

The Case for Refundable Tax Credits

Declaring our current system inefficient, **LILY BATCHELDER** proposes that incentives for charitable giving and other socially valued behaviors be equal for all taxpayers.



EACH YEAR THE FEDERAL INDIVIDUAL income tax code provides over \$500 billion worth of incentives intended to encourage socially beneficial activities, such as charitable contributions, homeownership, health insurance and education. This is an enormous investment, exceeding our budget for national defense and amounting to about four percent of GDP. The design of these tax incentives is therefore an immensely important policy matter. Yet despite their

efficiency rationale, little attention has been paid to the question of what economic efficiency implies about the form these tax incentives should take.

Currently the vast majority of tax incentives operate through deductions or other approaches that link the size of the tax break to a household's marginal tax bracket, which means that higher-income taxpayers receive larger incentives than lower-income taxpayers. Such an approach is often appropriate for provisions, such as deductions for business expenses, designed

to measure income or ability to pay. But we argue that it is inefficient for incentives intended to promote socially valued activities unless policymakers have specific knowledge that such households are more responsive to the incentive or that their engaging in the behavior generates larger social benefits. Absent such empirical evidence, all households should face the same set of incentives.

Purely on efficiency grounds, we therefore propose a dramatic change in how the government should provide tax incentives for socially valued activities. Under our proposal, which could be implemented on a revenue-neutral basis, the default for all such tax incentives would be a uniform refundable tax credit. These tax credits would be available to qualifying households even if they owe no income tax and would provide a much more even and widespread motivation for socially valued behavior than the current set of tax incentives. Moreover, they could further enhance economic efficiency by smoothing out fluctuations in household income and macroeconomic demand.

COMPARING TAX INCENTIVES

Currently approximately \$420 billion of the \$500 billion of tax incentives in the individual income tax code operate through deductions, exemptions, exclusions, or nonrefundable credits. Such tax incentives tie the size of the tax break to an individual's marginal tax bracket: A deduction of \$1, for example, is worth 35 cents to someone in the 35 percent marginal bracket but only 15 cents to someone in the 15 percent marginal bracket. Furthermore, these types of tax incentives fail to reach the increasingly significant share of low- and moderate-income individuals and families who do not have any federal income tax liability to offset in any given year. More than 35 percent of households during any given year have no income tax liability. These households are home to almost half of all American children.

Refundable tax credits represent a different approach. Since they are a credit, rather than a deduction or exclusion, they do not depend on a household's marginal tax bracket. A tax credit of \$1, for example, reduces taxes by \$1 and thus is worth the same to households in the 35 percent bracket or the 15 percent bracket. And since they are refundable, they provide benefits to all tax filers, regardless of whether they owe income taxes on net.

The tax code presently contains three main refundable tax credits: the Earned Income Tax Credit, the Child Tax Credit, and a small health insurance credit. The Earned

Income Tax Credit is the largest antipoverty program for the nonelderly in the country. In inflation-adjusted terms, the budgetary cost of the EITC has risen by a factor of nine since it was enacted in 1975, and it tripled between 1990 and 2000 alone. More recently, the partially refundable Child Tax Credit and a small, fully refundable health insurance credit were enacted, and the refundability of the Child Tax Credit was expanded and accelerated.

Refundable credits have grown substantially over the past three decades. In part, this is the result of policymakers' increasing reliance on the tax code, rather than direct government expenditures, to subsidize households and influence their behavior as a result of perceived or real incentives within the tax legislative process. Despite this growth, however, the vast majority of tax incentives still take the form of deductions or other approaches linked to marginal tax rates.

THE EFFICIENCY ARGUMENT

The efficiency benefits of refundable credits can be understood both intuitively and theoretically. Intuitively, if policymakers want to broadly promote socially valued behavior through the tax code, refundable credits are generally necessary. As illustrated at right, in any given year more than one-third of households do not have any federal income tax liability. About a quarter of tax units file a tax return but have no income tax liability, and another 13 percent do not file. Moreover, almost half of all children, and 80 percent of children in single-parent households, live in tax units with no income tax liability in any given year.

As a result, if policymakers want to create incentives through the individual income tax for all or most tax units to engage in certain behavior every year, such as saving or obtaining education for themselves or their children, refundability should not only be considered an acceptable instrument of tax policy—it is imperative.

At a more theoretical level, unless there is evidence that certain households are more responsive to the incentive than others or generate larger social benefits from engaging in the activity, tax incentives are most efficient if they provide the same incentive to all households—and that can only be accomplished in a straightforward manner through a uniform (and refundable) credit.

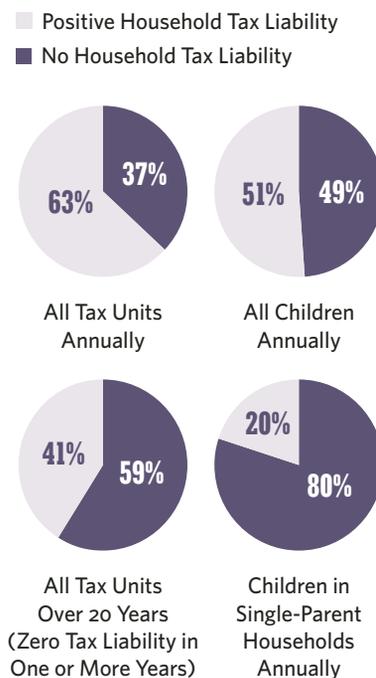
The reason that a uniform incentive is the most efficient approach in the absence of evidence regarding differences in responsiveness or social benefits is that a

small number of large mistakes in under- or oversubsidizing an activity is more costly in efficiency terms than a large number of small mistakes. Stated more technically, a uniform subsidy is the most efficient in the absence of such empirical evidence because the deadweight loss arising from an uncorrected externality rises with the square of the uncorrected externality. For example, imagine that certain behavior, perhaps charitable contributions, on average generates five cents of social benefits per dollar contributed per year and policymakers have determined to subsidize contributions by, on average, five cents per dollar. Imagine further that there is a 50 percent chance that a dollar of contributions by a high-income household generates 10 cents of social benefits, while a dollar of contributions by a low-income household generates none, and a 50 percent chance that this pattern is reversed. A uniform subsidy of five cents would leave five cents of lost social benefits in both cases. Meanwhile, a subsidy of 10 cents given to one group would result in 10 cents of lost social benefits in one case and none in the other. The uniform subsidy is more efficient—it minimizes the expected deadweight loss—because a small number of big errors (one case of 10 cents) is more costly in efficiency terms than a large number of small errors (two cases of five cents).

We acknowledge that many tax incentives may be bad policy regardless of whether they take the form of uniform refundable credits, perhaps because the behavior in question does not actually generate social benefits or because such social benefits are best addressed through direct government provision of the good or regulation. Even taking these limitations into account, however, assuming the continued existence of a tax incentive, our default structure is generally preferable. It minimizes the expected deadweight loss not only when a tax incentive undercorrects for an externality, but also when it overcorrects, either because the subsidy exceeds the social benefits generated or because the behavior actually is not socially beneficial.

We also acknowledge that tax incentives should not provide the same incentive to all households in all circumstances. If there is evidence that the associated social benefits vary systematically by income class, or that different income groups exhibit different levels of responsiveness to the subsidy, the tax incentive should not be identical for all households. Indeed, these differences between various income groups surely exist in

Federal Income Tax Liability



SOURCES: PETER R. ORSZAG & MATTHEW G. HALL, NONFILERS AND FILERS WITH MODEST TAX LIABILITIES, 100 TAX NOTES 723 (2003); AUTHORS' CALCULATIONS BASED ON A SIMPLE MODEL OF 2003 LAW AND LONGITUDINAL EARNINGS DATA FROM THE PANEL SURVEY OF INCOME DYNAMICS.

reality. But when, as is frequently the case, the evidence on these issues is nonexistent or inconclusive, the most efficient form for a tax incentive is a uniform refundable credit. The burden of proof should therefore be on those who prefer some other form of tax incentive to demonstrate that such deviations from a uniform refundable credit are justified by empirical evidence.

Even when there is empirical evidence suggesting that the optimal tax incentive should not be the same for all households, the most efficient incentive is almost certainly still some type of refundable credit. It is extremely unlikely that there is a sharp break in social benefits or responsiveness to an incentive exactly at the point of zero income tax liability, given that this threshold varies by family size and year-to-year. Yet these types of discontinuities are inherent in the application of all other basic forms of tax incentives.

Thus, if policymakers wish to use the tax system to create incentives for certain socially valued behavior, it makes no sense to exclude more than a third of American individuals and families from their reach,

or to provide smaller benefits to some households than others, absent evidence that those Americans would be relatively unresponsive or that their behavior generates fewer societal benefits.

ADDITIONAL BENEFITS

The potential efficiency benefits of uniform refundable credits are magnified further by a second feature: their ability to help smooth household income. That is, transforming existing tax incentives into uniform refundable credits would boost after-tax income during hard years, and thus help to cushion the blow of a drop in earnings, unemployment, or other hardships. Such income smoothing is desirable from an efficiency perspective for several reasons. It can reduce the costs associated with economic instability and offset failures in insurance markets. It also allows families to plan their expenditures more confidently and avoids the additional costs (moving costs, credit card debt) of financing constant changes in household living standards. Income smoothing is particularly beneficial for lower-income households because they generally don't have easy access to credit to make it through tough times, because they tend to have more volatile incomes than other families in general, and because income shocks can result in declines in their economic circumstances that persist over long periods of time and are passed on to their children.

A final efficiency benefit of uniform refundable credits is their ability to smooth

The U.S. spends 4% of GDP subsidizing socially valued activities through the tax code.

the macroeconomy. Like household income smoothing, macroeconomic smoothing can enhance economic efficiency. In particular, macroeconomic demand fluctuations make it difficult for companies to optimize their investment and production functions, resulting in adjustment costs. These difficulties can inhibit domestic and foreign investment, and thus economic growth. As a result, there is broad consensus in support of taxing and spending policies that are automatically countercyclical. Uniform refundable credits can help stabilize macroeconomic demand fluctuations by raising cash payments to families during recessionary periods, which then helps boost spending—precisely the desired response during such periods.



"I suppose one could say it favors the rich, but, on the other hand, it's a great incentive for everyone to make two hundred grand a year."

OPPOSING ARGUMENTS

Opponents of refundable credits typically raise four main objections to our proposal that the default structure for all tax incentives should be a uniform refundable credit. First, some question the extent to which government should engage in redistribution between different income groups. Second, some argue that the tax system should be used only to raise revenue, not to provide subsidies. Third, some believe that all Americans should pay at least some

tax, even if just one dollar, as a duty of citizenship and so that they feel some stake in governmental decisions. Finally, some argue that refundable credits would increase administrative and

compliance costs on net and are particularly subject to fraud and abuse.

Concerns about the extent of governmental redistribution, however, do not justify rejecting refundable credits that are enacted to enhance economic efficiency by subsidizing socially beneficial behavior. And concerns about delivering subsidies as a tax benefit instead of a transfer are generally objections to tax incentives overall, not to structuring tax incentives as refundable credits specifically.

The third objection—that all Americans should pay some tax—ignores the fact that most households claiming refundable credits pay a variety of federal, state and local taxes other than income taxes. Moreover, if one is interested strictly in federal

income taxes, it seems likely that most refundable credit beneficiaries pay a positive amount of federal income tax over time as a result of the income variations that people tend to experience over their lives. Indeed, a simplified model of 2003 federal income tax law using data from the Panel Survey of Income Dynamics suggests that about three quarters of tax units who are eligible for the refundable element of the EITC or CTC at some point during a 20-year period would nevertheless have positive net federal income tax liability over that period if historic earnings patterns are any guide. Thus, even if one accepts the principle that paying some income tax is necessary for feeling a stake in government decisions (which we do not), this principle would not necessarily preclude refundable credits once income tax liabilities are examined over longer time periods.

The final objection to refundable credits is that they could increase fraud and related compliance problems. Yet there is no reason in theory, and no empirical evidence in practice, why there should be a "cliff effect" in fraud precisely at the point of positive income tax liability. If anything, fraud may be easier to hide when it comes in the form of a deduction or exclusion, which reduces taxable income, as opposed to a refundable credit. Instead, reducing fraud and related compliance problems for all tax incentives, including refundable credits, requires structuring the incentives simply, relying on third-party reporting, and investing in enforcement staffing.

We recognize that increasing the prevalence of refundable credits may create incentives for tax units who are currently nonfilers to begin filing, thereby increasing administrative costs for the government and compliance costs for these households. These costs are real and should be taken into account. Nevertheless, they should not be overstated. Currently only about 13 percent of tax units are nonfilers. As a result, nonfilers represent a relatively small share of the roughly 70 to 80 percent of tax units who stand to gain from structuring tax incentives as uniform refundable credits. Moreover, all tax incentives are elective and, even for nonfilers, the administrative and compliance costs associated with claiming them are often likely to be swamped by the dollar value of the credit.

CONCLUSION

Uniform refundable tax credits are the most efficient structure for a tax incentive to encourage desired behavior when, as frequently occurs, evidence of how the desired behavior and its associated social benefits vary across the income distribution is unavailable or inconclusive. Indeed, refundable tax credits are generally the only way to ensure a tax incentive reaches the roughly two-fifths of tax units with no positive income tax liability in a given year. These efficiency benefits are magnified by the ability of refundable credits to help smooth income at a household level and by their ability, to a greater or lesser extent, to bolster the role of the tax system as an automatic stabilizer of macroeconomic demand. The United States spends almost four percent of GDP each year subsidizing socially valued activities through the tax code. Our proposal would dramatically improve the effectiveness and fairness of this substantial investment. □

LILY BATCHELDER is an assistant professor of law and public policy. She joined the faculty in 2005 from Skadden, Arps, Slate, Meagher & Flom, where she focused on tax aspects of transactional matters and tax policy issues arising before the Treasury Department, IRS and Congress. Her research centers on tax incentives, social insurance, income volatility and wealth transfer taxation. She is a member of the National Academy of Social Insurance. This excerpt, taken from "Efficiency and Tax Incentives: The Case for Refundable Tax Credits" by Lily L. Batchelder, Fred T. Goldberg Jr. & Peter R. Orszag, was published in the 2006 Stanford Law Review. When possible and appropriate, please cite to that version.

Reconciling Trademark Rights and Freedom of Expression

ROCHELLE COOPER DREYFUSS describes how strengthened trademark protections, such as the 2006 Dilution Revision Act, are on a collision course with technology and consumer behavior.



ON OCTOBER 6, 2006, PRESIDENT Bush signed the Dilution Revision Act, putting trademarks and free speech on a collision course. Classic trademark law regards trademarks as signals about price and quality. It protects them against the likelihood of consumer confusion in order to motivate producers to create goodwill and to reduce consumer search costs. Since allusive and referential uses of trademarks do not engender marketplace confusion, they have long been considered outside the scope of the right. The new dilution law's concern with the likelihood of "tarnishment" and "blurring" is thus a departure from traditional principles. Its ban on a wide swath of non-

confusing uses responds to the perception that marks now have intrinsic commercial value (as, for example, status symbols). Further, the new right facilitates "lifestyle marketing"—techniques that allow trademark holders to leverage their reputations from one product category to another and across geographic markets.

But just as trademark holders see greater potential in their trademarks, so too do consumers. Signifiers drawn from mythology, history, and literature are losing their potency in a world in which the populace lacks a shared vocabulary and a common intellectual tradition. Well-known marks have, to some extent, taken their place. Exploited as metaphors, similes, and metonyms, trademarks are becoming the lingua franca

of the communicative sphere. Likewise, trademarks play an expressive role on the Internet. Consumers “google” marks to find particular producers and producers use devices such as keywords, metatags, and pop-up ads to attract purchasers. Used as domain names, trademarks can draw visitors to websites for purposes ranging from political to parodic.

The result is a complicated picture. Images and trade symbols are increasing in cultural significance at exactly the time when protection is expanding. The exigencies of a global, online marketplace make stronger protection for trademarks necessary just when technology makes their widespread expressive use more desirable. Internet shopping requires both exclusivity and unrestricted availability—the former, to allow consumers to search effectively; the latter to permit cybermarkets to work efficiently. As the commercial/expressive duality of marks’ meanings becomes salient, so too does the expressive/commercial du-

ality of their use: a trademark-emblazoned T-shirt may be expressing a political point, but it is also a source of profits—profits that derive from the trademark but which can be channeled back into efforts to destabilize its meaning.

from cognitive and behavioral research can lead us to a legal regime that provides better protection for the interests of trademark holders and expressive users alike.

DOCTRINAL APPROACHES

STATUTORY LAW

Each of the three forms of trademark law (rights against passing off, dilution and cybersquatting) creates tools to balance the interests of trademark holders against the interests of expressive users; each tool is, however, of limited value.

Distinctiveness. The primary safeguard for speech interests lies in the distinctiveness requirement: marks are unprotectable if they are understood to be merely expressive or merely functional. In earlier work, I suggested a departure from the traditional all-or-nothing approach to distinctiveness and posited a concept of “expressive genericity” that would give proprietors marketing control—exclusive rights over signaling value—but leave other aspects—expressive

value—with the public. In fact, significant developments have occurred along precisely these lines. In the long run, however, this approach has proved incapable of dealing with today’s increasingly complex linguistic marketplace.

Actionable use. Another way to preserve expressivity is to refine the definition of wrongful use. Statutes variously articulate this requirement as “using in the course of trade,” “use in commerce,” or “commercial use in commerce”—phrases that could distinguish use to sell from use to persuade, entertain, affiliate, or navigate the Internet.

Michelin & Cie v. C.A.W.-Canada is an example of a case decided using this approach. In a unionization effort, CAW circulated pamphlets showing Bibendum, the famed Michelin tireman, crushing workers. Michelin’s suit for infringement failed, the court reasoning that even though the union stood to earn dues, the use was not commercial: “CAW is competing for the hearts and minds of...Michelin’s employees, not its customers.” Similarly, in *Mattel, Inc. v. MCA Records, Inc.*, a challenge to Aqua’s “Barbie Girl” song, Judge Kozinski limited the concept of commercial speech to speech that “does no more than propose a commercial transaction.” Since Aqua used

Barbie to lampoon the image of the ideal woman, and not to induce doll purchases, it escaped liability.

But despite cases like *Michelin* and *Mattel*, most courts reject this approach and equate “use in commerce” with “commercial use.” As the European Court of Justice put it in the case of *Arsenal Football Club Plc v. Reed*, a use that “takes place in the context of commercial activity with a view to economic advantage” is “use in the course of trade.”

Harm. Classically, the requirement of consumer confusion provides robust protection for both humorous and political usages. *Hormel Foods Corp. v. Jim Henson Productions, Inc.* is illustrative. The case concerned Hormel’s objection to a Muppet named Spa’am. The court declined to find infringement, reasoning that “consumers of Henson’s merchandise...are likely to see the name ‘Spa’am’ as the joke it was intended to be.” Some Internet cases are similarly decided. Thus, attempts to bar the use of marks on gripe sites (of the trademarkholdersucks.com variety) are sometimes rejected on the theory that viewers understand that these sites are unrelated to the trademark holder.

Unfortunately, the confusion tool cannot safeguard all expressive uses. Many courts believe that dissipating confusion at the point of sale is not the only goal: that even if confusion is eliminated by the time of purchase, “initial interest” confusion can also be harmful. Likewise, some courts consider “post-sale” confusion problematic because the purchaser’s unauthorized use could adversely influence those who view the goods when they are in use. Many courts also consider disclaimers ineffective, especially on the Internet where they may be misunderstood by foreign viewers.

As important, the new trademark torts, anti-cybersquatting and dilution, have different standards for harm. Anti-cybersquatting law is nominally aimed at preventing “bad faith.” But bad faith, like commercial use, is in the eye of the beholder. In one sense, dilution is quite narrow in that it protects only marks that are famous. However, this constraint is not much of a safeguard for speech, because famous marks are the ones that expressive users are most interested in utilizing. Further, because it suggests that all the value in a mark belongs to the trademark holder, dilution law strengthens the view that the goal is to eliminate free riding. While U.S. courts have proved fairly skeptical of this right of action, the revision President Bush just signed eliminates virtually all the ways in which claims have been cabined.

“I’m a Barbie girl, in my Barbie world Life in plastic, it’s fantastic...”

“BARBIE GIRL,” BY AQUA, ON AQUARIUM (MCA RECORDS, 1997)



"There's a Mr. Egg McMuffin here who says we've been using his name without permission."

Defenses. Certain high social-value uses, such as news reporting and commentary, are exempt from infringement liability. There are also exemptions for uses that play a unique role in the defendant's speech. Comparative advertising is one example—without using a rival's mark it is impossible to compare products. Similarly, a mark can be "fairly used" to describe such things as repackaged goods, replacement parts, and accessories for trademarked products.

These permitted uses are, however, highly circumscribed and usually require the user to persuade the court that the use is "fair" or "in good faith." The defenses may not cover expressively important users, such as parodists, who intentionally exploit the audience's understanding of the mark (and may even hope to cause a frisson of confusion).

CONSTITUTIONAL APPROACHES

In cases where statutory limitations fail, constitutive norms can be invoked to protect free speech concerns. For example, in *Laugh It Off Promotions CC v. SAB International* the Constitutional Court of South Africa considered a T-shirt parody of the Carling Beer mark, in which "Black Labour" replaced the term "Black Label." Laugh It Off could not avail itself of a noncommercial use defense because its business was selling parodic items. Nonetheless, the Court looked at the use through "the prism" of constitutional values and held it noninfringing.

The constitutional approach is inherently flexible and protects a broader range of uses. Relying on constitutional protection is, however, risky, as decisions are notoriously unpredictable. Also, courts do not always see trademark law as raising constitutional problems. For example, although courts understand that countenancing trespass in certain fora may be necessary to protect speech in real space, they have yet to appreciate the expressive implications of barring "trespass" (unauthorized use) of trademarks in information space.

A NORMATIVE ASSESSMENT

Given how deeply trademark rights can intrude on speech, particularly on the Internet, it can be hard to understand the lack of robust protection for expressive use. Interest group politics is, of course, one obvious culprit: trademark holders are better lobbyists than user groups. But standard public choice theory cannot fully explain this phenomenon. To some extent, conflicting uses are a wash—the Internet consumer who searched for merchant A but decided to buy from merchant B is canceled by the consumer who searched for B and found A.

Interestingly, many of the Internet cases reaching results antithetical to free-speech values are adjudicated in French courts and involve the activities of Americans. Here the issue may be cultural dominance—a particular concern on the Internet, where so much pop culture is disseminated. If

that is the problem, one solution would be to segment the Internet so that each nation can protect its own citizens as it sees fit. But it would be a mistake to support such an approach. Barring worldwide use would destroy a significant part of the Internet's value. At the same time, allowing the French view to dominate cripples navigation on the Internet and creates a solution that is out of all proportion to the problem.

Another possible reason for these outcomes is normative error. Some lawmakers see trademarks as a firm asset akin to real property, and thus consider every unauthorized use a free ride. Intellectual property law is not, however, about preventing free rides; evidence that does no more than show that a defendant enjoyed economic benefits has not, traditionally, sufficed to establish infringement. In the words of the *Laugh It Off* court, a parody can be a "take-off, not a rip off."

Finally, there is aspirational error. Lawmakers appear eager to eliminate all sources of confusion and dilution, and all unauthorized navigation. As the *Arsenal* court put it:

"For a trade mark to be able to fulfill its essential role in the system of undistorted competition..., it must offer a guarantee that all of the goods or services bearing it have been manufactured or supplied under the control of a single undertaking which is responsible for their quality."

It is clear, however, that this goal is unattainable in modern marketplaces. Because trademarks are territorial, the same mark may be used on similar goods in different trading regions; as regions merge, overlapping uses proliferate. Functionality convergence means that goods that were once different (e.g., computers and phonograph records) have, over time, become similar. If they used the same mark (e.g., Apple), there will now be ambiguity. The exceptions to infringement discussed above also result in some dual usages. Most important, not everyone is equally discerning—there will always be consumers who are confused. The time, therefore, has come to learn more about ambiguity—how it is caused; how it affects consumers; and, most important, how it is alleviated.

COGNITIVE AND BEHAVIORAL APPROACHES

Trademarks are not the only place where conflicting signals create confusion. I have two friends called Graeme; I live in Greenwich Village, shop in Greenwich, Connecticut, and visit friends on Greenwich Avenue (or do they live on Greenwich

Street?). Somehow, I manage to sort things out. In recent years, cognitive scientists have begun to study how that happens.

Admittedly, cognitive research supports certain concerns of trademark holders. The prevailing “activation theory” suggests that information is stored in “nodes” hierarchically arranged in “links.” The stronger each node and the stronger the links, the better information is retained and the quicker it is retrieved. Heavy advertising can turn a trademark into a strong node, facilitating quick recall. Indeed, some trademarks become so engrained in memory, they “dominate”—they stand for an entire product category (instance dominance) or are recalled whenever a product category is mentioned (category dominance). Instance dominance accounts for the use of trademarks as Internet search terms, but category dominance is particularly prized. Because information-processing is bounded, consumers display “satisficing” behavior: they may buy as soon

as they encounter the dominant brand, even if the choice fails to optimize their preferences.

ucts do not lend themselves to lifestyle marketing. Dilution may, indeed, be more of a theoretical problem than a real one. Even with confusion, arguably the wrong question is being asked: rather than try to eliminate confusion, the better approach is to ask how to deal with it. Cognitive research reveals various clues. First, consumers are not always confused by unfamiliar associations. Individuals monitor the activation process as it unfolds. When successive nodes fail to recall the source of an association, they eventually conclude that the association is unknown. To quote Donald Rumsfeld, “There are things we know we know. We also know there are known unknowns.” Yes, the moment of uncertainty may be frustrating. But it may also be crucial to the smooth operation of the marketplace. It is at that juncture that consumers seek out more information and learn about other considerations that should influence their purchasing decisions. Law that cuts off learning

about how visual information is integrated, what sorts of messages attract attention, and which are retained. With globalization of the marketplace, it would also help to study how people react when encountering text they do not understand. Are they confused or do they simply move on to products and websites they do understand?

In the final analysis, what trademark law especially needs is a better account of the reasonable consumer. Cognitive styles differ. If there is mobility within the various styles, then the law could be designed to encourage individuals to develop their facilities to focus and discern. Not all ambiguity will be eliminated in any event; creating incentives to deal with it will, in the long run, provide trademark holders with surer protection and give greater freedom to those who use marks expressively.

CONCLUSION

In an economy in which consumers have immediate access to products and services everywhere on the globe, in a legal environment in which symbols are protected in multiple ways, in a culture in which trademarks constitute a significant medium of expression, freedom from all sources of confusion or dilution is simply not achievable. What can be achieved is a marketplace in which consumers can decipher what they are experiencing. Legislation that is attentive to how encounters with multiple meanings are handled would protect both trade and creativity. A fuller understanding of how perception is shaped is likely to provide more durable protection for our shared expressive vocabulary. □

ROCHELLE DREYFUSS is the Pauline Newman Professor of Law. She holds a B.A. and an M.S. in chemistry and worked for several years as a research chemist before entering law school. She is the director of the Engelberg Center on Innovation Law and Policy and a past chair of the Intellectual Property Committee of the American Association of Law Schools. She has served as a member of two National Academy of Sciences committees and is currently a reporter for the American Law Institute project on Intellectual Property: Principles Governing Jurisdiction, Choice of Law, and Judgments in Transnational Disputes. This excerpt is from her chapter, “Reconciling Trademark Rights and Expressive Values: How to Stop Worrying and Learn to Love Ambiguity,” in *Trademark Law and Theory: A Handbook of Contemporary Research* (Graeme B. Dinwoodie and Mark D. Janis eds., Edward Elgar Press, 2007).

“I’m Barbie. No last name.... I sign it like this....’ She picks up Alice’s ballpoint pen and writes a carefully looped, upward slanting ‘Barbie™.’”

PIGS IN HEAVEN BY BARBARA KINGSOLVER (HARPERCOLLINS, 1993)

as they encounter the dominant brand, even if the choice fails to optimize their preferences.

Trademark holders are also right to worry that their marks could lose power. Consumers presented with extraneous information experience a “fan effect”—chains of nodes are activated and information retrieval slows. If confronted with too many choices, consumers may forgo purchase. Worse, when confronted with unauthorized associations, they may encode, and rely on, inaccurate information.

There are, however, strong countervailing considerations. Trademark holders who advertise marks to the point where they become dominant should, perhaps, be required to accept the risk that the marks will be used for expressive and navigational purposes. In an analogous context, the Court in *KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc.* held that “if any confusion results, that is a risk the plaintiff accepted when it decided to identify its product with a mark that uses a well known descriptive phrase.” In fact, less dominance could be in trademark holders’ own interest: marks that become too associated with specific prod-

ucts do not lend themselves to lifestyle marketing. Dilution may, indeed, be more of a theoretical problem than a real one.

imperils competition. It even undercuts the core rationale underlying trademark law: if a trademark holder can end the search, it has little need to further maintain goodwill. Second, cognitive research largely bears out the intuitions of the courts that are receptive to expressive interests. When confronted with ambiguity, people rely on their priors, just as the *Henson* court assumed when it found that experience with the Muppets would resolve consumer confusion regarding Spa’am. Contextual clues are extremely important, as the *Matel* court reasoned when it decided that people hearing a song about Barbie would understand it as entertainment, and not an offer to sell a doll. People use the gestalt of their experiences, just as the *Michelin* court imagined when it held that employees who were handed union pamphlets would not be thinking about buying tires.

But much remains to be learned. One issue is the effect of disclaimers. Although trademark courts discount their value, information is often conveyed through similar devices. Indeed, warning labels, washing instructions, and warranties are often required by law. It would be useful to know more

Separation of Parties, Not Powers

Arguing that the central dynamic of government is competition between political parties rather than between branches of government, **RICHARD H. PILDES** reenvision the law and theory of separation of powers.



DESCRIBING IN THE FEDERALIST Papers a set of “wholly new discoveries” in the “science of politics” that might enable democratic self-government to succeed in the American republic, Alexander Hamilton listed first the “balances and checks” that distinctively characterize the American system of separation of powers. In James Madison’s ingenious scheme of separated powers described in Federalist 51, “the interior structure of the government” would be “so contriv[ed]...as that its several constituent parts may, by their mutual relations, be the means of keeping each other in their proper places.” By institutionalizing a differentiation between executive and legislative powers (as well as by dividing the legislature bicamerally), the separation of powers would harness political competition into a system of government that effectively

checks, balances, and diffuses power. What is more, the system would be self-enforcing, relying on the mechanism of interbranch competition to police institutional boundaries and prevent tyrannical collusion. As Woodrow Wilson famously put it, in the Framers’ Newtonian vision, the separation of powers was to be “a machine that would go of itself.”

To this day, the idea of building self-sustaining political competition into the structure of government is frequently portrayed as the unique genius of the U.S. Constitution and largely credited for the success of American democracy. Yet the truth is closer to the opposite. The success of American democracy obviated the Madisonian design of separation of powers almost from the outset, preempting the political dynamics that were supposed to provide each branch with a “will of its own”

so that departmental “ambition [c]ould be made to counteract ambition.” What the Framers did not count on was the emergence of robust democratic competition, in government and in the electorate. Political competition and cooperation along relatively stable lines of policy and ideological disagreement quickly came to be channeled not through branches of government, but through an institution the Framers could imagine only dimly but nevertheless despised: political parties. As competition between the branches of government was displaced by competition between two major parties, the machine that was supposed to go of itself stopped running.

Few aspects of the founding generation’s political theory are now more clearly anachronistic than their vision of legislative-executive separation of powers. Nevertheless, few of the Framers’ ideas continue to be taken as literally or sanctified as deeply by courts and constitutional scholars as the passages about interbranch relations in Madison’s Federalist 51. Constitutional discourse to this day embraces Madison’s account of rivalrous, self-interested branches as an accurate depiction of political reality and a firm foundation for the constitutional law of separation of powers. In the Madisonian simulacrum of democratic politics that underwrites constitutional doctrine and theory, the branches of government continue to be personified as political actors with interests and wills of their own, entirely disconnected from the interests and wills of the officials who populate them or the citizens whom these officials represent. Acting on these interests, the branches are supposedly engaged in a perpetual competition to aggrandize their own power and encroach upon their rivals. The partisan political competition that structures real-world democracy and dominates political discourse is almost entirely missing from this picture.

As this article describes, the invisibility of political parties has left the constitutional discourse of separation of powers with no conceptual resources to understand or predict basic features of the American political system. It has also generated a set of judicial decisions and theoretical rationalizations that float entirely free of any functional justification grounded in the actual workings of separation of powers. Ignoring the reality of parties and fixating on the paper partitions between the branches, the law and theory of separation of powers is a perfect fit for the government the Framers designed. Unfortunately, it misses much of the government we actually have.

This article seeks to envision the law and theory of separation of powers by viewing it through the lens of party competition. What is revealed is that the degree and kind of competition between the legislative and executive branches will vary significantly, and may all but disappear, depending on whether the House, Senate, and presidency are divided or unified by political party. The practical distinction between party-divided and party-unified government rivals, and often dominates, the constitutional distinction between the branches in predicting and explaining interbranch political dynamics. Recognizing that these dynamics will shift from competitive when government is divided to cooperative when it is unified calls into question many of the foundational assumptions of separation of powers law and theory. But it also allows us, more constructively, to think about numerous aspects of legal doctrine, constitutional structure, comparative constitutionalism, and institutional design in a new and more realistic light.

FROM BRANCHES TO PARTIES

According to the political theory of the Framers, “the great problem to be solved” was to design governance institutions that would afford “practical security” against the excessive concentration of political power.

Madison’s vision of competitive branches balancing and checking one another has dominated constitutional thought about the separation of powers through the present. Yet it has never been entirely clear how the Madisonian machine was supposed to operate. From the modern perspective of consolidated democracy, it is hard to see how such a mechanism would arise. Even assuming that officeholders are driven by a “lust for self-aggrandizement,” the structure of democratic politics effectively channels those ambitions into a different set of activities than aggrandizing their departments or defending them against encroachments. Individual politicians gain and exercise power by winning competitive elections and effectuating political or ideological goals (their own or those of their constituents). Neither of these objectives correlates in any obvious way with the interests or power of branches of government. Madison’s will-based theory of separation of powers would seem to require government officials who care about the intrinsic interests of their departments more than they care about their personal interests or the interests of the citizens they represent. Democratic politics is unlikely to generate such officials.



“That’s just the trouble. I can’t lick them and they won’t let me join them.”

Indeed, this is just what happened: Madison’s design was eclipsed almost from the outset by the emergence of robust democratic political competition. Rather than tying their ambitions to the constitutional duties or power base of their departments, officials responded to the material incentives of democratic politics in ways that now seem natural and inevitable: by forming incipient organizations that took sides on contested policy and ideological issues and competing to marshal support for their agendas. These efforts led inexorably (though haltingly) to the organization of institutions that would facilitate alliances among groups of like-minded elected officials and politically mobilized citizens on a national scale: political parties.

The idea of political parties, representing institutionalized divisions of interest, was famously anathema to the Framers, as it had long been in Western political thought. Equating parties to other forms of nefarious “factions,” the Framers had attempted to design a “Constitution Against Parties.” But the futility of this effort quickly became apparent. By the end of the first Congress it had become clear that political competition organized around issues and programs had the potential to divide coalitions of officeholders and cut through the constitutional boundaries between the branches. The precursors of the modern political party had taken root, founded by the very Framers who had authored a Constitution against them. The rise of partisan politics worked a revolution in the American system of separation of powers, radically realigning the

incentives of politicians and officeholders. The emergence of a robust system of democratic politics tied the power and political fortunes of government officials to issues and elections. This, in turn, created a set of incentives that rendered these officials largely indifferent to the powers or interests of the branches per se. As a result, the political allies of a representative pursuing a policy agenda will often be other representatives pursuing the same agenda, not other representatives who happen to be affiliated with the same branch. A hawkish senator eager to assert American military power abroad may be happy to grant a like-minded president open-ended authority to fight a global war on terrorism—notwithstanding that this delegation “aggrandizes” the power of the president at the expense of Congress. The same senator may oppose delegating authority to a president inclined to stage humanitarian interventions—but for reasons of foreign policy, not of Federalist 51. Thus, in Madison’s terms, “the interests of the man” have become quite disconnected from the interests of “the place.” There is simply no way of grasping how the American system of government works in practice without taking account of how partisan political competition has reshaped the constitutional structure of government in ways the Framers would find unrecognizable.

Indeed, the American system functions dramatically differently depending on whether we have a government under unified or divided party control. Any understanding of the American system of

separation of powers should start from the recognition that both are possible; indeed, inevitable. Contrary to the foundational assumption of constitutional law and theory since Madison, the United States has not one system of separation of powers but (at least) two. When government is divided, party lines track branch lines, and we should expect to see party competition channeled through the branches. The resulting interbranch political competition will look, for better or worse, very much like the Madisonian dynamic of competitive, rivalrous branches—though, again, the underlying mechanism will be entirely different from the one Madison envisioned. On the other hand, when government is unified and the engine of party competition is removed from the internal structure of government, we should expect interbranch competition to dissipate. Smoothing over branch boundaries, intraparty cooperation (as a strategy of interparty competition) will override the central dynamic of the Madisonian model.

There are functional differences between these two systems of separation of powers, party-separated and party-unseparated, but the challenge to the constitutional law and theory of separation of powers should already be clear.

PARTY UNIFICATION AND DIVISION OF GOVERNMENT

The significance of whether government is unified or divided depends crucially on the internal ideological coherence and political distance between the two major parties. But the two major parties today are as coherent and polarized as they have been in perhaps a century, and for reasons that are likely to be enduring. The combination of a mature system of two-party competition nationwide, gerrymandered “safe” election districts, and somewhat more powerful party organizations in the last generation has led to the resurgence of more internally unified, ideologically coherent, and polarized parties than we have seen in many decades. And there is at least some reason to believe this will be the case for decades to come.

With cohesive and polarized parties, the functional differences between divided and unified government will be at their most pronounced. Partisan competition in government now means a Democratic Party dominated by liberals, with few moderates and no conservatives, pitted against a Republican Party dominated by conservatives, with few moderates and no liberals. Under divided governments, the absence of

a bloc of centrist legislators willing to cross party lines will make policy agreement more difficult and interbranch disagreement more intense. Under unified governments, smaller partisan majorities will be able to effect major policy change without the full range of checks of balances that are supposed to divide and diffuse power in the Madisonian system. These differences are immediately relevant to the goals and mechanisms of the constitutional separation of powers.

REENVISIONING, AND REFORMING, THE SEPARATION OF POWERS

We first explore numerous implications for constitutional doctrine from this more realistic understanding of separation of powers in practice, including for legal issues involving executive power, legislative power, and the relationship between them. But though constitutional lawyers and scholars are likely to focus first on doctrinal implications, a party-centered approach to separation of powers points in other directions as well. Particularly when it comes to confronting the unique

ernment’s policies are made highly visible. Other European democracies have gone further in organizing and empowering the minority party to play a more effective role in disrupting or influencing majoritarian governance. The French system, for example, entitles a minority in parliament to invoke judicial review to test the constitutionality of laws passed by parliament before they go into effect. This system of standing and constitutional review was created in 1974 precisely to place limits on the majority party in parliament.

The idea of minority opposition rights and institutions has not had much purchase in the American context, largely because American constitutional design assumes that political opposition, like political competition more generally, will be rooted in the branches. Congress is cast as the “opposition” to presidential governance, its “rights” are those associated with separation of powers, and “opposition rights” are thus subsumed into the ordinary workings of checks and balances. As we have seen, however, the structural position of the minority party under unified

There is simply no way of grasping how the American system of government works in practice without taking account of... partisan political competition....

challenge of strongly unified government, we might push beyond legal rules to think about separation of powers at the level of democratic institutional design.

MINORITY OPPOSITION RIGHTS

Viewing the absence of intragovernmental competition as a problematic feature of their parliamentary systems, many Western democracies have embraced the idea of “opposition rights” for minority parties out of power. In the Westminster system, for example, the majority party’s unified control over government is tempered by an elaborate set of rules and conventions that formally organize the minority party as “Her Majesty’s Official Opposition,” charged with organizing a shadow government to offer criticisms of and alternatives to the policies of the actual, majority government. The British opposition lacks any real agenda-setting, veto, or other co-governing power, but devices like opposition days and question times are designed to ensure that ongoing criticism of the gov-

ernment is more closely analogous to that of minority parties shut out of parliamentary governments than observers of the American system have recognized. Perhaps American constitutional theorists have something to learn from the comparative study of democratic oppositions in Europe and elsewhere.

Not surprisingly, constitutional democracies formed in the aftermath of totalitarian states since World War II have been the most acutely aware of the centrality of political party competition to robust democracy and, especially, the value of empowering democratic opposition. The German Constitution, probably the most imitated in recent decades, provides explicitly for the protection of parties, so that, as the Constitutional Court has held, German political parties have the “rank of constitutional institutions” and are “constitutionally integral units of a free and democratic system of government.” Opposition parties are empowered to participate in the Bundestag’s agenda-setting process, to chair a

number of standing committees, and to initiate constitutional review of legislation before the Constitutional Court whenever one-third of the members of parliament request. Combined with representation in the Bundesrat, this political leverage enables the opposition routinely to force compromises on legislation and to enact a significant part of its policy agenda.

American political institutions have never been self-consciously designed in this way to position minority parties as an “opposition” to unified government under majority control. Nevertheless, institutional structures that enable minority parties to engage in auditing, investigation, and information-gathering are one important means of maintaining checks and balances under unified party control of government.

Supermajority rules can also serve this purpose, by creating a *de facto* minority party veto and requiring bipartisan support for action to proceed. Consider, for example, the Senate filibuster in the context of judicial nominations, which might be viewed (sympathetically) as a *de facto* response to

might focus on insulating the administrative bureaucracy more fully from the partisan pressures of unified government. Here, one would need to take seriously the idea of the bureaucracy as a “fourth branch” of government, and to ensure that government is never fully unified by keeping this branch, at least, out of the hands of the majority party.

Indeed, one explanation for the rise of American bureaucracy is precisely that it came about as a response to the problem of the first overwhelmingly unified party government in American history and the threat such government was thought to pose to traditional American political practice. Although we often associate the rise of bureaucracy with more technocratic and instrumental imperatives—with the need for a specialized division of labor in the face of increasingly technical and complex regulatory problems—American bureaucracy was self-consciously constructed in the late 19th century, at least in part, for political reasons intimately connected to separation of powers ideals. The Civil War and Reconstruction established the domi-

highly skeptical of the bright-line distinctions between value judgments about ends and technocratic decision-making about means, and of the desirability of rule by the politically unaccountable “experts” that characterized Progressive- and New Deal-era administrative theory.

Were we to take more seriously the virtues of an independent bureaucracy, it is easy to imagine institutional modifications that would create greater political insulation. Longer or even life tenure for high-ranking administrators is an obvious example. Such reforms might demand greater flexibility than the Supreme Court’s formalistic Appointments Clause jurisprudence would presently afford. Other doctrinal reforms, such as rethinking separation of powers decisions that have facilitated greater political control over administrative agencies, prominently including Chevron, and resisting further efforts by presidents to exert greater control over agencies through the OMB and other strategies of presidential administration, would also help create the conditions for bureaucratic independence.

The faith of Progressives and New Dealers in neutral policy expertise may well be both naive and anachronistic. But the separation of powers case for politically insulated administration is neither. Politicization of the bureaucracy in the post-World War II era, whatever its benefits in terms of democratic accountability and political realism, has gradually eroded the capacity of bureaucratic institutions to check and balance unified party government. Perhaps constitutional and administrative lawyers and theorists should take a closer look at what has been lost.

POLITICAL PARTIES AND THE LAW OF DEMOCRACY

Thus far we have focused on mitigating the effects of unified parties and unified government, exploring institutions and reforms that might restore some measure of checks and balances under unified government. But we might also focus on measures to minimize the deleterious consequences of unified government by fragmenting or moderating the parties themselves, or by looking at ways to prevent strongly unified government from emerging in the first place. We cannot return to the Framers’ pre-modern vision of self-government without parties, but we might envision using legal rules and institutions to prevent parties from unifying government so strongly as to threaten Madisonian values. In the full article, we discuss various legal and policy changes that could lead to less polarized

Equating parties to other forms of nefarious “faction,” the Framers had attempted to design a “Constitution Against Parties.”

the erosion of a branch-oriented “Senate” check on presidential nominations during strongly unified government. The filibuster might be understood as the equivalent of a minority opposition right: a way of recreating the functional consequences of the constitutional requirement of presidential-senate consensus over judicial nominations in a political system dominated by partisan, as opposed to interbranch, competition, but only when the filibuster is applied during unified government. When party control is divided between the president and Senate, no longer is there a minority party to protect, and any nominee confirmable by a Senate majority will be effectively screened for ideological moderation. Adding a supermajority requirement when the president and Senate are divided might serve only to increase the costs of the confirmation process.

BUREAUCRACY AND THE CHECKING FUNCTION

Instead of empowering the opposition party to oversee or check the majority party under unified government (or in addition to doing so), constitutional engineering

nance of the national government over the states, leading to greater nationalization and centralization of political power than the country had ever experienced. At the same time, the postbellum ascendancy of the Republican Party vividly demonstrated the prospect of one-party dominance of the newly empowered national government. The resulting threat to checks and balances and a pluralist political culture drove even some who had been Jacksonian majoritarians in the antebellum period to see the necessity of new institutional forms. One such form was a professionalized bureaucracy.

In light of the separation of powers origins of the administrative state, it is ironic that American political culture over the last 40 years or so, on both right and left, has become far more skeptical than that of most European democracies about the possibility of a technical expertise that stands relatively independent of politics. The story of the increasing politicization of the administrative state since the New Deal—and the increasing acceptance of the inevitability, if not desirability, of this politicization—is a familiar one. Legal culture has become

political parties, while also noting various advantages for democracy that polarized political parties actually might have.

CONCLUSION

From nearly the start of the American republic, the separation of powers as the Framers understood it—and as contemporary constitutional law continues to understand it—had ceased to exist. The enduring institutional form of democratic political competition is political parties, not branches. Absorbing that insight is essential, not just to descriptive and historical analysis of the practice of democracy in America, but for normative thought about constitutional law and the design of democratic institutions today. If interbranch checks and balances remains a vital aspiration, the failure of the Framers' understanding of political competition raises the risk of a mismatch between constitutional structures and constitutional aims. Recognizing that failure and replacing it with an understanding of the actual mechanisms of political competition suggests new approaches to constitutional law and institutional design that would more effectively realize the aims of the separation of powers.

Such a project is all the more urgent as we come to terms with an emerging equilibrium of ideologically coherent and polarized political parties. Strong parties will accentuate the differences between unified and divided government, making constitutional law's conceptualization of a singular, static system of separation of powers all the more problematic. And when strong parties combine with extended periods of unified government, the challenge to the Madisonian picture of separation of powers, and to the values it is meant to protect, is stark. If the goal is a system of separation of powers that resembles the one Madison and subsequent generations of constitutional theorists imagined, it will have to be built not around branches but around the institutions through which political competition is in fact organized in modern democracies: political parties. □

RICHARD PILDES is the Sudler Family Professor of Constitutional Law and codirector of the Center on Law and Security. He is a specialist in legal issues concerning the design of democracy. He cowrote the casebook *The Law of Democracy: Legal Structure of the Political Process*, and was an editor of *The Future of the Voting Rights Act*. This excerpt is taken from an article of the same title, coauthored with Daryl Levinson, that was published in the *Harvard Law Review*.