite license.¹¹⁵ In most of the cases, the administrator might well have succeeded in denying the license (assuming the speaker had applied) under a properly drafted statute. The constitutional problem in the cases thus arose not from any administrative decision restricting speech, but from the wide authority that the statute gave to administrators to restrict speech, both then and in the future.¹¹⁶ Otherwise put, the rule established in these cases responded not to any actual abuse of governmental authority, but only to the potential for abusive conduct.

What could account for a rule of this sort? The best explanation, stated briefly for now, goes as follows. A law conferring standardless discretion effectively delegates to administrators the power to make decisions about speech on the basis of content. Such administrative decisions raise the same constitutional concern as do content-based laws: the danger of improper motive. Because courts cannot discover improper purpose directly, they use as a proxy the presumption against content-based distinctions. But likewise, because courts cannot easily determine, in the context of administrative action, when a content-based decision has occurred, they here add a further prophylactic rule, designed both to prevent and to detect content-based administration. The fundamental purpose of this rule barring standardless discretion thus resides in its capacity to assist in the campaign against impermissible motive.¹¹⁷

¹¹³ 334 US 558, 560 (1948).

¹¹⁴ See *Lovell v City of Griffin*, 303 US 444, 450-51 (1938) (leaflets); *Shuttlesworth v City of Birmingham*, 394 US 147, 150-51 (1969) (parades); *City of Lakewood v Plain Dealer Publishing Co.*, 486 US 750, 769-70 (1988) (newsracks).

¹¹⁵ See *Lovell*, 303 US at 448; *Lakewood*, 486 US at 754.

¹¹⁶ As the Court stated in *Freedman v Maryland*, "[i]n the area of freedom of expression . . . one has standing to challenge a statute on the ground that it delegates overly broad licensing discretion to an administrative office, whether or not his conduct could be proscribed by a properly drawn statute, and whether or not he applied for a license." 380 US 51, 56 (1965).

¹¹⁷ A similar argument explains at least in part the First Amendment's vagueness

To amplify this understanding of the rule, it is best to begin by rejecting the alternatives. The speaker-based model cannot explain the rule against standardless licensing schemes because such schemes do not necessarily curtail more speech than other, less constitutionally suspect modes of restricting expression. To see this point, contrast *Saia* to *Kovacs v Cooper*,¹¹⁸ decided seven months later. In *Kovacs*, the Court upheld an ordinance banning the use “upon the public streets” of any “sound truck, loud speaker, or sound amplifier.”¹¹⁹ The amount of speech that this ordinance limited must have exceeded the amount that the law in *Saia* curtailed; the latter, after all, allowed exemptions from the ban. The constitutional difficulty in *Saia* thus could not have stemmed from the extent of the restriction. The same is true of all standardless licensing schemes, which may restrict less, as well as more, speech than a cabined licensing system or a flat prohibition.¹²⁰ Nothing in the nature of standardless licensing indicates the amount of speech either actually or potentially restricted. If this is true, nothing in the nature of standardless licensing makes it peculiarly problematic from the standpoint of maximizing communicative outlets.

Similarly, the audience-based model fails to explain the rule against standardless licensing schemes. Standardless licensing, as I will show, may well encourage content-based decisions, which punish and deter certain ideas as compared to others.¹²¹

doctrine. Vague laws delegate to administering officials a kind of standardless discretionary authority—here, to interpret and apply unclear language. Such discretion raises much the same concerns of improper motive as I discuss in the text.

¹¹⁸ 336 US 77 (1949).

¹¹⁹ *Id* at 78.

¹²⁰ Another example comes from *Lakewood*, where the Court invalidated an ordinance regulating the placement of newspaper vending machines for failure to limit properly the relevant official’s discretion. 486 US at 750. All Justices assumed for purposes of the case that the city could have chosen to ban all such vending machines. *Id* at 762 n 7; *id* at 773 (White dissenting). The objection to the licensing scheme thus could not have turned on the extent to which it curtailed expressive opportunities.

¹²¹ The Court, in explaining why it recognizes facial challenges to standardless licensing schemes, often focuses on the way such schemes deter or “chill” certain ideas. As the Court said in *Lakewood*:

[The] licensor’s unfettered discretion, coupled with the power of prior restraint, intimidates parties into censoring their own speech, even if the discretion and power are never actually abused. . . . It is not difficult to visualize a newspaper . . . feeling significant pressure to endorse the incumbent mayor . . . to receive a favorable and speedy disposition on its permit application.

Hence, the claim might go, standardless licensing, in Meiklejohn's phrase, mutilates the community's thinking process. But I have reviewed this claim before and found it wanting, except to the extent that it reduces to a concern with improper purpose.¹²² As a conceptual matter, content-based actions as well may improve as mutilate the community's thinking process. So too, then, with the standardless licensing schemes that facilitate such actions. If there is reason to think that as a practical (rather than a conceptual) matter standardless licensing more often will distort than improve public debate, that reason relates to the fear of illicit motive on the part of licensing officials.¹²³ The real question, then, concerns governmental motive; it is whether a rule against standardless licensing will identify and reduce the incidence of improperly motivated administrative decisions.

The rule against standardless licensing indeed serves this function of flushing out bad motive by establishing a safeguard against administrative action based on the content of expression. I have discussed already how and why impermissible motive more often taints content-based than content-neutral decisions.¹²⁴ This taint may affect administrative actions based on content no less than their legislative counterparts. Enforcing a prohibition on improperly motivated action thus would counsel adopting a presumption against any administrative decision based on the content of speech, which indeed the Court has done. But even more is needed to enforce the stricture against improper motive, for content-based administrative action (unlike content-based legislative action) is itself hard to identify. Suppose an official denies a license to a speaker under a statute specifying no standards; in the absence of an admission, a court cannot easily determine whether the official based her decision on the content of the speech (let alone whether she allowed impermissible motive to infect the decision).¹²⁵ To enforce the prohibition on con-

¹²² See text accompanying notes 96-101.

¹²³ The Court's repeated invocation of chilling effects in this context, see note 121, thus derives from the fear of improper motive. Standardless licensing schemes chill certain ideas and not others, as each speaker considers what speech will advance and what speech will hinder her attempt to obtain a license. The reason to assume that the resulting disparate impact necessarily will have adverse effects on public discourse relates to the danger that officials will allow inappropriate factors—hostility, sympathy, or self-interest—to taint their decisions.

¹²⁴ See text accompanying notes 102-04.

¹²⁵ As the Court noted in *Saia*: "In this case a permit is denied because some persons were said to have found the sound annoying. In the next one a permit may be denied

tent-based action—which is itself a proxy for a finding of impermissible purpose—the Court needs yet a further prophylactic rule, commanding facial invalidation of all statutes providing for standardless licensing.

Such a rule decreases the danger of content-based licensing decisions, which in turn decreases the danger of improperly motivated licensing decisions, in two related ways.¹²⁶ First, by requiring that a licensing scheme contain specific, content-neutral standards, the rule directly promotes administrative decision making in accordance with those, and only those, criteria. Although licensing officials may take into account the content of speech even in the face of formal constraints, they are less likely to consider content if the authorizing statute spells out neutral criteria than if it is silent on the grounds for action. Second, the requirement of specific standards facilitates judicial review of the bases of licensing decisions. Although standards hardly ensure that courts will detect licensing decisions grounded in content, they at least provide, as the Court has noted, “guideposts” to aid the inquiry “whether the licensor is discriminating against disfavored speech”,¹²⁷ without such standards to serve as a measuring stick, “the difficulty of effectively detecting, reviewing, and correcting content-based censorship ‘as applied’” must increase.¹²⁸ Standards thus help to prevent and detect content distinctions and, even more critically, the impermissible motives that likely underlie them.

Simply put, the rule against standardless licensing statutes has emerged to prevent a legislature from displacing to administrators the power to make decisions regarding speech on the basis of a criterion—the content of expression—likely to involve impermissible motive.¹²⁹ The next rule I consider is similar, for

because some people find the ideas annoying. Annoyance at ideas can be cloaked in annoyance at sound.” 334 US at 562.

¹²⁶ See David A. Strauss, *The Ubiquity of Prophylactic Rules*, 55 U Chi L Rev 190, 196 (1988).

¹²⁷ *Lakewood*, 486 US at 758.

¹²⁸ Id at 759. At a later point, the Court recognized that the ultimate question was not whether the administrative action was based on content, but whether it was supported by an illegitimate motive. The Court stated that “without standards to bound the licensor, speakers denied a license will have no way of proving that the decision was unconstitutionally motivated, and, faced with that prospect, they will be pressured to conform their speech to the licensor’s unreviewable preference.” Id at 760. Note here how the issue of chill is related to the issue of motive. See notes 121, 123.

¹²⁹ These statutes also may contain a hint of illicit legislative motive. Legislators, after all, know the dangers of conferring unfettered discretion on administrators, and such discretion is not usually necessary to accomplish legitimate ends. Perhaps, then, a grant

it too demands strict scrutiny of a content-neutral law that effectively delegates authority to others to make content-based decisions, raising concerns of improper motive—although here the delegates are not so much administrative officials as members of the public.

2. Laws turning on communicative effect.

Some laws are content neutral on their face, but turn in operation on “communicative effect”—that is, on the way an audience responds to the content of expression.¹³⁰ A common example is a breach-of-the-peace statute, as applied to persons engaged in expressive activity.¹³¹ In the typical case, a speaker expresses certain ideas, an audience makes known its displeasure, and police officers, fearful of public disturbance, arrest the speaker. The statute under which the arrest occurs makes no reference to the content of speech; it applies to whatever speech provokes, or tends to provoke, a hostile reaction. Yet the courts act as if the statute referred in express terms to the ideas that prompted the response, upholding the conviction only on a showing of necessity.¹³² The courts effectively overlook the facial neutrality of the law and focus instead on what sparks its application—the reaction of an audience to a message.

At first glance, these cases seem mysterious under a motive-based theory, for breach-of-the-peace laws appear to raise few concerns of improper purpose on the part of the legislature. A desire to prevent violence counts as a legitimate, nonideological reason for restricting speech under the conception of improper motive used in this Article. Further, the risk seems small of hostility or self-interest creeping into the decision making process. If a statute were to prohibit specific ideas, on the ground that they provoked violence, we indeed would fear that dislike of those ideas influenced the legislature’s decision. But a breach-of-

of standardless discretion indicates a legislative desire for administrators to engage in ideological censorship, which itself counts as improper motive.

¹³⁰ For discussion of the notion of communicative effect, see Ely, *Democracy and Distrust* at 111-15 (cited in note 67). For discussion of the law’s treatment of content-neutral laws turning on communicative effect, on which my analysis is partly based, see Stone, 25 *Wm & Mary L Rev* at 234-39 (cited in note 15).

¹³¹ See *Gregory v City of Chicago*, 394 US 111 (1969); *Cox v Louisiana*, 379 US 536 (1965). See also *Edwards v South Carolina*, 372 US 229 (1963) (common law breach of peace); *Cantwell v Connecticut*, 310 US 296 (1940) (same).

¹³² See cases cited in note 131. In the hostile audience context, courts most often phrase the standard of review in terms of “clear and present danger,” but the standard differs not at all from that used to review content-based statutes.

the-peace statute, or other similar law, preselects no particular ideas as posing a danger, instead applying to whatever speech, in the actual event, threatens violence. (Indeed, such a law may apply not only to speech, but also to conduct posing a risk of disorder; such breadth usually decreases further the chance of illicit purpose.¹³³) Thus, if the focus is on legislators' motives alone, the hostile audience cases seem wrongly decided.

This conclusion, however, might come too fast, even if the focus remains on legislative motives. Legislators, after all, may well know what ideas will provoke a hostile response; at the least, legislators can guess that the ideas most likely to do so will be those most generally seen as controversial or offensive. Similarly, legislators may well know what ideas potentially falling within the statute most often will capture the notice of law enforcement officials. This broad sense of how the law will operate may taint any legislator's consideration of the law, whether she shares or disputes the views of the public and officials. Of course, a legislator also may anticipate the effects of other content-neutral laws, but these effects usually will be so numerous, disparate, and crosscutting as to minimize their potential to corrupt her decision. The case may be different as to a law that operates only against ideas that raise the ire of the public.

An even greater concern involves the motives of law enforcement officials. Hostile audience laws raise, though to a lesser degree, the same broad problem as standardless licensing schemes: the capacity of officials, under such laws, to take action based on their views of ideas. These laws, of course, do not authorize action on this basis. Further, enforcement officials may import into the administration of any restriction on speech such ideological considerations. But a bit of history suggests that hostile audience laws are especially prone to this abuse, as in case after case, decade after decade, police officers have responded hastily, to say the least, to the risk of disorder caused by disfavored speech.¹³⁴ Nor is this history very surprising, given the vague standards contained in most breach-of-the-peace statutes, which make such laws more than usually subject to discriminatory enforcement.¹³⁵ The facility with which improper motive may

¹³³ See Section III.D.

¹³⁴ See *Gregory*, 394 US at 111 (civil rights protest); *Cox*, 379 US at 536 (same); *Edwards*, 372 US at 229 (same); *Feiner v New York*, 340 US 315 (1951) (speech endorsing racial equality and criticizing public officials); *Chaplinsky v New Hampshire*, 315 US 568 (1942) (speech attacking religion and government); *Cohen v California*, 403 US 15 (1971) (speech attacking the draft).

¹³⁵ For example, the statute in *Cohen* prohibited any person from "maliciously and

taint—and the frequency with which it has tainted—the enforcement of hostile audience statutes thus provides another reason for treating such statutes with suspicion.

Most important, a hostile audience law, as applied, may constitute an improper delegation of authority to the public to suppress messages it disfavors. Consider how such a law operates in practice. A public disturbance of some kind triggers the restriction, but the content of the speech usually triggers the disturbance, and ideological considerations usually enter into the public's response to the speech's content. The constitutional difficulty grows out of this nexus between the government's action and the majority's viewpoint. First Amendment doctrine treats identically regulation stemming from incumbent officials' hostility toward ideas and regulation stemming from (or deferring to) the analogous attitudes of the majority.¹³⁶ As the government cannot itself censor citizens for ideological reasons, so too it cannot authorize one group of citizens to censor another.¹³⁷ This principle leaves the government free to restrict speech provoking a hostile response when necessary to prevent violence, but bars the government from restricting such speech in reflection of or deference to majoritarian bias. Stringent scrutiny of each application of a hostile audience law separates the one kind of action from the other; it ensures that the government has not simply empowered the public to restrict speech on the basis of content, with all the likelihood of ideological censorship such a restriction implies.¹³⁸

The key to the analysis of this Section is first, the functional equivalence between statutes referring to content and statutes turning on communicative impact and second, the relation between content discrimination and impermissible motive. The first point is that laws turning on communicative impact (much like laws establishing standardless licensing schemes), though con-

willfully disturbing the peace or quiet of any neighborhood or person . . . by . . . offensive conduct." 403 US at 16. Similarly, the statute in *Gregory* made it unlawful for any person to make "any improper noise, riot, disturbance, breach of the peace, or diversion tending to a breach of the peace." 394 US at 116 (Black concurring). See note 117 on the relationship between vagueness and improper motive.

¹³⁶ See text accompanying notes 49-51.

¹³⁷ See Geoffrey R. Stone, *Reflections on the First Amendment: The Evolution of the American Jurisprudence of Free Expression*, 131 Proc Am Phil Soc'y 251, 258 (1987).

¹³⁸ The fighting words doctrine constitutes an exception to the principle that statutes turning on communicative impact, though facially content neutral, receive heightened scrutiny because they delegate to the public the ability to restrict speech based on content. I consider in Section III.C.1 why fighting words laws receive minimal scrutiny.

tent neutral on their face, allow content-based actions in application, whether by enforcement officials or the public. The second point is that content-based action—whether on the face of a law or as it applies—raises the fear of improper motive. I now turn to one last case in which that fear comes into play, notwithstanding a statute's facial neutrality.

3. Laws "equalizing" the speech market.

In what has become one of the most castigated passages in modern First Amendment case law, the Court pronounced in *Buckley v Valeo* that "the concept that government may restrict the speech of some elements of our society in order to enhance the relative voice of others is wholly foreign to the First Amendment . . ." ¹³⁹ The Court made this statement in the course of invalidating expenditure ceilings in a campaign finance law, one justification for which was that they "equaliz[ed] the relative ability of individuals and groups to influence the outcome of elections." ¹⁴⁰ But the statement has ramifications far beyond the area of campaign finance. It applies as well to a wide variety of schemes designed to promote balance or diversity of opinion. So, for example, the statement could have explained the Court's earlier holding in *Miami Herald Publishing Co. v Tornillo* disapproving a "right of reply" statute that required a newspaper to grant space in its pages to any political candidate whose record the newspaper had criticized. ¹⁴¹ So too the statement could summarize the view of the four dissenting Justices in *Turner Broadcasting System, Inc. v FCC* that "must-carry" rules, requiring cable operators to provide access to local broadcast television stations, violate the First Amendment. ¹⁴²

¹³⁹ 424 US 1, 48-49 (1976). Not surprisingly, scholars who advocate a model of the First Amendment focusing on the quality of public debate have attacked the Court's statement most harshly. See Sunstein, *Democracy and the Problem of Free Speech* at 94-101 (cited in note 19); Fiss, 71 Iowa L Rev at 1423-25 (cited in note 97).

¹⁴⁰ *Buckley*, 424 US at 48. The Court also has refused to accept this justification (often quoting *Buckley*) in other cases involving campaign finance regulations. See, for example, *First National Bank v Bellotti*, 435 US 765, 790-91 (1978) (invalidating a statute restricting corporate contributions and expenditures in referendum campaigns); *Citizens Against Rent Control v City of Berkeley*, 454 US 290, 295-96 (1981) (invalidating an ordinance limiting contributions to committees taking positions on referenda).

¹⁴¹ 418 US 241 (1974).

¹⁴² 114 S Ct 2445, 2475-81 (1994) (O'Connor concurring in part and dissenting in part). It is possible to argue that the statement in *Buckley* applies to an expenditure limit, but not to a right-of-reply statute or a must-carry rule, because the former is a direct restriction on speech, whereas the latter are merely rules regulating access to expressive

The Court's commitment to the principle of *Buckley* has lapsed on some occasions. After the Court invalidated *Buckley's* expenditure ceilings (which applied to individuals and groups alike), it accepted as an adequate justification for a statute limiting a corporation's political expenditures that corporate wealth could cause "distortion" and "unfairly influence elections."¹⁴³ And a few years before the Court struck down *Tornillo's* right-of-reply statute, the Court approved the FCC's fairness doctrine, which operated in a similar manner and arose from identical concerns relating to diversity and balance.¹⁴⁴ Finally, a slim majority in *Turner* upheld the must-carry rules imposed on cable operators—although, notably, by refusing to view the rules as an attempt to achieve an appropriate mix of ideas and information.¹⁴⁵ Still, *Buckley's* antiredistribution principle has con-

channels. But this proposed distinction collapses on reflection. The so-called access regulations at issue in *Tornillo* and *Turner* prevent the speaker from using the space transferred to others for her own preferred expression and thus may be said to constitute a direct restriction. Conversely, the so-called direct restriction in *Buckley* operates to provide others with greater access to the sphere of campaign expression, while allowing the speaker some continuing ability to express herself, and thus may be said to constitute an access regulation. I do not wish to say that all access rules are like direct restrictions and vice versa. But wherever a given resource is limited, it does no good to try to draw too sharp a line between them; providing access will restrict speech and restricting speech will provide access. The rationale in *Buckley* applies to either.

¹⁴³ *Austin v Michigan Chamber of Commerce*, 494 US 652, 660 (1990). The Court tried to distinguish *Austin* from *Buckley*, principally on the ground that corporate wealth derives from privileges bestowed on corporations by the government. *Id.* But this argument fails, because individual wealth also derives from governmental action. What the Court recognized in *Austin* is only what is true in every case: direct regulation of speech occurs against a backdrop of law that, while not referring to speech, goes far toward structuring the sphere of public expression. The question in every case is whether the government may use direct regulation of speech to redress prior imbalances.

¹⁴⁴ *Red Lion Broadcasting Co. v FCC*, 395 US 367 (1969). The Court in *Tornillo* never mentioned *Red Lion*, but presumably believed the case did not apply because it involved broadcasting, rather than the print media. The *Red Lion* Court had reasoned that different First Amendment standards govern broadcasting because of the (then existing) scarcity of radio frequencies. 395 US at 386-88.

¹⁴⁵ The Court interpreted the must-carry rules as an attempt "not to favor programming of a particular subject matter, viewpoint, or format, but rather to preserve access to free television programming for the 40 percent of Americans without cable." *Turner*, 114 S Ct at 2461. Congress's concern, according to the Court, was that cable operators, in the absence of must-carry provisions, would refuse carriage to local broadcast stations and that this refusal would destroy the viability of free broadcast television. The dissent objected to this reading of the must-carry rules, arguing that they instead were an attempt to promote diversity of information on the airwaves, by favoring speakers who would provide significant amounts of educational, local news, and public affairs programming. *Id.* at 2476-77 (O'Connor concurring in part and dissenting in part). What matters here is that the Court understood the must-carry rules in such a way as to obviate any conflict between those rules and the *Buckley* principle. If the rules were not designed to improve the content of the speech market, then they could not conflict with a prohibition

tinuing importance: no Justice on the current Court would dispute the claim—even if some Justices would dispute applications of it—that the government may not restrict the speech of some to enhance the speech of others.¹⁴⁶ The question that remains concerns the basis of this principle: what view of the First Amendment accounts for the Court's refusal to allow, by means of restrictions, the redistribution of expression?

The audience-based conception of the First Amendment does not; indeed, that conception calls for the converse of the *Buckley* principle. If what is essential, to recall Meiklejohn's phrase, is that "everything worth saying shall be said,"¹⁴⁷ then the First Amendment often would permit—indeed require—the reallocation of speech opportunities. The realm of public expression may have too much of some kinds of speech, too little of others; some speakers may drown out or dominate their opposite numbers. Self-conscious redistribution of expressive opportunities seems the most direct way of correcting these defects and achieving the appropriate range and balance of viewpoint.

Neither does the speaker-based conception of the First Amendment explain the *Buckley* principle.¹⁴⁸ The speaker-based model counsels in every case a weighing of the scope of the restriction on speech against the importance of the asserted state interest. Under this analysis, the Court in *Buckley* would have asked whether the interest in promoting diversity of opinion outweighed the loss of expression resulting from expenditure ceilings. But the Court in *Buckley* followed no such analysis when it considered the government's redistributive interest.¹⁴⁹ The

on the government's undertaking such a project.

¹⁴⁶ Although the matter is not free from doubt, the antiredistribution principle of *Buckley* may well be subject to a "necessity" defense, allowing the Court to uphold redistributive speech policies upon a showing of need similar to that required for other suspect regulations of speech, such as content-based restrictions. The Court suggested as much in *Bellotti* when it said that if arguments relating to the "undue influence" of corporate speech "were supported by record or legislative findings that corporate advocacy threatened imminently to undermine democratic processes, thereby denigrating rather than serving First Amendment interests, these arguments would merit our consideration." 435 US at 789. Perhaps, too, the Court's decision in *Red Lion* emerged from the view that the fairness doctrine satisfied a test of necessity given the then existing scarcity of broadcast frequencies. See Geoffrey R. Stone, *Autonomy and Distrust*, 64 U Colo L Rev 1171, 1178 (1993). In any event, regulations thought suspect on the ground of their redistributive justification probably face a rebuttable, rather than an irrebuttable, presumption against them.

¹⁴⁷ Meiklejohn, *Political Freedom* at 26 (cited in note 15).

¹⁴⁸ Owen Fiss has suggested that the *Buckley* principle emerges from the speaker-based model, which he calls the autonomy principle. See Fiss, 71 Iowa L Rev at 1423 (cited in note 97). I disagree with this view of the *Buckley* principle's underpinnings.

¹⁴⁹ The Court did use this analysis in evaluating the distinct governmental interest in

Court said not that this interest was insufficient, but that its very assertion conflicted with fundamental premises of the First Amendment. No matter how little the ceilings limited expression or how much they promoted diversity of opinion, the Court could not countenance their existence. Such an analysis fits poorly with the speaker-based model because it rests not on the extent of, but on the justification for, the restriction.

The motive-based model of the First Amendment, last left standing though it is, appears at first to explain the *Buckley* principle no better than the others. It is true that *Buckley* is all about reasons; the decision bars the government from asserting or relying on an interest in redistribution to support a limit on expression. But the interest that *Buckley* declares improper is not congruent with—indeed, is almost the converse of—the concept of illicit motive I have developed. The justification barred in *Buckley*, far from relating to the decision maker's own ideology or self-interest, relates to her desire to construct a speech market in which citizens can choose, in an informed way, among many competing ideas, benefiting many different interests. How, then, does the *Buckley* principle fit with the concept of improper motive prevalent in the rest of First Amendment doctrine?

The answer to this question involves viewing the *Buckley* principle as an evidentiary tool designed to aid in the search for improper motive, much like the presumption against content-based restrictions. The *Buckley* principle emerges not from the view that redistribution of speech opportunities is itself an illegitimate end, but from the view that governmental actions justified as redistributive devices often (though not always) stem partly from hostility or sympathy toward ideas—or, even more commonly, from self-interest.¹⁵⁰ The *Buckley* principle thus serves a function of proof—once again, to uncover, in the absence of a

preventing corruption. There, the Court held, as an outgrowth of balancing, that the interest was "inadequate to justify" the restriction. *Buckley*, 424 US at 45. That approach is consistent with a speaker-based model. It also may represent an attempt, as with review of any content-neutral law, to discover the slight chance of impermissible motive.

¹⁵⁰ This interpretation of the *Buckley* principle explains why the Court—as suggested by the holding in *Red Lion*, 395 US at 367, and dictum in *Bellotti*, 435 US at 789—will accept a redistributive justification on a showing of necessity. See note 146. The ability to rebut the presumption of unconstitutionality flowing from a redistributive justification by showing that redistribution is desperately needed indicates that pursuing the goal of diversity is not itself illegitimate; the presumption of unconstitutionality must attach to the redistributive justification because that justification tends to be linked (in a way that a showing of need will disprove) to other, impermissible ends of government.

direct method for doing so and in a necessarily inexact manner, improperly motivated governmental actions.

When laws justified as redistributive devices draw content-based lines, the evidentiary function of the *Buckley* principle becomes clear, though also superfluous. Assume Congress enacts a law limiting speech in support of the Republican Party, on the ground that speech of this kind drowns out speech promoting other political views. Or assume Congress, to take a serious proposal, passes a law limiting sexually explicit speech promoting the subordination of women, on the ground that such speech dominates existing avenues of communication and, indeed, silences the speech of those opposed to it. In both cases, the Court might invoke the *Buckley* principle; and in both cases that principle would serve the same motive-hunting function as the (also applicable) presumption against content-based action. The justifications offered for these laws might be valid. But such laws, like all other content-based restrictions, pose a significant danger of arising from the different desire on the part of officials to insulate themselves or their ideas from challenge. In these cases, to state the slogan of *Buckley* is only to say that when officials restrict a particular idea—with the ostensible goal of increasing diversity, as with any other proper aim—the Court will not trust them to fend off improper considerations of hostility, sympathy, or self-interest.

The question remains, however, why the Court should treat as especially suspicious content-neutral regulations of speech—such as the regulations in *Buckley*—that are justified in terms of achieving diversity.¹⁵¹ If the *Buckley* principle serves

¹⁵¹ It is often difficult to classify schemes justified in terms of diversity as content based or content neutral. *Buckley* provides an example. I treat the expenditure ceilings in *Buckley* as content neutral, as is the norm, because they cover spending for expression supporting all political candidates. See Tribe, *American Constitutional Law* § 13-27 at 1132-36 (cited in note 67). It could be argued, however, that the ceilings imposed a subject-matter limitation because they applied only to spending in political campaigns. A similar question arises with respect to the right-of-reply laws in *Red Lion* and *Tornillo*, which also applied equally to all sides of a subject, but only to subjects of a certain kind. And cases involving diversity schemes often raise the problem of how to categorize so-called speaker-based restrictions on expression. The must-carry rules in *Turner*—like the corporate speech restrictions in *Bellotti* and *Austin*—primarily distinguished on the basis of the identity of the speaker, rather than the content of speech. Particular speakers, though, tend to say particular things; local broadcast stations, for example, usually carry programming on local issues. These facially speaker-based provisions thus raise the question—inadequately addressed in *Turner*—whether and when the Court should associate a given speaker with a given idea, thus treating the regulation as content based. See generally Stone, 25 *Wm & Mary L Rev* at 244-51 (cited in note 15). Of course, if the Court

as a means for flushing out illicit motive, then the answer must relate to some special characteristic(s) of these regulations that affect the motive inquiry. And in fact, the nature of these regulations, as compared with other content-neutral regulations, creates two problems (similar to those posed by standardless licensing schemes): first, that governmental officials (here, legislators) more often will take account of improper factors, and second, that courts will have greater difficulty detecting the presence of such tainted deliberations.

The increased probability of taint arises, most fundamentally, from the very design of laws directed at equalizing the realm of public expression. Unlike most content-neutral regulations, these laws not only have, but are supposed to have, content-based effects; their *raison d'être* is to alter the mix of ideas—or, at least, of speakers, who tend to be associated with ideas—in the speech market. Given this function, these laws will have not the diverse, diffuse, and crosscutting content-based effects usually associated with content-neutral laws, but a set of targeted and coherent effects on ideas and speakers. This set of focused effects renders a law directed at equalization nearly as likely as a facially content-based law, and much more likely than most facially content-neutral laws, to stem from improper motive. In considering such a law, a legislator's own views of the ideas (or speakers) that the equalization effort means to suppress or promote may well intrude, consciously or not, on her decision-making process. The law thus raises grounds for suspicion.

The likelihood of impermissible taint increases further because of the nature of the harm claimed to justify the action. This harm—a disproportionate access or undue influence—has an amorphous aspect to it that often (though not always) will make it peculiarly difficult to measure.¹⁵² This does not mean that legislators cannot debate rationally the size of the harm and cannot reach what appears to be—and in fact might be—a reasonable decision to counteract it. But because the harm has a fuzzy quality, the decision whether the harm justifies a limit on speech becomes especially susceptible to the infiltration of illicit factors. It is the rare person who can determine whether there is

treats identically all restrictions of speech justified in terms of diversity, regardless whether they are content based or content neutral, the Court will never have to confront these labeling issues. The question on the table is whether there is any other reason for providing the same treatment to all such restrictions.

¹⁵² See text accompanying note 100.

“too much” of some speech (or speakers), “too little” of other speech (or speakers), without any regard to whether she agrees or disagrees with—or whether her own position is helped or hurt by—the speech (or speakers) in question.

Indeed, all the laws directed at equalization that the Court has considered, whether classified as facially content based or content neutral, raise questions as to the motives of the enacting legislatures. Campaign finance laws like those in *Buckley* easily can serve as incumbent-protection devices, insulating current officeholders from challenge and criticism.¹⁵³ When such laws apply only to certain speakers or subjects, the danger of illicit motive becomes even greater; for example, the law in *First National Bank v Bellotti*,¹⁵⁴ which barred corporations from spending money in referendum campaigns, almost surely arose from the historic role of corporate expenditures in defeating referenda on taxation.¹⁵⁵ Similarly, a right-of-reply law like the one in *Tornillo*—applicable only to political candidates, albeit to all of them—may have stemmed not from the desire of officials to enhance the quality of public debate, but from their wish to get the last word whenever criticized.¹⁵⁶ If this law did not quite prohibit seditious libel, it came close. And the must-carry rules in *Turner* may have emerged more from the yen of politicians for local publicity than from their wish either to preserve free television or to enhance public discourse.¹⁵⁷ All these examples suggest the same point: there may be good reason to distrust the

¹⁵³ See, for example, Lillian R. BeVier, *Money and Politics: A Perspective on the First Amendment and Campaign Finance Reform*, 73 Cal L Rev 1045, 1080 (1985); *Austin*, 494 US at 692-93 (Scalia dissenting) (suggesting that “with evenly balanced speech incumbent officeholders generally win”).

¹⁵⁴ 435 US 765 (1978).

¹⁵⁵ See *id.* at 826-27 n 6 (Rehnquist dissenting). Even in voting to uphold the law, then Associate Justice Rehnquist admitted: “[A] very persuasive argument could be made that the General Court, desiring to impose a personal income tax but more than once defeated in that desire by . . . corporate expenditures in opposition to such a tax, simply decided to muzzle corporations on this sort of issue so that it could succeed in its desire.” *Id.* See also *Austin*, 494 US at 692 (Scalia dissenting) (noting that a limit on corporate expenditures in support of political candidates might have been “trying to give unincorporated unions . . . political advantage over major employers”).

¹⁵⁶ The fairness doctrine upheld in *Red Lion* included a similar provision allowing candidates a right to reply to any editorial endorsing an opponent. This provision, however, formed only part of a more general policy allowing any person or group to respond to personal attacks and requiring coverage of all sides of public issues. See *Red Lion*, 395 US at 373-75. The greater generality of this law, as compared with the one in *Tornillo*, suggests a lesser reason to distrust the legislative process.

¹⁵⁷ Local broadcast stations, which are the stations cable operators must carry, usually provide greater coverage of local politicians than other programmers do.

motives of politicians when they apply themselves to reconstructing the realm of expression.

The problem is compounded by the difficulty of detecting, through use of the standard of review applied to content-neutral laws, the presence of improper taint in statutes ostensibly aimed at equalizing expression.¹⁵⁸ The unusual impotence of the standard in this context relates to the absence of any clear criteria for deciding what state of public debate constitutes the ideal and how far current discourse diverges from it. If a court cannot make these fine judgments, then it also cannot conclude that a law supposedly enacted to ease distortion in fact fails to advance this interest or to do so in a narrowly tailored manner.¹⁵⁹ In almost any case, the government can defend a statute on the ground that it alleviates distortion; and in almost no such case—even when the asserted purpose is a pretext—will the standard of review applied to content-neutral laws command the statute's invalidation. The usual motive-hunting mechanism fails to operate.

The ease with which improper purpose can taint a law directed at equalizing expression together with the difficulty a reviewing court will have in detecting this taint account for the Court's approach in cases like *Buckley*. The reason for the approach is not that the goal of equalization itself conflicts with the First Amendment (though the Court often speaks in this manner); to the contrary, when Justice Scalia labeled the objective of equalization "unqualifiedly noble,"¹⁶⁰ he may have used the term tongue-in-cheek, but he also used it aptly. The reason for the approach is instead that the goal of equalization often and well conceals what does conflict with the First Amendment: the passage of laws tainted with ideological, and especially with self-interested, motivations. Again in Justice Scalia's words: "The incumbent politician who says he welcomes full and fair debate is

¹⁵⁸ As noted earlier, this standard usually works to flush out content-neutral laws tainted by illicit purpose. See text accompanying notes 108-09.

¹⁵⁹ See *BeVier*, 73 Cal L Rev at 1090 (cited in note 153). Yet another problem in testing the fit of a statute of this kind arises from the fact that the governmental interest at stake relates exclusively to speech. As I discuss in later Sections, when legislators can advance an interest both by regulations restricting speech and by regulations restricting nonexpressive conduct, courts can explore the bona fides of a limitation on speech by considering whether and how the government has tried to promote the interest in non-speech-related ways. But because the interest in equalizing the speech market is about nothing other than speech, this technique is unavailable to courts in the cases I have been discussing, thus further impeding the search for impermissible motive.

¹⁶⁰ *Austin*, 494 US at 692 (Scalia dissenting).

no more to be believed than the entrenched monopolist who says he welcomes full and fair competition.”¹⁶¹ The harsh treatment of laws directed at correcting distortion, even when these laws are framed in content-neutral language, arises from the fear that if the usual standards of review applied, legislators would use these laws as a vehicle for improper motive, and courts would bless what the First Amendment proscribes.

Laws directed at equalizing speech thus join the list of laws that, although facially content neutral, demand strict scrutiny because of heightened concerns relating to improper purpose. In all these cases, the laws in some way rest on or allow decisions based on content, though no reference to content appears on their face. In so doing, the laws raise the same suspicions of improper purpose as do facially content-based laws, and for much the same reasons. The Court thus treats these laws in a strict manner—presuming improper taint, though giving the government a chance to rebut this presumption. The doctrine, again, responds to the desire to discover impermissible purpose—and to the necessity of relying on overinclusive and underinclusive categorical rules to accomplish this object.

C. “Safe” Content-Based Laws

If some facially content-neutral laws pose a serious risk of stemming from improper purpose, perhaps the converse is also true: that some facially content-based laws pose only a slight danger of taint, so that weakening the presumption against them becomes appropriate. In this Section, I look at First Amendment law’s two great exceptions to the rule against content-based restraints—the recognition of low-value categories of speech and the doctrine of secondary effects—and ask whether a motive-based approach can explain them. I conclude that this approach, although not answering all questions in these two contexts, provides the most coherent general account of prevailing doctrine.

1. Low-value speech.

First Amendment law is replete with content distinctions that do not count as content distinctions because they disfavor speech found by the Court to have little (or no) constitutional value and thus to receive little (or no) constitutional protection.

¹⁶¹ *Id.*

The government may ban obscenity or child pornography wholesale without implicating First Amendment principles.¹⁶² So too for fighting words, as the *R.A.V.* Court reiterated.¹⁶³ Less simply, the regulation of commercial speech receives minimal constitutional scrutiny if the speech is false or misleading, intermediate scrutiny in all other cases.¹⁶⁴ Least simply, libel law is subject to a bewildering variety of constitutional standards, which (roughly speaking) become less exacting as the plaintiff and the subject matter become less “public.”¹⁶⁵ All of these disfavored categories are based on the content of speech; some, at least arguably, are based on its viewpoint.¹⁶⁶

These “low-value” categories have a flipside—a category of converse content, though of a more informal nature, which may place them in valuable perspective. The Court sometimes has stated that the First Amendment protects most strongly speech

¹⁶² See *Miller v California*, 413 US 15, 23 (1973) (obscenity); *New York v Ferber*, 458 US 747, 758 (1982) (child pornography). In addition, the Court has indicated that other sexually explicit speech is entitled to less than a full measure of constitutional protection, although the precise standard is uncertain. See *FCC v Pacifica Foundation*, 438 US 726, 744-48 (1978); *Young v American Mini-Theatres, Inc.*, 427 US 50, 62-63 (1976); *City of Renton v Playtime Theatres, Inc.*, 475 US 41, 48-50 (1986).

¹⁶³ The fighting words doctrine originated in *Chaplinsky v New Hampshire*, 315 US 568 (1942); although in recent years the doctrine's continued vitality seemed in doubt, the Court in *R.A.V.*, 505 US at 382-83, treated it as fully settled law. From the beginning, the Court defined the category of fighting words not by reference to specific content, but by reference to the typical addressee's response; as argued previously, however, this difference should have no operational significance, given that an audience's response to speech itself turns on content. See text accompanying notes 136-38.

¹⁶⁴ See *Central Hudson Gas & Elec. Corp v Public Serv Comm'n*, 447 US 557, 562-64 (1980).

¹⁶⁵ See *Gertz v Robert Welch, Inc.*, 418 US 323, 342-43 (1974); *Dun & Bradstreet, Inc. v Greenmoss Builders, Inc.*, 472 US 749, 760-61 (1985) (plurality opinion). Many complexities to one side, public officials and figures must prove actual malice on the part of defendants to recover damages; private figures complaining about a statement of public concern must show negligence to recover compensatory damages and actual malice to recover punitive and presumed damages; and private figures complaining about a statement of purely private concern need show only what state law requires for punitive and presumed damages and probably for compensatory damages as well.

¹⁶⁶ Catharine MacKinnon has argued, for example, that the categories of obscenity and child pornography reflect a kind of viewpoint discrimination because the speech falling within these categories likely expresses a single (disfavored) viewpoint about sexual matters. See Catharine A. MacKinnon, *Feminism Unmodified: Discourses on Life and Law* 212 (Harvard 1987). Cass Sunstein has strengthened the argument by noting that the test of obscenity invokes community standards of offensiveness. See Cass R. Sunstein, *Neutrality in Constitutional Law (With Special Reference to Pornography, Abortion, and Surrogacy)*, 92 Colum L Rev 1, 28-29 (1992). A similar argument of viewpoint bias might apply to the fighting words category, given that it is often the viewpoint (and not merely the generalized “content”) of speech that will prompt an average addressee to fight.

on political matters; it is in this area of expression, in the words of the Court, that the First Amendment achieves its “fullest and most urgent application.”¹⁶⁷ Statements of this kind, it might be argued, have had no real import. As a rule, no separate standard reigns for speech on political affairs—or even, to dilute the category a bit, speech on matters of public interest.¹⁶⁸ The Court’s pronouncements thus may have served only a packaging function: to provide added (but unnecessary) support for decisions that would have been reached regardless. But there is reason to think, contrary to this interpretation, that the political character of expression in fact affects the Court’s decisions, even if in a more subtle fashion than by raising the formal constitutional standard. In recent decades, the Court almost never has upheld a regulation of political speech. Perhaps more tellingly, almost all of the landmarks of First Amendment law—the classic cases that set the tone and provide the focus for analysis of free speech questions—arise out of governmental attempts to restrict speech of an obviously political nature.¹⁶⁹

¹⁶⁷ *Monitor Patriot Co. v. Roy*, 401 US 265, 272 (1971); *Minneapolis Star & Tribune Co. v. Minnesota Comm’r of Revenue*, 460 US 575, 585 (1983). See also *Bellotti*, 435 US at 776-77; *Buckley*, 424 US at 14-15; *Mills v. Alabama*, 384 US 214, 218 (1966); John Paul Stevens, *The Freedom of Speech*, 102 Yale L J 1293, 1306-07 (1993).

¹⁶⁸ The law on the speech of public employees provides one exception to this principle. The Court has ruled that the First Amendment prohibits discharging an employee because of her speech only when the speech relates to “matters of public concern.” See *Connick v. Myers*, 461 US 138, 142 (1983). The Court’s libel jurisprudence also has incorporated, in an on-again-off-again way, a separate standard for speech on matters of public interest. At one time, all (but only) statements of “general or public interest” received the highest level of constitutional protection from the prosecution of libel actions. *Rosenbloom v. Metromedia, Inc.*, 403 US 29, 48-49 & n 17 (1971) (plurality opinion). The Court in *Gertz* rejected this approach in favor of a method that makes the level of protection dependent on the nature of the plaintiff rather than the subject matter of the speech. The question whether the plaintiff is a public or a private figure, however, partly depends on whether the allegedly libelous speech stemmed from her participation in a “public controversy”—a decision scarcely different from the decision whether the allegedly libelous speech concerned a public issue. See *Gertz*, 418 US at 343-46. Moreover, in a later case, the Court reintroduced such an inquiry by dividing libel suits brought by private figures between those involving matters of “purely private concern” and those involving other subjects. See *Dun & Bradstreet*, 472 US at 758-60 (plurality opinion). See generally Cynthia L. Estlund, *Speech on Matters of Public Concern: The Perils of an Emerging First Amendment Category*, 59 Geo Wash L Rev 1 (1990).

¹⁶⁹ The list is a long one, including the great dissenting opinions of Holmes in *Abrams v. United States*, 250 US 616, 624 (1919), and *Gitlow v. New York*, 268 US 652, 672 (1925), the equally great concurring opinion of Brandeis in *Whitney v. California*, 274 US 357, 372 (1927), and the majority opinions in *Near v. Minnesota*, 283 US 697 (1931), *New York Times Co. v. Sullivan*, 376 US 254 (1964), *Brandenburg v. Ohio*, 395 US 444 (1969), *Cohen v. California*, 403 US 15 (1971), the Pentagon Papers case, otherwise known as *New York Times Co. v. United States*, 403 US 713 (1971), and the flag burning cases of *Texas v. Johnson*, 491 US 397 (1989), and *United States v. Eichman*, 496 US 310 (1990). For a

This schema—the formal delineation of low-value categories on the one hand, the informal elevation of political speech on the other—plausibly derives from more than one model of the First Amendment. Although the speaker-based conception fits awkwardly with the doctrine, both the audience-based model and the government- or motive-based model substantially, though not entirely, support it. Thus, this aspect of First Amendment law, unlike others I have noted, is not solely explicable as a tool for discovering improper motive. But as I explain, the audience-based account of low-value categories creates conflict—or at least tension—between the recognition of those categories and other fundamental aspects of First Amendment law. In contrast, the motive-based account of these categories renders them harmonious with the rest of the doctrine.

A speaker-based approach could explain low-value categories only if speech of the disfavored kinds confers less of value on a speaker than does speech receiving full protection. Low-value speech must promote autonomy or self-realization or self-development—the values, variously conceived and phrased, that a speaker gains from communicating—less well than does other speech. Further, political speech, to explain the flipside of the doctrine, must promote these values better.

This account of the doctrine seems peculiar. It is, after all, not clear what kind of speech does the greatest good for speakers, or best promotes their interests; it is not even clear what criteria we would use to think about this question.¹⁷⁰ One scholar has called the speech in *Chaplinsky v New Hampshire*¹⁷¹ “a personal catharsis, . . . a means to vent his frustration[,] . . . a significant means of self-realization.”¹⁷² Another proponent of the speaker-based model has written that “materials that most people view as pornographic may play an important role in some people’s self-fulfillment and self-expression.”¹⁷³ Given the uncer-

popular history of First Amendment doctrine that focuses on these cases, weaving them into a coherent First Amendment tradition, see Anthony Lewis, *Make No Law: The Sullivan Case and the First Amendment* (Random House 1991).

¹⁷⁰ See Sunstein, *Democracy and the Problem of Free Speech* at 141 (cited in note 19) (“If we protect speech because people want to talk, it is not easy to come up with standards by which to distinguish among different kinds of talk.”).

¹⁷¹ 315 US 568 (1942).

¹⁷² Martin H. Redish, *Freedom of Expression: A Critical Analysis* 56 (Michie 1984).

¹⁷³ C. Edwin Baker, *Human Liberty and Freedom of Speech* 69 (Oxford 1989). The same scholar has argued that commercial speech ought not to receive constitutional protection because such speech “reflects market forces” rather than manifests individual choice. *Id.* at 196. These are the kinds of distinctions that would have to be made in

tainty surrounding these issues, the most appropriate course would place in the speaker's own hands the question what kind of speech has value to her, by freeing her to choose among expressive activities.¹⁷⁴ Indeed, to the extent governmental officials influence this choice, they may stifle the very aspect of expression—the speaker's decision as to what kind of speech has greatest value to her—most closely related to self-realization. But even placing this claim to one side, it seems a dubious project to expropriate, on grounds only of the speaker's own interests, the speaker's decision, expressed through the act of speech itself, as to what forms of speech have value to her.

An audience-based perspective does better in accounting for First Amendment doctrine's categorization of speech according to value. If the goal of a free speech system is to provide individuals (especially in their roles as citizens) with the range of opinion and information that will enable them to arrive at truth and make wise decisions, then a tiered system of speech, of the kind the Court has created, seems appropriate. Some speech does not enrich (may even impoverish) the sphere of public discourse. Other speech contributes to reasoned deliberation on matters of public import. Under the audience-based approach, it would be perverse to treat these disparate forms of speech identically. Thus emerges a multitiered system.

And thus might emerge *this* multitiered system, which represents a plausible statement of the contribution of different kinds of speech toward creating a healthy sphere of discourse. The extreme respect shown to political speech comports with the strand of the audience-based model that highlights the need to give individuals the information required to fulfill the role of sovereign citizen. The treatment of libel rests on the premise that false statements of fact have no value because they prevent listeners from gaining accurate information; the varying standards applied to different kinds of libel suits then reflect roughly the value of the true speech that these suits chill.¹⁷⁵ The treatment

developing a tiered First Amendment through a speaker-based model.

¹⁷⁴ This does not mean that courts may not restrict the speech so chosen; they may do so if the speech poses the requisite danger of harm. But under a speaker-based approach, for the reasons stated in the text, courts usually should not assign varying values to varying forms of expression, so that disparate showings of harm are required.

¹⁷⁵ The correlation here is admittedly crude. If libel standards directly reflected the value of speech to public discourse, they would vary with the subject matter of the speech, rather than with the character of the plaintiff. One scholar who favors an audience-based approach proposes this reform of the law of defamation. See Sunstein, *Democracy and the Problem of Free Speech* at 159-62 (cited in note 19). As noted previously, however, the law

of false and misleading commercial speech arises from a similar sense that speech that operates through deception cannot serve audience interests. And the exclusion of obscenity and fighting words reflects a judgment that these forms of speech are themselves not "reasoned" and thus cannot aid reasoned discourse.

Some aspects of the current multitiered system, however, fit oddly with the audience-based model. The near absolute protection given to false but nondefamatory statements of fact outside the commercial realm is peculiar from this perspective; even a concern with chilling true speech would not explain such sweeping protection of speech that disserves understanding. Conversely, the reduced protection granted to truthful commercial speech, when compared with speech not only on politics but on other subjects, raises questions; to justify this treatment, the audience-based model needs a plausible theory of why learning about commercial matters and making sound commercial decisions has only insubstantial value to the public. Even the noncoverage of fighting words and obscenity is less than certain under this model. Perhaps personal invective adds something to public dialogue, precisely because of its earthy quality. Or perhaps some materials labeled obscene enrich understanding of sexual matters.

Even to the extent that the audience-based model makes sense of low-value categories, it has trouble bringing this doctrinal feature into line with the rest of First Amendment law. Many scholars have objected on this score to the creation and use of low-value categories. In the words of one, commenting on the reasoning of *Chaplinsky*, the inquiry whether certain kinds of speech are of "slight social value as a step to truth" compels the Court to make "value judgments concerned with the content of expression, a role foreclosed to it by the basic theory of the First Amendment."¹⁷⁶ Or, in the words of another, "the very concept of low-value speech is an embarrassment to first amendment orthodoxy. To say that government cannot suppress speech unless the speech is of low value sounds like a parody of free speech theory."¹⁷⁷ "Embarrassment" or "parody" may be too strong, but

of defamation already hinges in important ways on the subject matter of the allegedly libelous speech. See note 168. Moreover, even the inquiry as to whether a plaintiff is a public or private figure may serve as a rough proxy for the more amorphous question whether the case involves speech on matters of public interest. See Elena Kagan, *A Libel Story: Sullivan Then and Now*, 18 L & Soc Inquiry 197, 214-15 & n 39 (1993).

¹⁷⁶ Thomas I. Emerson, *The System of Freedom of Expression* 326 (Random House 1970), quoting *Chaplinsky*, 315 US at 572.

¹⁷⁷ Steven H. Shiffrin, *The First Amendment, Democracy, and Romance* 44 (Harvard

these categories do present a puzzle. Much of the law I have discussed operates to prevent the government from elevating certain forms of speech over others on the basis of their relative value. The government usually cannot say that one kind of speech, as compared with another, is more (or less) worthy, useful, or important. Why, then, may the government say exactly this of fighting words or obscenity? The dearth of public value in such speech may explain the presence of low-value categories, but only by creating a tension within the doctrine of free expression.

A motive-based approach, in contrast, could use a multitiered system in a way consonant with the other central features of First Amendment law. Under this approach, the government would have unusual leeway over so-called low-value speech not because the speech in fact has low value, but because regulation of the speech has a low probability of stemming from illicit motive.¹⁷⁸ If a regulation of a certain kind of speech carries a reduced suspicion of taint, the Court should adopt a standard of review that places a reduced burden of justification on the government; the Court thus should lower the usual strong presumption against regulation or even switch the presumption in the opposite direction. If such a regulation poses no, or almost no, concern of impermissible motive, the Court should allow the legislature to regulate at will, adopting an irrebuttable presumption that the regulation comports with the First Amendment. In this way, so-called low-value categories of speech would appear, but these categories now would fit into, rather than obtrude from, the overall panorama of First Amendment doctrine.

This understanding of a tiered system of expression plausibly explains many, though not all, of the content-based categories of speech that the Court has created. At one end of the spectrum, the regulation of speech about political issues poses the greatest risk of stemming from improper purpose.¹⁷⁹ Political speech often implicates the self-interest of governmental officials; likewise

1990).

¹⁷⁸ Cass Sunstein's proposed four-part test of low-value speech includes the inquiry whether "government is unlikely to be acting for constitutionally impermissible reasons . . ." Cass R. Sunstein, *Pornography and the First Amendment*, 1986 Duke L J 589, 604. This matches the approach I am taking; it has nothing to do with the "value" of low-value speech, either to the speaker or to the audience. Sunstein's other three factors do relate to such value. See *id.* at 603-04.

¹⁷⁹ See Thomas I. Emerson, *Toward A General Theory of the First Amendment* 9-10 (Random House 1966); Meiklejohn, *Political Freedom* at 25-27 (cited in note 15); T.M. Scanlon, Jr., *Freedom of Expression and Categories of Expression*, 40 U Pitt L Rev 519, 534-35 (1979).

(and perhaps synonymously), political speech often implicates their most strongly held views and opinions. Because this is so, courts view the regulation of political speech with special disfavor (even if not through a different formal standard), requiring the government to make an extraordinary showing to dissipate the suspicion of improper motive.

At the other end of the spectrum, the regulation of speech falling within low-value categories often raises fewer concerns than usual about improper purpose. The Court's defamation law illustrates the point. The First Amendment provides greatest protection in suits brought by public officials and figures—suits in which the speech most likely addresses matters of public concern and thus most likely implicates the views and interests of decision makers.¹⁸⁰ The First Amendment provides least protection in suits brought by private figures concerning speech of purely private interest—suits in which improper factors are least likely to influence libel judgments.

Similarly, the treatment of commercial speech may respond to the chance that improper motive will taint regulation; in this sphere, the Court may believe, the government less often acts for self-interested or ideological reasons.¹⁸¹ The government, after all, daily regulates commercial transactions; and in reviewing this regulation, the Court presumes the legitimacy of the government's reasons. A like presumption seems natural when courts review regulation of speech soliciting commercial transactions—itself a kind of commercial activity. Similar legitimate reasons support such regulation, and the danger of taint appears correspondingly slim. The Court thus lowers the burden placed on the government to demonstrate that regulation of commercial speech has a legitimate basis. Here, officials freely may regulate false and misleading expression—as they also may in the “safe” context of private defamation, but as they may not in the “unsafe” context of political discourse. And as to truthful commercial speech, where a legitimate reason for regulating is not so obvious, officials may act so long as they meet an intermediate stan-

¹⁸⁰ See Frederick Schauer, *Public Figures*, 25 *Wm & Mary L Rev* 905, 921-29 (1984). Schauer argues that fear of governmental overreaching explains the treatment of libel suits brought by public officials, but not by public figures. But if the inquiry into whether the plaintiff is a public figure represents a more easily administrable version of the inquiry into whether the speech concerns a matter of public interest, see note 175, then the fear that impermissible factors will influence judges or jurors also explains the cases involving public figures.

¹⁸¹ See Scanlon, 40 *U Pitt L Rev* at 541 (cited in note 179).

dard. The rigor of review again follows the danger that illicit motive has tainted a law.

The analysis proceeds in a slightly different way for categories of speech entitled to no protection. In creating these no-value categories, the Court may be stating an irrebuttable presumption that certain prohibitions arise from legitimate motives. Here, the Court in effect predecides that a ban on all speech within the category will result not from hostility toward the ideas restricted, nor from self-interest, but instead from a neutral decision that the harm the speech causes justifies prohibition. This predecision amounts to much the same thing as a post hoc decision in the usual case that a regulation meets the requisite standard of review. Both are conclusions about the absence of illicit motive; the difference is only that in the former case the constitutional review and the conclusion as to motive are submerged in the creation of an "unprotected" category. What the Court is saying is that any law framed in the way the Court has defined the category closely enough fits a sufficiently significant interest as to negate any worry of improper purpose.

This explanation fits the Court's creation of a fighting words category. Though hazy in its boundaries, this category seems to embrace direct face-to-face insults that would cause the average addressee to respond with violence. In holding that a legislature may prohibit fighting words, the Court is doing no more than approving a governmental response to an immediate danger of violence. The premise of the category, no less than of a clear-and-present-danger test, is that the government would respond to such a danger no matter what its views of the ideas affected.

A motive-based account, however, fits poorly if at all the Court's treatment of obscene speech. As an initial matter, the formal test for determining obscenity suggests that motive is not the key. That test mandates an inquiry into the value of the materials, which suggests that the concerns of the audience-based model here predominate. The test also demands a finding of community offense, which constitutes, rather than disproves, improper motive. And even if the formal test did not include these attributes, the probability of taint infecting an obscenity law seems severe. Such a law might stem from a neutral evaluation of the harm these materials cause, such as sexual violence. But hostility toward certain ideas about sexual mores—otherwise stated, the desire to maintain status quo ideas about sexuality free from challenge—are likely to color this evaluation or trump it entirely. (Consider how officials often respond to speech show-

ing or approving homosexuality.) In this area, a motive-based model thus fails to explain the doctrine.¹⁸²

But placing sexually explicit speech to one side, the term “low-value” category may be a misnomer. Perhaps what sets these categories apart is not that the speech within them is low value, but that regulation of the speech within them is low risk. No matter that a regulation of these categories is content based, even viewpoint based; the government need not satisfy the usual standard because the courts do not suspect, to the usual extent, that the government’s asserted legitimate interest is a pretext. The categories—call them now “low-risk categories”—thus function as yet another part of the broad evidentiary mechanism designed to flush out actions based on impermissible purpose.

This explanation makes the holding of *R.A.V.*—invalidating a partial prohibition of fighting words—consonant with the theory that underlies the fighting words category.¹⁸³ This category rests on a presumption that, given the state’s great and legitimate interest in maintaining order, a ban on fighting words stems from proper motive. The Court has no reason to apply this presumption when the government regulates only part of the category, on the basis of content (or viewpoint) extraneous to the category’s boundaries.¹⁸⁴ In that case, there is an underinclusion problem—a lack of fit between the regulation and

¹⁸² The motive-based model also may not explain the category of child pornography, though here the analysis is more complicated. The formal test for child pornography does not include any assessment of the material’s offensiveness or value. See *Ferber*, 458 US at 754-65. In addition, the harms caused by this material to children would seem amply to justify governmental action. But here too, motives are complex, and it would be surprising if disapproval of certain ideas about sexuality and children did not enter into the government’s calculation.

¹⁸³ See *Kagan*, 60 U Chi L Rev at 899-900 (cited in note 61).

¹⁸⁴ The situation would be different, as Justice Scalia noted in *R.A.V.*, if the government regulated only part of the category, but on the basis of the characteristics (and corresponding state interests) that define the category. 505 US at 387-88. Justice Scalia used the example of a law that prohibited only the most prurient of all obscenity. *Id* at 388. In his words:

When the basis for the content discrimination consists entirely of the very reason the entire class of speech at issue is proscribable, no significant danger of idea . . . discrimination exists. Such a reason, having been adjudged neutral enough to support exclusion of the entire class of speech from First Amendment protection, is also neutral enough to form the basis of distinction within the class.

Justice Scalia’s conclusion that the St. Paul law did not fit within the contours of this principle is open to question. See *id* at 403-04 (White concurring); *id* at 422-24 (Stevens concurring). But his broader point is on the mark: it understands the critical issue, in evaluating regulation, as relating to the government’s motive.

the reasons supporting the fighting words category. This lack of fit provides grounds to suspect that the government is acting for other reasons, which the Court has yet to identify, let alone find legitimate. So whereas a full-scale ban on fighting words is "low risk," a partial, content-based ban on fighting words is not. The Court's treatment of each relates to the probability of taint, but that single gauge points in opposite directions.

A second recent case—*City of Cincinnati v Discovery Network, Inc.*¹⁸⁵—also highlights the connection between governmental motive and so-called low-value categories. The case began when city officials banned newsracks disseminating commercial handbills—62 of the approximately 1,500 newsracks cluttering Cincinnati's sidewalks. The city asserted, as its basis for action, interests in safety and aesthetics. Of course, commercial and noncommercial newsracks look the same and thus implicate these interests in the same way. The city noted, however, that commercial speech had "low value."¹⁸⁶ If this were true, the city claimed, then its interests could justify a ban on commercial newsracks, but not any others. This claim, accepting the premise that commercial speech has low value, is unexceptionable; and the Court accepts exactly this premise in commercial speech cases. Yet the Court rejected the city's argument. Why?

Said the Court: Because the distinction between commercial and noncommercial speech "bears no relationship *whatsoever* to the particular interests [in safety and aesthetics] that the city has asserted."¹⁸⁷ Well, yes—the city had admitted as much. But why should this matter if commercial and noncommercial speech have different values, such that a single interest can outweigh the one kind of speech but not the other? Said the Court: "In our view, the city's argument attached more importance to the distinction between commercial and noncommercial speech than our cases warrant and seriously underestimates the value of commercial speech."¹⁸⁸ Did this mean commercial speech henceforth would receive full constitutional protection? In cases after *Discovery Network*, the Court has continued to treat such speech as falling within a low-value category.¹⁸⁹ And in *Discovery Network*

¹⁸⁵ 507 US 410 (1993).

¹⁸⁶ Id at 418-19.

¹⁸⁷ Id at 424. See also id at 426-31.

¹⁸⁸ Id at 419. See also id at 428-31.

¹⁸⁹ See *Rubin v Coors Brewing Co.*, 115 S Ct 1585, 1589 (1995).

itself, the Court applied not strict scrutiny, but the intermediate standard of review devised for commercial speech cases.

The only way to reconcile *Discovery Network* with the Court's commercial speech doctrine is to see that doctrine as emerging from a judgment of risk, rather than a judgment of value. Once the doctrine is viewed in this way, the Court can apply relaxed scrutiny to regulations of commercial speech at the same time as it instructs governmental officials to treat this kind of speech as having the worth of any other. The problem in *Discovery Network* was that Cincinnati had mistaken the basis of the commercial speech category. Cincinnati had treated the creation of this category as a kind of substantive statement, addressed to courts and legislatures alike. The city had argued: Commercial speech has low value and therefore may be regulated on a reduced finding of harm. The Court in *Discovery Network* responded that the creation of the category was instead an evidentiary statement, addressed only to courts. The Court, that is, had said: Regulation of commercial speech has low risk and therefore should be held to a reduced standard of review. *Discovery Network* thus well illustrates the true nature—contrary to the appellation—of low-value categories. On this understanding, low-value categories fall into line with the rest of First Amendment law; they become another way of focusing and refining the search for motive.

2. Secondary effects doctrine.

The Court in recent years has toyed with an approach, known as the "secondary effects" doctrine, that would treat another swath of content-based regulations as if content neutral. This doctrine, in focusing on the nature of governmental justification, appears at first glance to prove my theory. The reality is more complex. The motive-based model and the secondary effects doctrine operate in considerable tension with each other. Even so, this model comes closer than its competitors to explaining the doctrine of secondary effects. This explanation turns, once again, on the relative "safety," with respect to motive, of certain kinds of content-based actions.

The essence of the secondary effects doctrine runs as follows: facially content-based regulations of speech that "are *justified* without reference to the content of the regulated speech" should be treated as if they made no facial distinctions on the basis of content.¹⁹⁰ In the seminal case, *City of Renton v Playtime The-*

¹⁹⁰ *City of Renton v Playtime Theatres, Inc.*, 475 US 41, 48 (1986), quoting *Virginia*

aters, the Court approved, under the standard of review usually used to test facially content-neutral action, a zoning ordinance applying to (and only to) theaters that showed sexually explicit movies.¹⁹¹ The Court explained its choice of standard by reference to the aim of the zoning ordinance. The ordinance, the Court emphasized, was not intended to “suppress the expression of unpopular views”¹⁹² or to “restrict[] the message purveyed by adult theaters.”¹⁹³ Rather, the law was “designed” to achieve certain so-called secondary effects—specifically, “to prevent crime, protect the city’s retail trade, maintain property values, and generally ‘protect and preserv[e] the quality of [the city’s] neighborhoods.’”¹⁹⁴ Because the ordinance aimed at these objects, it did not demand heightened scrutiny.

This doctrine contravenes the strictures of both the speaker-based and the audience-based models of the First Amendment. Consider two laws, one restricting sexually explicit speech to preserve neighborhood character, the other restricting the identical speech to suppress the message of sexual libertinism. The two laws curtail to the same extent both a speaker’s opportunity to express a message and a listener’s opportunity to consider its merits. Under each model, then, the difference between these two laws should have no constitutional significance. The secondary effects doctrine makes that difference relevant because, unlike these effects-based models and despite its name, it concerns itself with a regulation’s reasons.

Stated in this way, the secondary effects doctrine seems to offer a particularly fine example of the way in which the search for motive structures the law of the First Amendment. Indeed, an understanding of this kind might explain the Court’s creation of the doctrine. To the Court, the doctrine may have seemed but a natural—even an essential—aspect of First Amendment law’s focus on reasons.

In fact, however, the secondary effects doctrine fits uneasily with the rest of First Amendment jurisprudence. Although the

Pharmacy Bd v Virginia Citizens Consumer Council, Inc., 425 US 748, 771 (1976) (emphasis in *Renton*). The Court has not yet determined whether the secondary effects doctrine applies in the realm of fully protected speech. The Court has used the doctrine only when dealing with sexually explicit expression. The Court, however, has not made the presence of arguably low-value speech a definite condition of the doctrine’s application and once has suggested to the contrary. See *Boos v Barry*, 475 US 312, 320-21 (1988).

¹⁹¹ 475 US 41, 54-55 (1986).

¹⁹² *Id.* at 48.

¹⁹³ *Id.*, quoting *Young v American Mini Theatres, Inc.*, 427 US 50, 82 n 4 (1976).

¹⁹⁴ *Id.*, quoting the appendix to the Jurisdictional Statement.

doctrine reflects the First Amendment's concern with illicit motive, it disdains the usual method by which that concern is effected. Rather than using objective tests as proxies for examining motive, the doctrine of secondary effects insists on directly addressing this question. Construed in one way, the doctrine requires courts to make independent determinations of the government's reasons, notwithstanding the difficulties of proof entailed in this effort. Construed in another, more practical way, the doctrine counsels courts to accept without investigation plausible assertions as to governmental motive. Under either approach, courts often will countenance what they should not, mistaking the pretextual reason for the real and the tainted motive for the pure. The secondary effects doctrine thus seems to run counter to—and thereby negate the effect of—all the indirect techniques for flushing out illicit purpose that the Court has developed.

Is, then, the secondary effects doctrine a simple mistake—a doctrine meant to address the danger of improper purpose, but designed so as to undermine its own objective? Before reaching this conclusion, a closer examination of the doctrine is in order. Such an inspection may show the secondary effects doctrine to be a more sensible response to the fear of improper purpose than the I have just suggested. Or, less optimistically, this investigation at least may suggest ways of interpreting and shaping the doctrine so as to make it an aid, rather than a hindrance, in the effort to detect improper motive.

The distinction that drives the secondary effects doctrine, at least in its most sensible incarnation, is a distinction between the communicative (primary) impact and the noncommunicative (secondary) impact of expression. An example focusing on the regulation of child pornography will illustrate the point. Assume that two municipalities enact an identical child pornography statute. City A does so on the ground that viewing child pornography increases the probability that an individual will commit an act of child molestation. City B does so on the ground that manufacturing child pornography itself involves acts of child molestation. Which statute (both, either, none) should the Court treat as content neutral because based on secondary effects?

On one conceivable view of the secondary effects doctrine, both statutes would fall within the doctrine's scope. Neither statute, after all, arises from a censorial purpose—from the simple desire, in the words of *Renton*, to “suppress expression of unpopu-

lar views.”¹⁹⁵ Both statutes instead arise from the desire to prevent real-world harms of a serious nature; hence, both provoke application of the doctrine of secondary effects. But this understanding of the secondary effects doctrine would revolutionize free speech law. It would turn every First Amendment case into a secondary effects case, given that the government almost always can proffer a justification based on harm.

A second view of the secondary effects doctrine avoids this outcome by narrowing the doctrine's scope. Under this view, City B's justification for regulating child pornography is based on a secondary effect of the expression, but City A's is not. The difference lies in whether the harm the government is seeking to prevent arises from the expressive aspect of the communication—or, stated in another way, whether the harm results from a listener's hearing the content of speech and reacting to it. The harm justifying City A's ordinance arises from such a communicative effect: the content of the speech moves a listener to engage in hurtful behavior. The harm justifying City B's ordinance does not arise from the content of communication in this way. Although the harm relates to expression—more, to expression of a certain content—the harm does not grow out of a listener's response to a message. In the language of three Justices who appear now to represent the view of the Court on this issue, the government justifies its regulation of content on the ground that “the regulatory targets happen to be associated with” this content, rather than on the ground that the comixture between the content and the listener itself brings the regulatory target into being.¹⁹⁶

¹⁹⁵ *Renton*, 475 US at 48. Chief Justice Rehnquist, the author of *Renton*, may well hold this broad view of the secondary effects doctrine. All of the language in *Renton* distinguishes between a desire to suppress speech because of the harm it causes and a desire to suppress speech because of the ideas it espouses. By contrast, none of the language distinguishes between harms arising from the communicative aspects of speech and harms arising from the noncommunicative aspects. See *id.* at 47-49. Indeed, the Chief Justice, along with several other Justices, refused to join a portion of a later opinion attempting to narrow application of the secondary effects doctrine to regulations of speech based on noncommunicative impact. See *Boos*, 485 US at 338 (Rehnquist concurring).

¹⁹⁶ *Boos*, 485 US at 320 (O'Connor opinion). Secondary and primary effects, defined in this way, are sometimes difficult to tell apart. *Renton* provides one example: is increased crime merely “associated with” sexually explicit speech, or does it result from the effects of such speech on an audience? Or consider a statute drawing a content-based distinction on the ground that speech of a certain content draws large crowds and thus causes congestion. Does the congestion arise from the communicative effect of the speech or is the congestion merely associated with the expression? Many similar quandaries could be devised. But in this context, as in many others, the ability to concoct cases that might fit into one or another category does not gainsay the usefulness of the categories generally. In most instances, the distinction between primary and secondary effects, drawn in the

Why should this distinction between communicative and noncommunicative impact count as crucial in delimiting the doctrine of secondary effects? On one explanation, the distinction matters because it marks the precise divide between permissible and impermissible motive in the regulation of expression. The government, that is, may not restrict expression for any reason relating to communicative impact; the government may limit speech only for reasons independent of the response of listeners to a message.¹⁹⁷ Thus follows naturally the role of communicative impact in shaping the doctrine of secondary effects: the Court, in asking whether a restriction is justified in terms of communicative impact, is doing no more and no less than asking whether a restriction is tainted with improper motive.

But this explanation is unsatisfying. I have argued before that the concept of impermissible purpose posited in this account conflates dissimilar governmental motives and conflicts with the Court's First Amendment decisions.¹⁹⁸ More important for current purposes, the explanation just offered fails to address, much less to dissolve, the criticism that the doctrine of secondary effects hinders, rather than aids, the effort to uncover improper purpose. If all reasons relating to the communicative impact of speech are improper, then the current doctrine of secondary effects indeed turns on the concept of illicit motive. But in that case, the doctrine defeats its own mission by discarding the (necessary) indirect means of discovering motive and substituting a (hopeless) direct inquiry.

A second and better explanation for making communicative impact the key to the doctrine of secondary effects is that the resulting doctrine refines courts' efforts to flush out the narrow class of motives I have labeled improper (not all reasons relating to communicative impact, but only reasons of ideology or self-interest). On this view, the concept of communicative impact, instead of defining what counts as legitimate purpose, plays a quasi-evidentiary role, signaling a change in the standard of review a court needs to uncover the presence of improper motive. Assume here that improper motive, as I have defined it, is easier to detect—and less likely, in the first instance, to exist—when the justification for a statute relates to noncommunicative, rather

way described in the text, can be applied in a sensible manner.

¹⁹⁷ I have discussed this view of impermissible purpose, suggested in the writings of John Hart Ely and Laurence Tribe, at text accompanying notes 67-71.

¹⁹⁸ See text accompanying notes 60-66, 69-71.

than communicative, impact. In this event, a relaxed standard of review would suffice in secondary effects cases to separate proper and improper motives, and a court should discard its usual "sledgehammer" standard for a daintier constitutional instrument. The secondary effects doctrine, it might be argued, accomplishes just this result, thus rendering still more precise the complex mechanism established by First Amendment law to uncover improper motive.

This theory, of course, rests on a nonobvious premise: that courts can detect improper purpose more easily, and legislators will resort to it less often, when the justification for a restriction refers to secondary, rather than primary, effects. Why would this be? The key point is that because the harm in secondary effects cases derives from a thing only contingently related to expression, courts and legislators in these cases possess, to a greater degree than is usual, two testing devices for stripping away pretexts and revealing motives.

First, a court usually can check for improper motive in a secondary effects case by asking whether the government has tried to regulate the affected speech in the absence of the harm asserted. Suppose, as in *Discovery Network*, that the government justifies a restriction on commercial publications by pointing to the aesthetic deficiencies of newsracks; then the restriction can go no further than to limit this one means of distribution.¹⁹⁹ Or suppose, as in *Simon & Schuster, Inc. v Members of the New York Crime Victims Bd*, that the government justifies a law impounding profits from published descriptions of crime as a way to compensate the crime's victims; then the provision can apply only when "a victim remains uncompensated."²⁰⁰ These examples suggest that secondary effects cases often involve narrower restrictions on speech than do primary effects cases.²⁰¹ But the critical point concerns not the breadth of the restriction, but the ease of deciding whether the justification offered for it is real or pretextual. In almost any secondary effects case, a control group

¹⁹⁹ 507 US at 425-26. The restriction at issue was so limited.

²⁰⁰ 502 US 105, 122 (1991). The Court found the statute overinclusive in part because it applied to works that did not implicate the interest in compensation. *Id.*

²⁰¹ See Scanlon, 40 U Pitt L Rev at 528 (cited in note 179). This is usually, but not necessarily, true. A flat prohibition on speech may derive from a secondary justification. See, for example, *New York v Ferber*, 458 US 747 (1982) (ban on child pornography based on harm to the child photographed). Conversely, a limited regulation may derive from a primary justification. See, for example, *Nebraska Press Association v Stuart*, 427 US 539 (1976) (restrictions on press coverage of trial based on potential response of public).

will exist consisting of the identical speech unaccompanied by the (nonexpressive) agent causing the asserted harm. A check of this kind will exist more rarely in primary effects cases because there the harm derives from the substance of the speech, rather than from features severable from it.²⁰² That difference may make the difference in the effort to flush out improper motive.

Second, a court usually can check for improper motive in a secondary effects case by asking whether the government has regulated conduct, as well as other speech, that causes the same harm as the affected expression. The Court, for example, in *Discovery Network* asked whether the government had prohibited newsracks not only for commercial, but also for noncommercial publications, given that the two posed identical aesthetic issues.²⁰³ Similarly, in *Simon & Schuster*, the Court asked why the statute covered only the profits criminals gained from expressive activity, when the profits gained from nonexpressive activity also could have compensated victims of crime.²⁰⁴ Again, the ability to check the government's justification in this way depends in part on whether the case involves primary or secondary effects. If the government invokes primary (communicative) effects, then a court often (though not always) will have no nonspeech analogues by which to test the restriction; nonexpressive activity, or even expressive activity of a different content, will rarely implicate in the same way the same governmental interest. But if the government invokes secondary (non-communicative) effects, then a court usually (though again not always) will have a reference point in the government's treatment of nonexpressive or other expressive activity; this is so precisely because the asserted harm in the case comes from a non-communicative, hence severable, aspect of the expression.

Because of the common availability of these testing mechanisms in secondary effects cases, courts arguably do not need to apply the most stringent standard of review in order to flush out improper motive. Even the standard of review applicable to con-

²⁰² There are exceptions to each side of this comparison. First, if a noncommunicative aspect of speech attaches to all speech of a certain content, then the control group I am positing in secondary effects cases will not exist. I doubt, however, that there are many cases of this kind. Second, if a communicative aspect of speech causes harm only in particular circumstances, then a control group might exist in primary effects cases consisting of the same speech occurring in all other circumstances. I doubt, however, that such a verification device will arise as often or prove as easy to apply as the checking device available in secondary effects cases.

²⁰³ See 507 US at 425-26.

²⁰⁴ See 502 US at 119-20.

tent-neutral action bars any law that (1) restricts speech not implicating the governmental interests asserted, or (2) discriminates against speech relative to nonexpressive activities implicating the same interests.²⁰⁵ If the testing mechanisms I have discussed make this standard an effective filter in secondary effects cases, then use of a stricter standard would be overkill. When the nature of the asserted justification makes it easy to impeach, then courts do not need the powerful weapons designed to achieve this object in other circumstances.

I do not wish to stake very much on the strength of this motive-based explanation of the doctrine of secondary effects. Perhaps I have overstated the difference between primary and secondary effects cases with respect to the difficulty of discrediting pretexts and finding real reasons. If so, the doctrine of secondary effects constitutes at once a paradox and an error. The paradox is that the doctrine of First Amendment law most concerned with evaluating reasons is the doctrine least reconcilable with the motive-based model. The error, which creates the paradox just mentioned, lies in the decision to evaluate reasons by asking questions about them. In fact, the proper way to resolve the issue of motive is to pose other questions. If the secondary effects doctrine indeed is incompatible with the motive-based approach, it is because the former brings to the surface what the latter knows should reside beneath it.

But if the doctrine of secondary effects has any sound foundation, it relates to refining the search for improperly motivated governmental actions. More specifically, the doctrine emerges from the view that it is relatively easy in cases involving secondary effects to isolate the role played by hostility, sympathy, or self-interest. No other account of the doctrine of secondary effects makes better (or indeed, any) sense. And this account provides a guide for judges applying the doctrine. In highlighting the way in which a secondary justification affects the search for improper

²⁰⁵ The standard applicable to content-neutral regulations, although variously articulated, requires that a law extend only to speech implicating the asserted interest. See *Ward v Rock Against Racism*, 491 US 781, 798 (1989) (requiring that a regulation "be narrowly tailored to serve the government's legitimate content-neutral interests"); *O'Brien*, 391 US at 377 (demanding that a restriction be "no greater than is essential to the furtherance of [the asserted] interest"). Although not included in statements of the applicable standard of review, the principle that the government may not enact content-neutral laws discriminating against speech, as compared to nonexpressive activity, appears in (and alone explains) many cases. See *Martin v City of Struthers*, 319 US 141, 142-44 (1943); *Schad v Borough of Mount Ephraim*, 452 US 61, 73 (1981); *Minneapolis Star & Tribune Co. v Minnesota Comm'r of Revenue*, 460 US 575, 581-85 (1983).

motive, the account focuses judicial attention on that search, while revealing the checking devices available to conduct it. The doctrine of secondary effects thus adds further intricacy to the essential mechanism of First Amendment law: a mechanism devised to flush out improper purpose on the part of the government in the face of serious, but variable, problems of proof.

D. The Distinction Between Direct and Incidental Restrictions

Consider again the facts of *R.A.V.*, but this time with a twist, simple in form yet profound in consequence. Recall that in the actual case, St. Paul charged a juvenile who allegedly had burned a cross on the property of an African-American family with violating the city's hate-speech ordinance, a law interpreted to punish, and to punish only, certain kinds of expressive activity. Now suppose the conduct of the juvenile remains the same, but the prosecutor makes a different charging decision. Rather than resort to the city's hate-speech ordinance, the prosecutor relies on a generally applicable law—a law not specifically directed toward speech or other expressive activity. Any of a number of laws may come to mind—a statute prohibiting trespass or arson or the infliction of damage on another person's property.²⁰⁶ If the prosecution had proceeded under one of these laws—say, the trespass ordinance—the trial court would have dismissed a defense based on the First Amendment. Even assuming the burning of a cross qualifies as fully protected expressive activity, the case would have started and ended as a trespass prosecution.

This result arises from a distinction as important as any in First Amendment law: the distinction between direct and incidental restrictions on speech or, otherwise phrased, the distinction between actions targeting expression alone and actions applying generally, to both nonexpressive and expressive activity.²⁰⁷ The distinction received its most famous articulation in

²⁰⁶ The Court in *R.A.V.* noted the possibility that the conduct at issue violated statutes of this kind, thus implying that a prosecution brought under such a statute would have raised no First Amendment problems. 505 US at 380 n 1.

²⁰⁷ See generally Frederick Schauer, *Cuban Cigars, Cuban Books, and the Problem of Incidental Restrictions on Communications*, 26 Wm & Mary L Rev 779 (1985); Stone, 54 U Chi L Rev at 105-14 (cited in note 110). This distinction, of course, assumes another: between expressive and nonexpressive activity, sometimes loosely termed speech and conduct. I have noted elsewhere the need for First Amendment law to view expressive and nonexpressive activity as meaningfully different, even though drawing a line between the two raises hard questions. See Kagan, 60 U Chi L Rev at 883-84 (cited in note 61). For purposes of this Article, the more important point is that the Court always has distinguished between the two, usually by asking whether the activity in question is, in purpose

O'Brien, when the Court upheld the conviction of an antiwar protester who had burned his draft card for violating a federal statute prohibiting the intentional destruction of draft registration certificates. The Court explained that the prosecution comported with the First Amendment because the statute (and the governmental interest supporting it) applied regardless whether the conduct enjoined was expressive; hence the restriction on speech was incidental.²⁰⁸ As the Court more recently stated the proposition in *R.A.V.*, “[w]here the government does not target conduct on the basis of its expressive content, acts are not shielded from regulation merely because they express a[n] [] idea or philosophy.”²⁰⁹ So long as the law applies generally, it can apply to expression. Thus the distinction between prosecuting a cross burner under a trespass law and prosecuting him under a hate-speech ordinance.

Courts usually treat the application of a general law, even to activity concededly expressive, as raising no First Amendment issue whatsoever. So, for example, courts will not see a constitutional question if the government convicts for vandalism a person who draws swastikas on a synagogue wall; or applies taxation, labor, or antitrust laws to the publisher of a newspaper; or uses a residential zoning law to prevent the opening of a bookstore. On occasion, courts will apply a form of intermediate scrutiny to an incidental restraint, as the Court did in *O'Brien*,²¹⁰ still more rarely, a court will subject to strict scrutiny the use of a generally applicable statute.²¹¹ I will have more to say later about these more or less exceptional cases. But for now, the critical point is that incidental restraints on expression usually receive more deferential treatment than direct restraints on the same expression.²¹² The question is why this is so.

and function, primarily expressive. See *Spence v Washington*, 418 US 405, 410-11 (1974) (asking about the “intent to convey a particularized message” and the likelihood that “the message would be understood by those who viewed it”). I use this view of the difference between expressive and nonexpressive activity in this discussion.

²⁰⁸ *O'Brien*, 391 US at 376-82. Earlier statements of this principle appear in *Associated Press v NLRB*, 301 US 103, 132 (1937) (applying labor laws to press), and *Associated Press v United States*, 326 US 1, 7 (1945) (applying Sherman Act to press).

²⁰⁹ 505 US at 390.

²¹⁰ 391 US at 376-77. See text accompanying notes 225-34.

²¹¹ See *United States v Eichman*, 496 US 310, 318 (1990). See also text accompanying notes 235-43.

²¹² It is on account of this principle—and this principle alone—that the Court is right to say that “[t]he Government generally has a freer hand in restricting expressive conduct than it has in restricting the written or spoken word.” *Texas v Johnson*, 491 US 397, 406 (1989). The “freer hand” is a function not of any difference between verbal and nonverbal

If the law concerning direct and incidental restraints seems too obvious to merit this discussion, consider as a preliminary matter that the distinction first arose, and then operated for years, in a form converse to that of modern doctrine. In *Gitlow v New York*, the Court contrasted a statute directly targeting certain forms of advocacy to a statute “prohibit[ing] certain acts involving the danger of substantive evil, without any reference to language itself.”²¹³ The Court noted that it had greater call to review an application of the general statute to speech than to review an application of the targeted statute. In the latter case, the Court seemed to reason, the legislature already had made a considered judgment that the speech at issue posed the requisite danger; in the former case, the legislature had made no such judgment, and might have concluded to the contrary, had it ever considered the matter.²¹⁴ The unintentionality of the incidental restraint worked against it; the purposefulness of the direct restraint worked in its favor. The end result was a doctrine that treated the restriction of speech through a generally applicable law as more, rather than less, problematic than the restriction of speech through direct legislation.

What accounted for the Court’s eventual shift in understanding? Again, I argue that modern doctrine acquired its structure as an attempt to discover actions tainted with ideological motive. What changed between the old case law and the new was the Court’s understanding of its role in policing the inputs of governmental action affecting expression. Before I press this point, however, I again consider alternative arguments.

The inadequacy of a speaker-based model in explaining current doctrine emerges from a simple form of hypothetical, contrasting a direct restriction on speech to a generally applicable regulation that covers everything affected by the direct restriction, but more in addition. Suppose, for example, that one city makes it illegal to deface synagogues with swastikas, while another enjoins all acts of vandalism. Or that one city imposes a tax on corporations that publish newspapers, while another imposes the tax on corporations generally. Or, finally, that one city prohibits bookstores in an area, while another zones the area for

expression, but of the difference between laws targeting expression and laws applying more broadly; the government may restrict expressive conduct more easily than the written or spoken word because expressive conduct more often falls within the terms of generally applicable regulations.

²¹³ 268 US 652, 670 (1925).

²¹⁴ *Id.* at 670-71.

residential use only.²¹⁵ If what mattered were the effect of a regulation on a speaker's expressive opportunities, then courts would review these municipal acts in identical fashion; after all, the one city impairs speech no more than the other.²¹⁶ Of course, the generally applicable law may have a stronger state interest supporting it than does the targeted restriction. But this difference, even when it exists, would not justify separate standards of review; at most, it would explain how application of the same standard to these laws could produce different outcomes. The current doctrinal structure thus cannot result from an inquiry into the effects of a law on expressive opportunities.

Neither can the distinction between direct and incidental restrictions derive from a concern with the way in which governmental actions distort public discourse. Even supposing that incidental restrictions were subject to the same level of review as other content-neutral laws, this account of the doctrine would fail to persuade. It then would suffer from all the same difficulties as plague the attempt to explain, on grounds of distortion, the distinction between content-based and content-neutral laws generally.²¹⁷ Most notably, the account would ignore that any incidental restraint works against a backdrop of other laws—themselves functioning as incidental restraints—that may render the restraint a means of perpetuating not balance in discourse, but distortion. That laws affecting speech “only” incidentally may play a large role in shaping the speech market, either for good or for ill, should not by now be in question.

But using an audience-based model here poses further problems, for it cannot explain why current doctrine in fact treats incidental restrictions more deferentially than other (speech-specific) content-neutral restrictions. Even assuming that a law distorts public discourse to the extent it alters the *ex ante* distribution of opinion,²¹⁸ there is no reason to think that content-neutral direct restraints cause greater distortion than incidental

²¹⁵ The last example comes from *Stone*, 54 U Chi L Rev at 105 (cited in note 110). See also *Schauer*, 26 Wm & Mary L Rev at 779 (cited in note 207) (listing similar examples).

²¹⁶ Indeed, the city using general restrictions may curtail more speech than the city using direct restrictions. The general tax, for example, covers broadcast stations as well as newspapers, and the general zoning law covers not only bookstores, but video outlets.

²¹⁷ See text accompanying notes 91-100.

²¹⁸ I am indulging this assumption, which is contrary to all prior argument, to cast in the best possible light the audience-based explanation of the doctrine of incidental restraints. If the assumption is dropped, the explanation will fail regardless whether generally applicable laws or targeted content-neutral laws more greatly alter the existing distribution of opinion.

restrictions. Indeed, if either were to skew debate more than the other, incidental restrictions would count as the culprit. Whereas the defining characteristic of content-neutral direct restrictions is what I will call “horizontal sweep” (application to different kinds of speech), the defining characteristic of incidental restraints is “vertical sweep” (application to speech and conduct), which may or may not have a horizontal dimension. Consider, for example, the law in *O’Brien*, which in prohibiting the destruction of draft cards interfered with one viewpoint only. Given such cases, an approach focusing on distortion might turn current doctrine inside-out and require courts to review incidental restraints more closely than direct restrictions.²¹⁹

Perhaps the explanation of current doctrine lies solely in a set of practical constraints.²²⁰ If all laws incidentally restricting speech were subject to First Amendment review, then (almost) all laws would be subject to First Amendment review. This is to say no more than what I have suggested before: that many laws not specifically directed at speech have effects on expressive activity. The need to address the constitutionality of all such laws would impose significant costs. If, as seems likely, most of the laws would pass constitutional muster, incurring these costs does not seem worthwhile.²²¹ Better to assume from the outset that these laws raise no serious problem.

But an alternative or additional rationale is available: that the law of incidental restraints arises from a focus on governmental motive.²²² Consider that a generally applicable law by definition targets not a particular idea, nor even ideas broadly speaking, but an object that need not, and usually does not, have any association with ideas whatsoever. Recall the examples used

²¹⁹ I discuss below why a motive-based account would not do so, even given that differential effects are one indicator of improper motive. See text accompanying note 223.

²²⁰ See Schauer, 26 *Wm & Mary L Rev* at 784 (cited in note 207); Stone, 54 *U Chi L Rev* at 107 (cited in note 110).

²²¹ If any significant portion of the laws would fail constitutional scrutiny, a different—but perhaps no less trenchant—objection to judicial review would arise. The core issue then would concern the propriety of interpreting the First Amendment to impose substantial limits on the government’s power to enact or apply regulations not specifically addressed to expression. Resolution of this issue would depend on the selection of a theory of the First Amendment; the theory posited in this Article, for reasons that follow, holds that the First Amendment should not be understood to impose these limits.

²²² See Schauer, 26 *Wm & Mary L Rev* at 783 (cited in note 207); Schauer, *Free speech* at 100-01 (cited in note 39). The doctrine, viewed in this way, has an analogy in equal protection law, where the Court has held that facially race-neutral laws pose a relatively minor risk of stemming from impermissible motive and hence should receive relaxed scrutiny. See *Washington v Davis*, 426 US 229, 247-48 (1976).

above: laws prohibiting vandalism, imposing a tax, or setting a zoning restriction. The breadth of these laws makes them poor vehicles for censorial designs; they are instruments too blunt for effecting, or even reflecting, ideological disapproval. Thus, incidental restrictions receive minimal scrutiny because of the likelihood that they also will be accidental restrictions in the relevant sense—that they will result from a process in which officials' hostility (or partiality) toward ideas played no role.

This analysis explains why the Court treats incidental restraints more deferentially even than direct restraints of a content-neutral nature. It is true that the breadth of content-neutral direct restraints also lessens the likelihood of illicit motive. Indeed, to the extent that effects constitute evidence of intent, the argument I made earlier about horizontal and vertical sweep might support not the laxer, but the harsher treatment of incidental restraints than of content-neutral direct restrictions. But the difference in the facial terms of the two kinds of regulations—that, by definition, one goes to speech, the other to conduct—cuts sharply in the direction of existing doctrine.²²³ When a proposed law is addressed to expression, a legislator cannot help but consider, consciously or not, whether and how the law will affect particular messages; this is to say little more than that when a law is about speech, the legislator will consider its impact on speech—a proposition neither deep nor shocking. But when a proposed law, by its terms, focuses on nonexpressive conduct, restricting speech only as an incidental and thus a covert matter, the probability increases that a legislator will consider the regulation divorced from hostility or sympathy toward particular messages.

A law of general application, of course, may have such dramatic—and apparent—effects on expressive activities that it might as well target those activities in express terms. I soon note that it is in just these circumstances that courts treat generally applicable laws as if they were direct restrictions on expression. But even were this escape hatch unavailable, a decision to treat incidental restraints with a degree of deference not given to content-neutral direct restrictions would have an adequate justification. If it is not always true, it is true often enough that bias toward ideas will infect the former less than the latter.

²²³ I similarly argued in Section III.A that the facial terms of a regulation—there, whether the regulation was content based or content neutral—provides a better indication of governmental motive than does any measurement of content-differential effects.

The Court's abandonment of *Gitlow* thus results from its adoption of the foundational principle of modern doctrine. The thing that *Gitlow* counted in a law's favor—the purposeful and considered judgment of the desirability of repressing expression—now appears a grave danger, because it so easily may become tainted with ideological factors. The thing that *Gitlow* counted against a law—the nonobviousness of its relation to, or effects on, expression—now appears a great boon, because when legislators do not consider the question of restricting speech, *a fortiori* they do not consider it in an improper manner. The key point concerns the relation of general applicability to neutrality and of neutrality to motive.²²⁴ Generality implies neutrality among ideas as such; neutrality of this kind signifies and, indeed, defines the absence of impermissible motive.

If, in this way, a concern with motive accounts for the doctrine of incidental restraints, then the same concern ought to aid in explaining exceptions to this doctrine: the cases in which courts apply intermediate or even strict scrutiny to generally applicable laws. A brief review of the cases confirms this thesis.

The Court in *Arcara v Cloud Books, Inc.*²²⁵ attempted to explain why some incidental restraints, rather than falling outside the First Amendment's sphere of influence, receive the scrutiny usually given to content-neutral direct restrictions.²²⁶ According to *Arcara*, generally applicable laws warrant First Amendment scrutiny in only two circumstances. First, such scrutiny is appropriate when the law, although not specifically referring to speech, "has the inevitable effect of singling out those engaged in expressive activity."²²⁷ The Court cited as an example the imposition of a tax on newsprint, which would fall disproportionately—indeed, almost exclusively—on speakers. Second, review is proper when the conduct that draws the sanction has a significant expressive element. The Court used *O'Brien* as an example, noting that the act regulated in that case—the destruction of a draft card—itself "carr[ie]d] a message."²²⁸ The Court cited in contrast

²²⁴ For discussion of this relation in the context of religious liberty, see *Church of the Lukumi Babalu Aye, Inc. v City of Hialeah*, 508 US 520, 532-36 (1993).

²²⁵ 478 US 697 (1986).

²²⁶ Viewed solely in historic terms, this question has proved unimportant. Although the Court sometimes has subjected incidental restrictions to intermediate scrutiny, the Court never has overturned a restriction reviewed under this standard. See Schauer, 26 *Wm & Mary L Rev* at 787-88 (cited in note 207). When applied to incidental restraints, intermediate scrutiny has acquired a peculiarly toothless quality. Still, the decision to subject certain incidental restraints to heightened scrutiny has the potential to matter.

²²⁷ *Arcara*, 478 US at 707.

²²⁸ *Id.* at 702-03. Of course, destruction of a draft card does not always carry a mes-

cases in which the sanction, although affecting expressive activity, stemmed from nonexpressive conduct—as where a city closed a bookstore or fined a newspaper for violations of a zoning law or minimum wage ordinance.

The relevance of hugely disproportionate impact to the level of scrutiny is, under a motive-based approach, no great mystery. What separates direct from incidental restraints is breadth: whether the law applies to more than just speech. If an incidental restraint has no such sweep, effectively regulating only speech, then the danger it poses of illicit motive approaches the level associated with direct restraints, and the same standard of review thus should obtain.

A similar purpose-based analysis applies to the second factor the Court deemed important: whether the legal sanction results directly from, or merely impedes, expressive activity. The significance of this distinction is by no means transparent, nor does it fully explain the case law.²²⁹ But if the distinction matters, it does so because it relates to motive—here, of administrative and judicial actors. When expressive activity triggers application of a law, the expression and the legal violation become ineluctably

sage; if it did, banning the act would constitute a direct, rather than incidental, restraint on expression. The point here is only that the conduct drawing a sanction in the particular case expresses a message. The Court in *Arcara* cited two other cases as similar to *O'Brien* in this respect, though only one in fact is so. In *Clark v Community for Creative Non-Violence*, 468 US 288 (1984), the Court considered the application of a ban on sleeping in Lafayette Park to demonstrators who sought to sleep in the park in order to call attention to the plight of the homeless. Here, the banned activity—sleeping—indeed had an expressive element. In *United States v Albertini*, 472 US 675 (1985), the Court considered the constitutionality of a statute barring a person from reentering a military base after being ordered not to do so, as applied to a person who wished to reenter a base in order to distribute leaflets. In this case, contrary to the Court's analysis, the conduct that drew the sanction—the reentry—was not itself expressive, although the policy regulating reentry interfered with the leafletter's ability to engage in expressive activity. Another explanation is thus necessary to explain the decision. See text accompanying notes 230-34. A more recent case, *Barnes v Glen Theatre, Inc.*, 501 US 560 (1991), conforms to the *Arcara* analysis. There, the Court considered whether a state could apply a public indecency statute to a nude-dancing establishment. Because expressive activity (nude dancing) drew the legal penalty, the Court (except for Justice Scalia) understood the case to raise a First Amendment question. See *id.* at 565-66 (plurality opinion); *id.* at 581 (Souter concurring); *id.* at 587 (White dissenting). But see *id.* at 576-78 (Scalia concurring).

²²⁹ The Court has subjected to First Amendment review some restrictions that merely burden, rather than come down directly upon, expressive activity; *Albertini* presents an example. See note 228. Conversely, courts find First Amendment review unnecessary in some cases where expressive activity draws the sanction. I doubt, for example, that courts would engage in First Amendment review of a city's decision to apply a vandalism law to a person who draws swastikas on a synagogue wall. And I am sure courts would decline to apply First Amendment scrutiny to the conviction (under a statute prohibiting the destruction of property) of a political terrorist for blowing up the Statue of Liberty.

intertwined; this linkage is usually both less visible and less tangible when application of the law merely burdens expression. To the extent this entanglement occurs, the danger grows that prosecutors and fact finders alike will consider the nature of the expression in discharging their functions. The danger of discriminatory enforcement in such a case becomes roughly equivalent to that present when a speaker arguably violates a direct content-neutral restriction on speech, such as an antileafletting ordinance. Because the closeness of the link between sanction and expression affects the danger of improper motive, so too does it shift the appropriate constitutional standard.

In the end, however, the Court's decision to apply intermediate review to certain incidental restrictions may result not so much from use of the *Arcara* test as from a visceral sense that an illicit factor entered into a governmental decision—whether legislative, administrative, or judicial. Consider when the Court has treated an incidental restraint as raising a cognizable First Amendment question calling for intermediate scrutiny. First, in *O'Brien*, the law at issue, viewed in light of both legislative and extralegislative history, gave many indications, acknowledged (though trivialized) by the Court, of censorial motivation.²³⁰ Next, in *United States v Albertini*²³¹ and *Clark v Community for Creative Non-Violence*,²³² administrative officials interfered with the expression of persons (like the speaker in *O'Brien*) engaged in protest against governmental policies.²³³ Finally, in *Barnes v Glen Theatre, Inc.*, the state itself created suspicion by announcing that it would apply the allegedly general law to some, but not all expressive activities.²³⁴ In short, in all of the cases in which the Court has tested an incidental restraint against an intermediate standard of review, there have been signs of impermissible motive.

²³⁰ See *O'Brien*, 391 US at 385-86. See generally Dean Alfange, Jr., *Free Speech and Symbolic Conduct: The Draft-Card Burning Case*, 1968 S Ct Rev 1.

²³¹ 472 US 675 (1985).

²³² 468 US 288 (1984).

²³³ In *Albertini*, the speaker wished to protest the nation's nuclear-arms policy; he previously had destroyed military documents. 472 US at 677-78. In *Community for Creative Non-Violence*, the speakers wished to protest the government's treatment of the homeless. 468 US at 289. An additional concern of improper motive in that case arose from the circumstances surrounding the adoption (not just the administration) of the facially general regulation. See *id* at 315-16 (Marshall dissenting).

²³⁴ 501 US 560, 590 (1991) (White dissenting). The State asserted that its public indecency statute prohibited nude (erotic) dancing but did not prohibit nudity in theatrical or operatic productions.

In other cases, the Court has gone further, subjecting apparently incidental restraints to strict scrutiny—again, for reasons relating to motive. One set of these laws I have mentioned before: content-neutral laws whose application turns on the communicative effects of expression.²³⁵ Many of these laws are generally applicable, in the sense I use the term; for example, a law barring breaches of the peace applies to both expressive and nonexpressive activities that disturb public order. I have suggested earlier why strict scrutiny applies in these circumstances: because of the way in which the content of ideas triggers the application of such a law, the ideological views of officials and members of the public likely will influence its enforcement. Asking whether the application of a facially general law turns on communicative impact thus serves as a way of deciding whether bias tainted the application.²³⁶

The Court also will apply strict scrutiny to a law of general application when either the asserted justification or the only rational justification for the law (or an application of it) relates to the communication of a message. *O'Brien* suggested this approach in asking whether the governmental interest asserted is related “to the suppression of free expression,” properly understood to mean whether the interest is related to the suppression of specific messages.²³⁷ If the asserted interest or the only rational interest for an action is of this kind, then a court can assume that the official taking the action indeed considered the desirability of restricting certain messages. And when this is true, the probability is high that bias tainted the decision. This is not because all interests relating to the content of ideas are improper; to the contrary, many such interests reflect the cognizable harm an idea produces. But when officials self-consciously consider the merits of restricting ideas—whether because a law on its

²³⁵ See Section III.B.2.

²³⁶ At first glance, these cases may look no different from a case such as *Community for Creative Non-Violence*, 468 US at 288, where expressive activity also triggered the legal sanction. But in these cases, the content of expression, rather than the simple fact of expression, produces the sanction; the difference is between application of a breach-of-the-peace statute to very loud expression and application of the same statute to expression whose provocative content has stirred public ire.

²³⁷ 391 US at 377. This interpretation follows as a simple matter of logic. The test *O'Brien* proposes for a law whose justification does *not* relate to the “suppression of free expression” is essentially the test applied to content-neutral regulations of speech. To merit a stricter standard of review, a law would need to have a justification relating not to the restriction of speech generally (which all content-neutral laws have), but to the restriction of speech of a certain content.

face effects this restriction or because the law, though written generally, has this restriction as its stated or obvious object—bias easily influences the evaluation of neutral harm-based criteria. An asserted or inescapable content-based interest, like content-based language, thus demands use of strict scrutiny.²³⁸

The point may become clearer by considering the legality of laws banning flag burning.²³⁹ Note first, as illustration of a critical point of this Section, that the government may stop protesters from burning flags by enacting a general restriction—say, a ban on lighting fires in public places. But the government may not specifically proscribe the burning of flags for purposes of protest. The effects of these laws are of course no different. The dissimilar treatment arises, and must arise, from the disparate concerns about governmental motive raised by one law that sweeps broadly and another that focuses on expressive activity.

But now consider a harder case—a case involving the constitutionality of a law prohibiting any person from knowingly mutilating the flag of the United States.²⁴⁰ The greater difficulty of the case again demonstrates the importance of motive, for the statute, when compared to the other two mentioned, has no different effect on expressive activity. What makes the case hard is that the statute falls near the line between incidental and direct restraints and thus raises a question about how far to suspect the motives of government. On one argument, the law applies on its face to an activity generally, regardless whether expressive in purpose or function; the law covers alike the person who burns a flag to protest a war and the person who uses flags for kindling.²⁴¹ But on the contrary argument, all of the rational inter-

²³⁸ The same kind of analysis applies to a facially general law that applies to nothing, or almost nothing, but speech of a certain content. Just as a facially general law that operates to restrict only speech (and not conduct) ought to confront intermediate scrutiny, see text accompanying notes 227-29, so a facially general law that operates to restrict only speech of a particular kind ought to confront the strictest review. Arguably, the law in *O'Brien* was of this nature and ought to have been treated as if it exclusively related to antiwar protest, rather than encompassed as well “the odd soul who burns a draft card just to stay warm or to light up his campsite” or to deliver some other message. Tribe, 1993 S Ct Rev at 34 (cited in note 78). The same might be true of all the variations of flag-burning statutes I discuss below.

²³⁹ The Court’s two recent cases on the subject are *Eichman*, 496 US 310, and *Johnson*, 491 US 397.

²⁴⁰ The hypothetical law provided here is a simplified version of the Flag Protection Act of 1989, 18 USC § 700 (1994), at issue in *Eichman*, 496 US at 314.

²⁴¹ This feature of the Flag Protection Act made *Eichman* more difficult than *Johnson*, in which the Court a year earlier had invalidated Texas’s flag-burning statute. The Texas law prohibited flag burning only when it would cause offense to others—that is, only when it functioned as communication. See *Johnson*, 491 US at 400 n 1.

ests underlying the law relate to the restriction of a message.²⁴² In such a case, an indicator of illegitimate motive (content-based justification) undermines the indicator of legitimate motive (general application).²⁴³ This is why the Court, in invalidating the statute, made the correct decision.

A brief discussion of another question recently decided by the Court—the constitutionality of hate-crimes laws—may serve as a summary of all these issues. Hate-crimes laws, as usually written, provide for the enhancement of penalties when a specified crime (say, assault) is committed because of the victim's race, religion, or other listed characteristic.²⁴⁴ In *Wisconsin v Mitchell*,²⁴⁵ the Court unanimously ruled that these laws present no First Amendment issue. The analysis so far helps to explain why the Court reached (and was right to reach) this decision.²⁴⁶

The key lies in the fact that the typical hate-crimes law—unlike a hate-speech law, as in *R.A.V.*—is an incidental restraint. On its face, the hate-crimes law targets not only speech, but a range of activity; it applies regardless whether the conduct at issue expresses a message. In this way, a hate-crimes law functions in the same way as any discrimination law—for example, in the sphere of employment relations.²⁴⁷ When an

²⁴² See *Eichman*, 496 US at 315-16. Reasonable people can disagree—in fact, have disagreed—with this conclusion. See, for example, Geoffrey R. Stone, *Flag Burning and the Constitution*, 75 Iowa L Rev 111, 119 (1989) (proposing, with respect to a carefully drafted flag-burning statute, a reasonable interest unrelated to the communication of a particular message). I do not wish to rehash this debate, although I find the Court's position more persuasive. The critical points here concern the consequences of finding that the interests supporting a facially general law relate to the suppression of ideas, and the reason why those consequences follow.

²⁴³ Again, the content-based justification is only an indicator of impermissible motive, not the thing itself. Many justifications relate to the suppression of messages, but are permissible because they relate to material harms the speech causes; consider, for example, a justification for a flag-burning statute that focuses on the tendency of such speech to provoke unlawful conduct. The problem of motive arises from the difficulty of evaluating this legitimate interest uninfluenced by ideological views of the speech in question. This problem will be evident in any case where the government's asserted interest in a law refers to a particular message, just as it is where the law itself makes this reference.

²⁴⁴ See, for example, Cal Penal Code § 422.7 (West 1988 & Supp 1995); Wis Stat Ann § 939.645 (West Supp 1994).

²⁴⁵ 508 US 476 (1993).

²⁴⁶ For an excellent treatment of this question, see Tribe, 1993 S Ct Rev at 4-11 (cited in note 78). See also Kagan, 60 U Chi L Rev at 884-87 (cited in note 61).

²⁴⁷ The Supreme Court in *Mitchell* recognized the analogy between Title VII and a hate-crimes statute. See *Mitchell*, 508 US at 487. It is noteworthy that both laws apply not only irrespective of whether the discrimination at issue expresses a message, but also irrespective of whether the discrimination stems from particular beliefs. If, for example, discrimination laws prohibited discharges or assaults motivated by racial hatred—rather than based on race—they would pose a more severe First Amendment problem.

employer fires an employee on the basis of race, the government may impose sanctions whether or not speech accompanies or itself effects the discharge; whatever speech is involved is incidental to the activity (race-based discharge) that the law condemns. The government may do the same when one person assaults another on the basis of race, again whether or not speech accompanies the conduct; a penalty enhancement may follow because it is pegged to conduct (race-based assault) that the state is attempting to prevent irrespective whether it has an expressive component. In both cases, the generality of the law provides a qualified assurance that disapproval of ideas qua ideas played no causal role in the legislative process.

One objection to this analysis might focus on the extent to which speech—more, speech of a certain kind—accompanies the prohibited conduct: if racist expression always or almost always is associated with race-based assault, then proscribing the activity amounts to proscribing the expression, with all the constitutional issues such a policy raises. In such a case, the so-called generally applicable law is not generally applicable after all; because the law applies only to certain expressive activity, the reasons for trusting the law disappear.²⁴⁸ But I do not think this description accurately characterizes hate-crimes laws, which ban conduct that may and often does occur independent of expression; indeed, persons committing race-based crimes may try hard to conceal, rather than express, the racism inherent in the conduct.²⁴⁹ Here, communication is neither so integral to nor so coincidental with the condemned activity as to reverse the usual presumption supporting generally applicable regulations.

A second objection to the analysis might point to the uneven way in which a hate-crimes law affects speech (when it does affect speech), effectively barring racist ideas and not others. There is no way to deny this skewing effect, and if it matters, then it calls for reversal of *Mitchell*. But I think it should not make such a difference. Many generally applicable laws affect speech in an asymmetrical way, as the conduct proscribed captures the expression of only certain messages.²⁵⁰ Consider, for example, what kind of speech is likely to accompany a race-based discharge.²⁵¹ But if the law applies to conduct generally, the

²⁴⁸ See text accompanying notes 227-29, 238.

²⁴⁹ Tribe, 1993 S Ct Rev at 9-10 (cited in note 78).

²⁵⁰ See text accompanying notes 219-24.

²⁵¹ Examples can proceed ad infinitum. Consider, to note two more, what kind of

critical barrier to the intrusion of illicit motive remains intact. That barrier is the focus of the law on acts irrespective of expression—a focus that usually prevents attitudes toward a message from influencing the legislative outcome. Again, then, the usual principles applicable to incidental restrictions seem to hold with respect to hate-crimes statutes.

The last objection to the analysis also is the strongest: that the only rational justification for a hate-crimes law relates to the message the proscribed activity conveys. The Court rightly saw flag-burning laws in this light—as an effort, underneath the cover of an incidental restraint, to suppress communication of a message.²⁵² Perhaps the same argument applies to hate-crimes laws; indeed, the Court in *Mitchell*, though upholding such a law, understood it in much this way, pointing to interests the government had in restricting expression of racist messages.²⁵³ But this view of hate-crimes laws is not necessary. The government may have a non-speech-related interest for sanctioning race-based assault, no less than race-based discharge: an interest in eradicating racially based forms of disadvantage—in preventing disproportionate harm from falling, by virtue of status alone, on members of a racial group. Given this interest, existing apart from any speech, the Court correctly treated hate-crimes laws as laws of general application.

speech falls victim to a law banning treason or prohibiting the assassination of the President.

²⁵² For an example whose structure parallels a hate-crimes law, consider a penalty enhancement provision applicable to persons who obstruct voting on the basis of a voter's membership in the Republican Party. In this case too, the government's interest in the law cannot but relate to favoring or disfavoring particular viewpoints. But now consider a broader law applying enhanced penalties to persons who obstruct voting based on the voter's affiliation with *any* political party. Such a law could have emerged from a governmental interest in protecting persons from suffering disproportionate harm as a result of their political views, analogous to the interest I will soon note in protecting persons from suffering disproportionate harm as a result of their race. Accordingly, this law would meet constitutional standards: it applies regardless whether the conduct communicates a message, and the government has a credible interest in the law not related to favoring or disfavoring particular messages.

²⁵³ The Court noted that race-based crimes were more likely than other crimes "to provoke retaliatory crimes, inflict distinct emotional harms on their victims, and incite community unrest." See 508 US at 488. But these effects arise, largely if not entirely, from the expressive component of a race-based crime—the aspect of the activity that communicates, and is meant to communicate, a message. The interest in preventing these effects is of course not illegitimate. But it is an interest related to restricting a message. Thus, if this interest alone lies behind hate-crimes laws, then courts, however they rule on the legitimacy of the laws, should regard them as restrictions of expression.

But this analysis of hate-crimes legislation is no more than an illustration. The basic points I wish to make concern not whether courts ought to regard any particular piece of legislation, including hate-crimes laws, as falling within the category of incidental restraints, but instead what rationale lies behind that category and what limits that rationale sets on the category's boundaries. The doctrine of incidental restraints, like so much of First Amendment law, arises from a desire to flush out impermissible purposes on the part of the government. The reason why incidental restraints usually receive no First Amendment review relates to the low risk that hostility or partiality toward ideas tainted these restrictions. The circumstances in which incidental restraints confront a measure of scrutiny—and the level of scrutiny then employed—also connect to the risk of taint in certain laws. Once again, the doctrine acts as a complex mechanism to provide review where necessary, and of the kind necessary, to invalidate improperly motivated governmental actions.

IV. THE UNDERPINNINGS OF MOTIVE ANALYSIS

The primary task of this Article is to lay out a descriptive theory of the First Amendment—an explanation of the state of First Amendment doctrine, an account of the wellspring of the law. I cannot here proceed much beyond the aims of this project to resolve all of the normative questions this theory raises. In particular, I cannot here provide a full justification for structuring First Amendment law around the question of motive, rather than around the question of effects, whether on a speaker or an audience. But I also cannot conclude this Article without considering briefly what might lie behind the law's focus on motive—why, that is, motive might (or, at least, might be thought to) matter. In this Section, I turn to the normative underpinnings of motive analysis.²⁵⁴

The importance of motive in First Amendment analysis, as in other spheres of constitutional analysis, is in many ways mysteri-

²⁵⁴ Many scholars have discussed in detail the appropriate role of governmental motive in constitutional law, especially in cases involving the Equal Protection Clause. See, for example, Lawrence A. Alexander, *Introduction: Motivation and Constitutionality*, 15 San Diego L Rev 925 (1978); Brest, 1971 S Ct Rev at 95 (cited in note 77); Theodore Eisenberg, *Disproportionate Impact and Illicit Motive: Theories of Constitutional Adjudication*, 52 NYU L Rev 36 (1977); Ely, 79 Yale L J at 1205 (cited in note 40); Michael J. Perry, *The Disproportionate Impact Theory of Racial Discrimination*, 125 U Pa L Rev 540 (1977). For a brief but excellent discussion, focusing on the role of motive in First Amendment analysis, see Tribe, 1993 S Ct Rev 1 (cited in note 78).

ous. Consider two laws with identical real-world consequences, either as to the aggregate quantity of speech or as to the quality of the speech market. An analysis focusing on motive may force us to treat these laws differently, even though they affect us in the same manner. Or consider two laws with diverse impacts, one restricting the sum total of speech more than the other or distorting the sphere of discourse in more dramatic fashion. An analysis focusing on motive may force us to treat these laws identically, even though we experience the one as more confining than the other. To say this much is only to acknowledge the inevitable outgrowth of motive analysis—only to recognize the potential disjunction between the reasons for doing a thing and the results of having done it. The question underlying motive-based analysis is why we should focus in this way on what led to an action, not what the action accomplished.²⁵⁵

One way to approach this issue is to consider the kinds of justifications we usually offer for rules of process. (A prohibition relating to motive is, after all, nothing more than such a rule, operating to exclude certain factors from the decision-making process.) First, we may adopt such a rule because it will promote good consequences, where the criterion for deciding what constitutes a good consequence comes from outside the rule itself. Second, and alternatively, we may adopt a rule of process for its own sake, because it possesses certain attributes or expresses certain norms, the correctness of which renders any outcome it produces correct as well. The distinction is between rules of process whose justification derives from the appropriateness of the results they promote, as independently defined, and rules of process whose justification derives from internal attributes, which themselves define what results count as appropriate.²⁵⁶

²⁵⁵ The famous Holmesian quip respecting the importance of motive tends to glide over this problem. “[E]ven a dog,” Holmes wrote, “distinguishes between being stumbled over and being kicked.” Oliver Wendell Holmes, *The Common Law* 7 (Belknap 1963) (Mark DeWolfe Howe, ed). But the injured dog might think that this distinction, however easy to make, ought not to carry weight. It is one thing to distinguish and another to make the distinction matter in a legal system. The critical question is why this distinction should matter given that the resulting bruise seems the same. Holmes’s epigram is perhaps suggestive of an answer, as I later discuss, but does not itself provide it.

²⁵⁶ The discussion in the preceding paragraph follows John Rawls’s typology of systems of procedural justice. Rawls defines “imperfect procedural justice” as a situation in which procedures gain their correctness from their ability to promote certain independently justified outcomes. He defines “pure procedural justice” as a situation in which procedures lend their own (internal) correctness to the outcomes they produce. See John Rawls, *A Theory of Justice* 85-86 (Belknap 1971). This distinction correlates roughly to one between consequential and deontological justifications for a rule—that is, justifications

Using this framework, two kinds of theories may explain the prevalence in First Amendment law of motive-based analysis. First, courts may focus on motives because doing so will promote good outcomes, as defined by some independent set of criteria. On this theory, the gulf I have presupposed between the reasons for an action and the effects of an action in an important sense closes. The reason to think about reasons has to do with the likelihood that the consideration of certain reasons will systematically and predictably lead to actions that have adverse consequences.²⁵⁷ Second, courts may focus on motives because motives, in and of themselves, are what matter—so much that any actions deriving from improper motives also become improper, by a kind of automatic motion. On this theory, the division I have posited between reasons and effects closes in a different but no less fateful manner. Now, the only effects that matter are the effects (whatever they may be) that emerge from particular reasons for action.²⁵⁸ Both of these theories are as yet mere outlines; the remaining questions concern how to give them content.²⁵⁹

A consequentialist theory of motive analysis must provide an account of how certain motives foster adverse outcomes, as defined by an independent criterion of value. Assume here that we should assess outcomes in the fashion of the audience-based model: the optimal state of public discourse is that most illuminating to and desired by an ideally curious and engaged audience. Now consider how the concept of improper motive I have developed relates to this view of desirable (and undesirable) outcomes. When self-interest or ideological hostility enters into a restriction on speech, the odds increase that the resulting action will impoverish the sphere of public discourse.²⁶⁰ By happen-

that refer to the desirability of producing some independently justified value and justifications that operate autonomously of any particular set of consequences.

²⁵⁷ This closure, I hasten to add, does not undermine the thesis of this Article. As I explain, the distinction between motive-based analysis and effects-based analysis remains all-important for purposes of constructing (and explaining) First Amendment doctrine.

²⁵⁸ The apparent oddity that good motive can save a law with detrimental effects (and bad motive doom a law with beneficial effects) thus disappears. On this account, in its purest form, the presence or absence of impermissible motive itself defines whether effects are detrimental or beneficial.

²⁵⁹ These two accounts can coexist, if not in their purest forms, then in some modified versions. In a combined approach, the insistence on a rule of process would derive from both the value inherent in the rule and the connection of the rule to certain outcomes.

²⁶⁰ See text accompanying notes 100-01. The argument here is that improperly motivated action will tend to distort the speech market, not that content-based action will tend to distort the speech market. I have criticized the latter argument many times, noting that content-based regulation sans bad motive (say, if conducted on a random

stance, some improperly motivated restrictions will enhance, and some purely motivated restrictions will mutilate, the thinking process of the community. But in general, a system in which the government freely may restrict ideas on the ground that they challenge the power or wisdom of officials will produce a less healthy debate than a system in which the government has no such ability. A rule proscribing actions arising from censorial motive thus will promote the set of outcomes that the audience-based model deems desirable.

But even assuming this claim is accepted, the reason to focus on reasons is still not self-evident. If the ultimate question relates to the effects of an action on public discourse, then courts seemingly should make just this inquiry. Asking about improper motive so as to hazard a guess about untoward effects appears a strangely circuitous way of addressing the issue. For the consequentialist theory under discussion to work, an inquiry into audience-based effects must be infeasible; only then might discovering reasons provide the most judicially manageable way to evaluate an action's impact on public discourse.

An argument of this kind is indeed plausible, given the nature of the inquiry into effects and the limits of judicial capacity. The criteria that the audience-based model provides by which to judge the results of a speech regulation may be insufficiently definite and detailed to lend themselves to direct application. Consider that we do not possess a fully developed sense of what an optimal marketplace of ideas would look like.²⁶¹ We have instead a set of hazy generalities (rich, robust, balanced, diverse) by which we refer to the ideal state. These standards, although sometimes of concrete assistance, often cannot tell us whether a governmental action, the very operation of which may be uncertain, will impair or improve public discourse. And even if they always could do so in the hands of, say, Hercules, they seem far too diffuse—far too manipulable—to trust to a judge with her own set of interests and biases. The problem with an effects-based standard is one of judicial administration. The questions it forces judges to ask about what ideas are over- or under-repre-

basis) is as likely to improve as impair the speech market. All I am claiming here is that when the government restricts ideas because officials see them as threatening or distasteful, the action more often will debilitate than strengthen public discussion.

²⁶¹ We also may not possess precise ways to measure or describe the current distribution of ideas, so that we could determine the difference between it and the ideal. For fuller discussion of these issues, see Strauss, 1993 U Chi Legal F at 202-10 (cited in note 20).

sented, about who has talked too much or too little, about when “drowning out” has occurred, are not subject to unbiased, reliable evaluation.

If this is so, the focus on governmental motive that marks First Amendment doctrine may function as a kind of proxy for an inquiry into the effects of a restriction of speech on an audience. This does not mean that motive-based analysis plays a subsidiary role to effects-based analysis or that the two meld into each other.²⁶² To the contrary, it is the search for improper motive that drives the doctrine. That search generates a set of doctrinal rules different from the rules that would flow from a direct effects-based inquiry. Those rules then spawn a set of results different from the results that would emerge from effects-based doctrines. Yet beneath the operation of these rules indeed may lie a concern with consequences. The focus on motive, on this account, provides an indirect way of identifying actions with untoward effects on public discourse. This identification mechanism is necessarily imprecise—both over- and underinclusive. But given the difficulties of inquiring directly into effects, it may be the best such instrument that courts can find.

The notion here should seem familiar, for I have based most of this Article on a similar form of argument. What I have said already goes something as follows: We wish to discover improper motive, but cannot do so by making a direct inquiry; we instead construct a set of rules, turning on the facial terms of legislation, to identify motive indirectly; we realize these rules will prove imprecise, capturing too much and too little, but we use them because we can think of no better way to discover improper motive. What I am positing now adds an anterior, but similarly constructed, line of reasoning to this one. Why do we wish to discover improper motive? Perhaps because we wish to discover adverse effects, but cannot do so directly; because we know that actions tainted with certain motives tend to have such consequences; because although a focus on motive will prove imprecise, we can think of no better way to gauge the effects of an action on the state of public discourse. Hence we emerge with a set of First Amendment rules serving as “double proxies”—first, and more

²⁶² As Larry Alexander has noted, there is a difference between motive theories and effects theories even when, as is often the case, the proscribed motives are “selected because of the social effects associated” with them. Alexander, 15 San Diego L Rev at 931 (cited in note 254). The critical aspect of motive theories is that “in the final analysis, where motives and effects are inconsistent, the motives, not the effects, govern.” *Id.*

proximately, for an inquiry into a certain kind of motive; then, and more remotely, for an inquiry into a certain kind of effect. At each step, something gets lost; the rules function imperfectly to flush out improper motive, still more imperfectly to identify adverse effects. But on this theory, we use the rules because they, better than any others, allow the discovery of improper motive—in order to achieve the discovery of adverse impact.

In contrast to this consequentialist account, the second explanation for the law's focus on motive is purely internal. Here, the reasons underlying a law do not provide evidence of something else, but themselves constitute the ultimate issue. But this raises the question why motives ought so to matter—why motives, for their own sake and irrespective of material consequence, should determine the legitimacy of governmental action.

The answer may begin to emerge if we recognize that two actions having similar material outcomes may express different values and have different meanings.²⁶³ This contrast, I think, is what Holmes meant to highlight when he distinguished between stumbling over and kicking a dog.²⁶⁴ The former may suggest a lack of optimal care, but the latter suggests contempt or hatred.²⁶⁵ The same dynamic between reasons, actions, and meanings often arises. Consider the difference between a policy that intentionally excludes African-Americans from employment and a test that as effectively, but unintentionally, prevents them from gaining desired positions. There are, of course, (effects-based) reasons to treat these actions identically; but it is wholly intelligible to say that the former is worse than the latter because it conveys an attitude of disrespect or malevolence. (Indeed, equal protection law, for this reason, treats deliberate discrimination

²⁶³ See Elizabeth Anderson, *Value in Ethics and Economics* 33-34 (Harvard 1993) (“[T]he distinction[] . . . between foreseeing and intending certain consequences . . . mark[s] [a] distinction[] in the expressive significance of actions.”); Joseph Raz, *The Morality of Freedom* 378 (Clarendon 1986) (Certain acts have “meaning regardless of their actual consequences . . . expressing disregard or even contempt.”); Richard H. Pildes, *Avoiding Balancing: The Role of Exclusionary Reasons in Constitutional Law*, 45 *Hastings L J* 711, 728 (1994) (“How an action comes about shapes what it means and therefore what it is.”).

²⁶⁴ See note 255.

²⁶⁵ The dog also may fear the kick more than the stumble for related, but entirely consequentialist reasons. Even if the bruise from the two is the same, the contempt inherent in the kick increases the likelihood that yet another kick will come and yet another bruise follow. This reasoning suggests a connection between First Amendment law's focus on motive and its amenability to slippery slope arguments. We care about bad reasons because they entail repeated bad results; we fear the occurrence of repeated bad results because we suspect the existence of bad reasons.

having trivial consequences as more problematic than incidental discrimination having great impact.²⁶⁶) An action acquires meaning in part through motive, and the meaning of an action in part defines it. Hence, what an action *is* derives not only from what it does, but from where it comes from; so too, then, whether the action is legitimate involves this matter.

The doctrine of impermissible motive, viewed in this light, holds that the government may not signify disrespect for certain ideas and respect for others through burdens on expression. This does not mean that the government may never subject particular ideas to disadvantage. The government indeed may do so, if acting upon neutral, harm-based reasons. But the government may not treat differently two ideas causing identical harms on the ground that—thereby conveying the view that—one is less worthy, less valuable, less entitled to a hearing than the other. To take such action—in effect, to violate a norm of ideological equality—would be to load the restriction of speech with a meaning that transcends the restriction's material consequence.

The First Amendment's focus on motive, on this account, serves as an analogue in the speech context to the principle that the government must treat all persons with equal respect and concern.²⁶⁷ This principle, which may well explain much of equal protection law,²⁶⁸ holds in part that the government may not treat some persons differently from others because they are deemed less intrinsically worthy. If such impermissible considerations intrude into the decision-making process, the results of that process likewise become improper; this is so even if the same results could have stood had hostility not infected the process.²⁶⁹ So too here, except that the principle of impartiality applies not to persons, but to ideas. In determining whether to restrict speech, the government may not rank the worth or "rightness" of messages; to do so would be to register a kind of disrespect that automatically renders the action improper.

²⁶⁶ See, for example, *Washington v Davis*, 426 US 229, 244-48 (1976).

²⁶⁷ See, for example, Ronald Dworkin, *Taking Rights Seriously* 272-73 (Harvard 1977) (proposing that government may not treat persons unequally "on the ground that some citizens are entitled to more because they are worthy of more concern").

²⁶⁸ See Cass R. Sunstein, *Public Values, Private Interests, and the Equal Protection Clause*, 1982 S Ct Rev 127, 143.

²⁶⁹ We might think of officials in such a case as having violated the rules of a game. It is no excuse to a charge of cheating to say that the same results would be unobjectionable in a game played honestly. The fact of cheating renders the result improper, whatever the result might be. See Rawls, *A Theory of Justice* at 86 (cited in note 256).

But this account has left open a crucial issue: why is it improper for the government, through restrictions on speech, to show contempt for contemptible ideas, independent of the harm they cause? One answer to this question negates its premise: this answer insists that in life, as in law, "there is no such thing as a false idea," neither an abhorrent one;²⁷⁰ or, to put the point somewhat less baldly (in the way Justice Holmes flirted with it), the very, and only, definition of truth and wisdom is what emerges from free discussion.²⁷¹ But this answer entails an extreme skepticism, unacceptable to most of us because incompatible with a host of our considered judgments. It explains the First Amendment principle of equality only by assuming a world of moral indeterminacy (thankfully) impossible to recognize. A second and better answer to the question refers to the probability that the government will err, as a result of self-interest or bias, in separating the true and noble ideas from the false, abhorrent ones; a scheme of neutrality thus provides the surer means to make this distinction. But this answer largely returns us to the consequentialist basis for focusing on motive; again, what is stressed is the connection between distrusting government and achieving the best possible public discussion. What I am trying now to explore is a different rationale for the focus on motive.

Two versions of this rationale seem possible, both referring to the locus of decision-making authority in our political system, but one sounding in terms of individual rights, the other in terms of popular sovereignty.²⁷² In the first, the prohibition of ideological motive, and the correlative duty of equal respect for ideas and their proponents, serves as just one application of a general ban on subjecting people to disadvantage for reasons that do not relate to harm, but instead arise from judgments of moral value (or from official self-interest). The government, in the view of many liberal theorists, cannot disadvantage a person because the way she lives is immoral or repellant, even if it is so—or because others view it as immoral or repellant, even if they do.²⁷³ (Nor

²⁷⁰ *Gertz v Robert Welch, Inc.*, 418 US 323, 339 (1974).

²⁷¹ See *Abrams v United States*, 250 US 616, 630 (1919) (Holmes dissenting); *Gitlow*, 268 US at 672-73 (Holmes dissenting).

²⁷² Richard Pildes has suggested that motive inquiries throughout constitutional law serve "to set the boundaries between separate spheres of authority"—to aid in the "differentiation of political authority that is crucial to liberalism and the Constitution." See Pildes, 45 *Hastings L J* at 713, 715 (cited in note 263). Both of the following accounts can be understood in this light.

²⁷³ See John Stuart Mill, *On Liberty*, in J.M. Robson, ed, 18 *Collected Works: Essays on Politics and Society* 213, 223-24 (Toronto 1977); Ronald Dworkin, *A Matter of Principle*

can the government impose such a disadvantage, in the absence of any harm, merely because public officials thereby would gain.) If this is so, it follows that the government cannot disadvantage a person because what she thinks or says is immoral or repellant or because others view it as such.²⁷⁴ The narrower (speech-related) principle inheres in the broader; both are aspects, so the argument goes, of the appropriate relationship between the government and individuals within a liberal society.²⁷⁵

The second kind of nonconsequentialist account for the prohibition of ideological motive relates more exclusively to expression, emphasizing the place of such activity in a democracy.²⁷⁶ On this view, the prohibition of ideological motive, and its concomitant principle of equality, lies at the core of the First Amendment because it lies at the core of democratic self-government. The democratic project is one of constant collective self-determination; expressive activity is the vehicle through which a sovereign citizenry engages in this process by mediating diverse views on the appropriate nature of the community. Were the government to limit speech based on its sense of which ideas have merit, it would expropriate an authority not intended for it and negate a critical aspect of self-government.²⁷⁷ Democracy demands that sovereign citizens, through each generation, retain authority to evaluate competing visions and their adherents—to decide which ideas and officials merit approval. Hence democracy bars the

191, 203 (Harvard 1985); Raz, *The Morality of Freedom* at 420 (cited in note 263).

²⁷⁴ For the most persuasive statement of this position, see Dworkin, *A Matter of Principle* at 353 (cited in note 273) ("People have the right not to suffer disadvantage . . . just on the ground that their officials or fellow-citizens think that their opinions about the right way for them to lead their own lives are ignoble or wrong.").

²⁷⁵ Joseph Raz has argued that a liberal government may make moral judgments in the absence of harm by means other than coercion. In his words, the harm principle holds that "[w]hile [moral] ideals may indeed be pursued by political means, they may not be pursued by the use of coercion except when its use is called for to prevent harm." See *The Morality of Freedom* at 420 (cited in note 263). In the First Amendment context, such an argument might support a decision to apply purpose analysis more strictly when the government restricts speech than when the government funds speech or speaks itself.

²⁷⁶ The thoughts in this paragraph owe much to the work of Robert Post and Geoffrey Stone. See Robert Post, *Managing Deliberation: The Quandary of Democratic Dialogue*, 103 *Ethics* 654, 660-61 (1993); Robert C. Post, *Racist Speech, Democracy, and the First Amendment*, 32 *Wm & Mary L Rev* 267, 290-91 (1991); Stone, 131 *Proc Am Phil Soc'y* at 253 (cited in note 137).

²⁷⁷ This is true even when the government is acting as representative of the majority of citizens. The notion of democracy used here means something different from simple majoritarian decision making. The notion invokes a continuing, evolving process by which sovereign citizens accommodate and reaccommodate diverse views and, in so doing, create and recreate their society. See Post, 32 *Wm & Mary L Rev* at 279-83 (cited in note 276); Stone, 131 *Proc Am Phil Soc'y* at 253 (cited in note 137).

government from restricting speech (as it also bars the government from limiting the franchise) on the ground that such activity will challenge reigning beliefs or incumbent officials. The government must treat all ideas as contingent, because subject to never-ending popular scrutiny. On this view, the prohibition of certain motives again serves as a way to delineate the proper sphere of authority, hereby preventing a democratic state from contravening key principles of self-government and thereby undermining its foundation.²⁷⁸

All of these explanations go to the question why governmental motive is important. A separate and harder question concerns why motive should be all-important, such that an inquiry into motive precludes an inquiry into effects. I have considered this question briefly in discussing audience-based effects, noting that an inquiry into motive tests these effects better, though more circuitously, than a frontal inquiry could do. I have neglected this question altogether in discussing speaker-based effects.²⁷⁹ But for two reasons, I will not attempt here to address this question further. First, I have never claimed that current law wholly excludes an inquiry into effects; I have claimed that motive plays the dominant, not the exclusive, role in the doctrine. Second, I have never proposed to show that the most sensible system of free expression would focus on issues of governmental motive to the extent our system does, let alone to the exclusion of all others. I have posited only that our system of free expression focuses on motive—and, to buttress that claim, noted the normative commitments that underlie this election. I leave for another day the question whether our doctrine, in attempting to discover improper motive, has neglected too much else of importance.

CONCLUSION

In his opinion for the Court in *R.A.V.*, Justice Scalia distinguished the case before the Court from several of his own inven-

²⁷⁸ An argument of this kind also might support applying motive analysis differently to governmental efforts to restrict private speech and governmental efforts to participate in the speech market. See note 275. It is arguable (though subject to many limitations and exceptions) that the government does not expropriate a sovereign citizenry's ultimate authority to decide which ideas are worthy when the government acts not to limit a debate, but to engage in it.

²⁷⁹ One explanation for relegating speaker-based effects to a secondary position is that we can and do expect the political process, in the absence of impermissible motive, to protect against inordinate restrictions on expressive opportunities.

tion.²⁸⁰ If the case had involved secondary effects, Justice Scalia noted, a different issue would have been presented. So too if the case had involved an incidental restriction. In these and other circumstances, Justice Scalia wrote, the Court would have countenanced governmental action burdening only certain messages. The concurring opinions in the case castigated Justice Scalia's list of distinctions and exceptions. Justice Stevens wrote that the Court had "offer[ed] some ad hoc limitations" on its holding in order to contain the "perversities" it would engender.²⁸¹ Justice White similarly noted the Court's effort to "patch[] up its argument with an apparently nonexhaustive list of ad hoc exceptions."²⁸² But Justice Scalia's opinion did contain a rationale for his catalogue of distinctions. What he was trying to separate from the St. Paul ordinance, the Justice intimated, were laws that, although imposing differential burdens on ideas, "refute[d] the proposition" that they were "even arguably 'conditioned upon the sovereign's agreement with what a speaker may intend to say.'"²⁸³ In essence, Justice Scalia was pointing toward a set of rules and categories (whether he got them exactly right is not what matters) that attempt to sort out, even if implicitly, regulations based solely on neutral determinations of harm from regulations tainted with ideological motive.

The reaction of Justice White and Justice Stevens is not surprising—nor is the more general sense of rebellion against the increasingly technical, complex classificatory schemes of First Amendment law, which Justice Scalia's opinion highlighted. One commentator wrote more than a decade ago of the "elaborate codification of the First Amendment," warning against "excess categorization."²⁸⁴ Since then, the doctrine has become only more intricate, as categories have multiplied, distinctions grown increasingly fine, and exceptions flourished and become categories of their own. Little wonder that Justice Stevens could refer to the *R.A.V.* Court's "adventure[s] in . . . doctrinal wonderland."²⁸⁵ Or that a recent commentator on Supreme Court jurisprudence, including *R.A.V.* and other First Amendment cases,

²⁸⁰ 505 US at 388-90.

²⁸¹ *Id.* at 423 (Stevens concurring).

²⁸² *Id.* at 407 (White concurring).

²⁸³ *R.A.V.*, 505 US at 390, quoting *Metromedia, Inc. v City of San Diego*, 453 US 490, 555 (1981) (Stevens dissenting in part).

²⁸⁴ Frederick Schauer, *Codifying the First Amendment: New York v. Ferber*, 1982 S Ct Rev 285, 288, 316.

²⁸⁵ 505 US at 418 (Stevens concurring).

could criticize the Court for “devolv[ing] into conceptualism and technicality” and exhibiting “an almost medieval earnestness about classification and categorization.”²⁸⁶

But before we dismiss this conceptual scheme, we should at least explore what lies behind it. This means more than engaging in the now familiar debate about the relative merits of rules and standards. It means more than asking in the abstract whether formalism might have its uses. It means, instead, examining in concrete terms how these rules function and what they accomplish. That is what this Article has tried to do; for only when we know why the doctrine has emerged and what purposes it serves will we know whether and how to modify it.

What I have argued in this Article is that most of First Amendment doctrine constitutes a highly, but necessarily, complex scheme for ascertaining the governmental purposes underlying regulations of speech. The Court could not—and knows it could not—discover these motives through direct inquiry; in all but the most unusual case, the government could offer a permissible reason for its action, and the Court could not tell whether this reason was real or pretextual. Hence, the Court (whether consciously or not is unimportant) has constructed and relied upon a set of rules and categories, focusing on the facial aspects of a law, that operates as a proxy for this direct inquiry. Because these rules operate at a step removed, they are both over- and underinclusive. But they do well, if not perfectly, what could not be done in their absence—ferreting out and then invalidating impermissibly motivated governmental actions. The categories, the distinctions, and the rules of First Amendment law thus have a rationale and purpose not immediately apparent; if courts could evaluate motive directly, they could remove the lion’s share of the First Amendment’s doctrinal clutter.

The presence of this underlying principle, explaining and rationalizing First Amendment doctrine, does not make the doctrine self-evidently correct. We may believe that the doctrine cares too much about motive or that it cares too little about other things. But the principle does make the doctrine internally consistent and coherent. If the current doctrinal formulations are wrong, they are largely wrong as a whole and for the same reasons. And those reasons would relate—as most of the law re-

²⁸⁶ Morton J. Horwitz, *The Supreme Court, 1992 Term—Foreword: The Constitution of Change: Legal Fundamentalism Without Fundamentalism*, 107 Harv L Rev 30, 98 (1993).

lates—to the decision to treat the question of governmental motive as the preeminent inquiry of the First Amendment.