2008

# Data Mining and the Security-Liberty Debate

Daniel J. Solove
*George Washington University Law School*, dsolove@law.gwu.edu

# Data Mining and the Security-Liberty Debate

*Daniel J. Solove*†

## INTRODUCTION

Ever since the horrific images of September 11 were seared into the public consciousness, the longstanding clash between security and liberty has been at the forefront of law and politics. Generally, law enforcement is primarily investigative, focusing on apprehending perpetrators of past crimes. When it comes to terrorism, law enforcement shifts to being more preventative, seeking to identify terrorists before they act. To this end, the government has become interested in data mining—a new technological tool to pinpoint the terrorists burrowed in among us. Data mining involves creating profiles by collecting and combining personal data, and analyzing it for particular patterns of behavior deemed to be suspicious.[1] In this way, data mining helps predict who might be likely to conduct a future terrorist attack.

In 2002, the Department of Defense, under the guidance of Admiral John Poindexter, began developing a data mining project called Total Information Awareness (TIA). Under the TIA program, the government would assemble a massive database consisting of financial, educational, health, and other information on US citizens, which would later be analyzed to single out people matching a terrorist profile. According to Poindexter, "[t]he only way to detect [ ] terrorists is to look for patterns of activity that are based on observations from past terrorist attacks as well as estimates about how terrorists will adapt to our measures to avoid detection."[2] The program sparked public outrage, and the Senate denied it funding. But TIA did not die; instead, it exists in various projects with obscure names such as Basketball, Genoa II, and Topsail. Unlike TIA, which had its own website, these projects are significantly more clandestine.[3]

---

† Associate Professor, George Washington University Law School. Thanks to Chris Hoofnagle, Paul Schwartz, Michael Sullivan, and Tal Zarsky for helpful suggestions, and to Sheerin Shahinpoor for research assistance. 

1 See Technology and Privacy Advisory Committee, Report, *Safeguarding Privacy in the Fight against Terrorism* ("TAPAC Report") 2–5 (Mar 1, 2004), online at http://www.cdt.org/security/usapatriot/20040300tapac.pdf (visited Jan 12, 2008) (discussing the Department of Defense's and other government agencies' use of data mining as a means of intelligence gathering).

2 John M. Poindexter, *Finding the Face of Terror in Data*, NY Times A25 (Sept 10, 2003).

3 Shane Harris, *TIA Lives On*, Natl J 66, 66–67 (Feb 25, 2006).

The government has also been developing other data mining programs. The Technology and Privacy Advisory Committee (TAPAC), tasked with examining government data mining, noted that "TIA was not the tip of the iceberg, but rather one small specimen in a sea of icebergs."[4] Following September 11, the Transportation Security Administration (TSA), with the help of the FBI, has been developing a program to mine data about airline passengers to determine who should be allowed to fly, selected for extra screening, or denied the right to board an aircraft.[5] In May 2006, the media revealed that the National Security Administration (NSA) had gathered a massive amount of telephone customer records to create the "largest database ever assembled in the world."[6] Various states have used the Multistate Anti-terrorism Information Exchange (MATRIX), a shared database of personal information subject to data mining analysis.[7] Countless other data mining programs are being used or developed—about 200 according to one government report in 2004.[8]

Data mining is one issue in a larger debate about security and privacy. Proponents of data mining justify it as an essential tool to protect our security. For example, Judge Richard Posner argues that "[i]n an era of global terrorism and proliferation of weapons of mass destruction, the government has a compelling need to gather, pool, sift, and search vast quantities of information, much of it personal."[9] Moreover, proponents of security measures argue that we must provide the executive branch with the discretion it needs to protect us. We cannot second guess every decision made by government officials, and excessive meddling into issues of national security by judges and oth-

4 TAPAC Report at 5 (cited in note 1).

5 See Daniel J. Solove, Marc Rotenberg, and Paul M. Schwartz, *Information Privacy Law* 603–04 (Aspen 2d ed 2006). Several successive versions of the screening program have been developed and abandoned. For example, after September 11, the government sought to develop the Computer Assisted Passenger Prescreening System II (CAPPS II)—a successor to the screening program in place before September 11. CAPPS II was abandoned because it grew too far beyond its original purpose; it was later replaced with a program called Secure Flight. This was later abandoned as well. The TSA continues to develop passenger screening data mining systems.

6 Leslie Cauley, *NSA Has Massive Database of Americans' Phone Calls; 3 Telecoms Help Government Collect Billions of Domestic Records*, USA Today 1A (May 11, 2006) (quotation marks omitted). But see Susan Page, *Lawmakers: NSA Database Incomplete; Some Who Were Briefed about the Database Identify Who Participated and Who Didn't*, USA Today 2A (June 30, 2006) (explaining that the database of domestic phone call records is incomplete because certain telecommunications providers did not provide the NSA with call records).

7 For more details on the MATRIX program, see Jacqueline Klosek, *The War on Privacy* 51–53 (Praeger 2007); GAO, *Data Mining: Federal Efforts Cover a Wide Range of Uses*, GAO-04-548, 5 (May 2004), online at http://www.gao.gov/new.items/d04548.pdf (visited Jan 12, 2008).

8 GAO, *Data Mining* at 2 (cited in note 7).

9 Richard A. Posner, *Not a Suicide Pact: The Constitution in a Time of National Emergency* 141 (Oxford 2006).

ers lacking expertise will prove detrimental. For example, William Stuntz contends that "effective, active government—government that innovates, that protects people who need protecting, that acts aggressively when action is needed—is dying. Privacy and transparency are the diseases. We need to find a vaccine, and soon."[10] Stuntz concludes that "[i]n an age of terrorism, privacy rules are not simply unaffordable. They are perverse."[11]

We live in an "age of balancing," and the prevailing view is that most rights and civil liberties are not absolute.[12] Thus, liberty must be balanced against security. But there are systematic problems with how the balancing occurs that inflate the importance of the security interests and diminish the value of the liberty interests. In this essay, I examine some common difficulties in the way that liberty is balanced against security in the context of data mining. Countless discussions about the tradeoffs between security and liberty begin by taking a security proposal and then weighing it against what it would cost our civil liberties. Often, the liberty interests are cast as individual rights and balanced against the security interests, which are cast in terms of the safety of society as a whole. Courts and commentators defer to the government's assertions about the effectiveness of the security interest. In the context of data mining, the liberty interest is limited by narrow understandings of privacy that neglect to account for many privacy problems. As a result, the balancing concludes with a victory in favor of the security interest. But as I will argue, important dimensions of data mining's security benefits require more scrutiny, and the privacy concerns are significantly greater than currently acknowledged. These problems have undermined the balancing process and skewed the results toward the security side of the scale.

## I. THE SECURITY INTEREST

Debates about data mining begin with the assumption that it is an essential tool in protecting our security. Terrorists lurk among us, and ferreting them out can be quite difficult. Examining data for patterns will greatly assist in this endeavor, the argument goes, because certain identifiable characteristics and behaviors are likely to be associated with terrorist activity. Often, little more is said, and the debate pro-

10 William J. Stuntz, *Secret Service: Against Privacy and Transparency*, New Republic 12, 12 (Apr 17, 2006).

11 Id at 14.

12 See T. Alexander Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 Yale L J 943, 965–72 (1987) (analyzing the evolution and acceptance of the balancing of interests in a wide array of different constitutional issues and discussing some of the problems that have since arisen).

ceeds to examine whether privacy is important enough to refrain from using such an effective terrorism-fighting tool.

Many discussions about security and liberty proceed in this fashion. They commence by assuming that a particular security measure is effective, and the only remaining question is whether the liberty interest is strong enough to curtail that measure. But given the gravity of the security concerns over terrorism, the liberty interest has all but lost before it is even placed on the scale.

### A. The Deference Argument

Judge Richard Posner argues that judges should give the executive branch considerable deference when it comes to assessing the security measures it proposes. In his recent book, *Not a Suicide Pact: The Constitution in a Time of National Emergency*,[13] Posner contends that judicial restraint is wise because "when in doubt about the actual or likely consequences of a measure, the pragmatic, empiricist judge will be inclined to give the other branches of government their head."[14] According to Posner, "[j]udges aren't *supposed* to know much about national security."[15] Likewise, Eric Posner and Adrian Vermeule declare in their new book, *Terror in the Balance: Security, Liberty, and the Courts*,[16] that "the executive branch, not Congress or the judicial branch, should make the tradeoff between security and liberty."[17] Moreover, Posner and Vermeule declare that during emergencies, "[c]onstitutional rights should be relaxed so that the executive can move forcefully against the threat."[18]

The problem with such deference is that, historically, the executive branch has not always made the wisest national security decisions. Nonetheless, Posner and Vermeule contend that notwithstanding its mistakes, the executive branch is better than the judicial and legislative branches on institutional competence grounds.[19] "Judges are generalists," they observe, "and the political insulation that protects them from current politics also deprives them of information, especially information about novel security threats and necessary responses to those threats."[20] Posner and Vermeule argue that during emergencies, the

13 Richard A. Posner, *Not a Suicide Pact: The Constitution in a Time of National Emergency* (Oxford 2006).

14 Id at 27.

15 Id at 37.

16 Eric A. Posner and Adrian Vermeule, *Terror in the Balance: Security, Liberty, and the Courts* (Oxford 2007).

17 Id at 5.

18 Id at 16.

19 See id at 6.

20 Id at 31.

"novelty of the threats and of the necessary responses makes judicial routines and evolved legal rules seem inapposite, even obstructive."[21]

"Judicial routines" and "legal rules," however, are the cornerstone of due process and the rule of law—the central building blocks of a free and democratic society. At many times, Posner, Vermeule, and other strong proponents of security seem to focus almost exclusively on what would be best for security when the objective should be establishing an optimal balance between security and liberty. Although such a balance may not promote security with maximum efficiency, it is one of the costs of living in a constitutional democracy as opposed to an authoritarian political regime. The executive branch may be the appropriate branch for developing security measures, but this does not mean that it is the most adept branch at establishing a balance between security and liberty.

In our constitutional democracy, all branches have a role to play in making policy. Courts protect constitutional rights not as absolute restrictions on executive and legislative policymaking but as important interests to be balanced against government interests. As T. Alexander Aleinikoff notes, "balancing now dominates major areas of constitutional law."[22] Balancing occurs through various forms of judicial scrutiny, requiring courts to analyze the weight of the government's interest, a particular measure's effectiveness in protecting that interest, and the extent to which the government interest can be achieved without unduly infringing upon constitutional rights.[23] For balancing to be meaningful, courts must scrutinize both the security and liberty interests.

With deference, however, courts fail to give adequate scrutiny to security interests. For example, after the subway bombings in London, the New York Police Department began a program of random searches of people's baggage on the subway. The searches were conducted without a warrant, probable cause, or even reasonable suspicion. In *MacWade v Kelly*,[24] the United States Court of Appeals for the Second Circuit upheld the program against a Fourth Amendment challenge. Under the special needs doctrine, when exceptional circumstances make the warrant and probable cause requirements unnecessary, the search is analyzed in terms of whether it is "reasonable."[25]

---

21 Id at 18.

22 Aleinikoff, 96 Yale L J at 965 (cited in note 12).

23 See Daniel J. Solove, *The Darkest Domain: Deference, Judicial Review, and the Bill of Rights*, 84 Iowa L Rev 941, 954–55 (1999).

24 460 F3d 260 (2d Cir 2006).

25 Id at 267–68.

Reasonableness is determined by balancing the government interest in security against the interests in privacy and civil liberties.[26]

The weight of the security interest should turn on the extent to which the program effectively improves subway safety. The goals of the program may be quite laudable, but nobody questions the importance of subway safety. The critical issue is whether the search program is a sufficiently effective way of achieving those goals that it is worth the tradeoff in civil liberties. On this question, unfortunately, the court deferred to the law enforcement officials, stating that the issue "is best left to those with a unique understanding of, and responsibility for, limited public resources, including a finite number of police officers."[27] In determining whether the program was "a reasonably effective means of addressing the government interest in deterring and detecting a terrorist attack on the subway system,"[28] the court refused to examine the data to assess the program's effectiveness.[29]

The way the court analyzed the government's side of the balance would justify nearly any search, no matter how ineffective. Although courts should not take a know-it-all attitude, they should not defer on such a critical question as a security measure's effectiveness. The problem with many security measures is that they are not wise expenditures of resources. A small number of random searches in a subway system of over four million riders a day seems more symbolic than effective because the odds of the police finding the terrorist with a bomb are very low. The government also argued that the program would deter terrorists from bringing bombs on subway trains, but nearly any kind of security measure can arguably produce some degree of deterrence. The key issue, which the court did not analyze, is whether the program would lead to deterrence significant enough to outweigh the curtailment of civil liberties.

If courts fail to question the efficacy of security measures, then the security interest will prevail nearly all the time. Preventing terrorism has an immensely heavy weight, and any given security measure will provide a marginal advancement toward that goal. In the defer-

26 Id at 269.

27 Id at 273 (quotation marks omitted).

28 Id (quotation marks omitted).

29 The court declared:

> We will not peruse, parse, or extrapolate four months' worth of data in an attempt to divine how many checkpoints the City ought to deploy in the exercise of its day-to-day police power. Counter-terrorism experts and politically accountable officials have undertaken the delicate and esoteric task of deciding how best to marshal their available resources in light of the conditions prevailing on any given day. We will not—and *may not*—second-guess the minutiae of their considered decisions.

Id at 274.

ence equation, the math then becomes easy. At this point, it is futile to even bother to look at the civil liberties side of the balance. The government side has already won.

Proponents of deference argue that if courts did not defer, then they would be substituting their judgment for that of executive officials, who have greater expertise in understanding security issues. Special expertise in national security, however, is often not necessary for balancing security and liberty. Judges and legislators should require the experts to persuasively justify the security measures being developed or used. Of course, in very complex areas of knowledge, such as advanced physics, nonexperts may find it difficult to understand the concepts and comprehend the terminology. But it is not clear that security expertise involves such sophisticated knowledge that it would be incomprehensible to nonexperts. Moreover, the deference argument conflates *evaluating* a particular security measure with *creating* such a measure. The point of judicial review is to subject the judgment of government officials to critical scrutiny rather than blindly accept their authority. Critical inquiry into factual matters is not the imposition of the judge's own judgment for that of the decisionmaker under review.[30] Instead, it is forcing government officials to explain and justify their policies.

Few will quarrel with the principle that courts should not "second guess" the decisions of policy experts. But there is a difference between not "second guessing" and failing to critically evaluate the factual and empirical evidence justifying the government programs. Nobody will contest the fact that security is a compelling interest. The key issue in the balancing is the extent to which the security measure furthers the interest in security. As I have argued elsewhere, whenever courts defer to the government on the effectiveness of a government security measure, they are actually deferring to the government on the ultimate question as to whether the measure passes constitutional muster.[31] Deference by the courts or legislature is an abdication of their function. Our constitutional system of government was created with three branches, a design structured to establish checks and balances against abuses of power. Institutional competence arguments are often made as if they are ineluctable truths about the nature of each governmental branch. But the branches have all evolved considerably throughout history. To the extent a branch lacks resources to carry out its function, the answer should not be to diminish the power of that branch but to provide it with the necessary tools so it can more

30 See Solove, 84 Iowa L Rev at 1019 (cited in note 23).

31 See id at 958–59, 967–68.

effectively carry out its function. Far too often, unfortunately, discussions of institutional competence devolve into broad generalizations about each branch and unsubstantiated assertions about the inherent superiority of certain branches for making particular determinations.

It is true, as Posner and Vermeule observe, that historically courts have been deferential to the executive during emergencies.[32] Proponents of security measures often advance what I will refer to as the "pendulum theory"—that in times of crisis, the balance shifts more toward security and in times of peace, the balance shifts back toward liberty. For example, Chief Justice Rehnquist argues that the "laws will thus not be silent in time of war, but they will speak with a somewhat different voice."[33] Judge Posner contends that the liberties curtailed during times of crisis are often restored during times of peace.[34] Deference is inevitable, and we should accept it without being overly concerned, for the pendulum will surely swing back.

As I argue elsewhere, however, there have been many instances throughout US history of needless curtailments of liberty in the name of security, such as the Palmer Raids, the Japanese Internment, and the McCarthy communist hearings.[35] Too often, such curtailments did not stem from any real security need but because of the "personal agendas and prejudices" of government officials.[36] We should not simply accept these mistakes as inevitable; we should seek to prevent them from occurring. Hoping that the pendulum will swing back offers little consolation to those whose liberties were infringed or chilled. The protection of liberty is most important in times of crisis, when it is under the greatest threat. During times of peace, when our judgment is not clouded by fear, we are less likely to make unnecessary sacrifices of liberty. The threat to liberty is lower in peacetime, and the need to protect it is not as dire. The greatest need for safeguarding liberty is during times when we least want to protect it.

### B. Assessing the Security Threat

In order to balance security and liberty, we must assess the security interest. This involves evaluating two components—the gravity of the security threat and the effectiveness of the security measures to address it. It is often merely assumed without question that the secu-

---

32 See Posner and Vermeule, *Terror in the Balance* at 32 (cited in note 16).

33 William H. Rehnquist, *All the Laws But One: Civil Liberties in Wartime* 225 (Knopf 1998).

34 See Richard A. Posner, *Law, Pragmatism, and Democracy* 304 (Harvard 2003).

35 See Daniel J. Solove, *Melville's* Billy Budd *and Security in Times of Crisis*, 26 Cardozo L Rev 2443, 2457–59 (2005); Michael Sullivan and Daniel J. Solove, *Can Pragmatism Be Radical? Richard Posner and Legal Pragmatism*, 113 Yale L J 687, 711–13 (2003).

36 Sullivan and Solove, 113 Yale L J at 712 (cited in note 35).

rity threat from terrorism is one of the gravest dangers we face in the modern world. But this assumption might be wrong.

Assessing the risk of harm from terrorism is very difficult because terrorism is such an irregular occurrence and is constantly evolving. If we examine the data from previous terrorist attacks, however, the threat of terrorism has been severely overstated. For example, many people fear being killed in a terrorist attack, but based on statistics from terrorism in the United States, the risk of dying from terrorism is miniscule. According to political scientist John Mueller,

> [e]ven with the September 11 attacks included in the count . . . the number of Americans killed by international terrorism since the late 1960s (which is when the State Department began its accounting) is about the same as the number killed over the same period by lightning, or by accident-causing deer, or by severe allergic reactions to peanuts.[37]

Add up the eight deadliest terrorist attacks in US history, and they amount to fewer than four thousand fatalities.[38] In contrast, flu and pneumonia deaths are estimated to be around sixty thousand per year.[39] Another forty thousand die in auto accidents each year.[40] Based on our experience with terrorism thus far, the risk of dying from terrorism is very low on the relative scale of fatal risks.

Dramatic events and media attention can cloud a rational assessment of risk. The year 2001 was not just notable for the September 11 attacks. It was also the summer of the shark bite, when extensive media coverage about shark bites led to the perception that such attacks were on the rise. But there were fewer shark attacks in 2001 than in 2000 and fewer deaths as well, with only four in 2001 as compared to thirteen in 2000.[41] And regardless of which year had more deaths, the number is so low that an attack is a freak occurrence.

It is certainly true that our past experience with terrorism might not be a good indicator of the future. More treacherous terrorism is possible, such as the use of nuclear or biological weapons. This complicates our ability to assess the risk of harm from terrorism. Moreover,

---

37 John Mueller, *Overblown: How Politicians and the Terrorism Industry Inflate National Security Threats, and Why We Believe Them* 13 (Free Press 2006).

38 See Bruce Schneier, *Beyond Fear: Thinking Sensibly about Security in an Uncertain World* 239 (Copernicus 2003).

39 See, for example, Arialdi M. Miniño, Melonie P. Heron, and Betty L. Smith, *Deaths: Preliminary Data for 2004*, 54 Natl Vital Stats Rep 19, 28 table 7 (2006), online at http://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_19.pdf (visited Jan 12, 2008).

40 See id.

41 See Jeordan Legon, *Survey: "Shark Summer" Bred Fear, Not Facts* CNN.com (Mar 14, 2003), online at http://www.cnn.com/2003/TECH/science/03/13/shark.study/ (visited Jan 12, 2008).

the intentional human conduct involved in terrorism creates a sense of outrage and fear that ordinary deaths do not engender. Alleviating fear must be taken into account, even if such fear is irrationally high in relation to other riskier events such as dying in a car crash. But enlightened policy must not completely give in to the panic and irrational fear of the moment. It should certainly attempt to quell the fear, but it must do so thoughtfully.

Nevertheless, most policymakers find it quite difficult to assess the threat of terrorism modestly. In the face of widespread public panic, it is hard for government officials to make only moderate changes. Something dramatic must be done, or political heads will roll. Given the difficulty in assessing the security threat in a more rational manner, it is imperative that the courts meaningfully analyze the effectiveness of security measures. Even if panic and fear might lead to the gravity of the threat being overstated, we should at least ensure that the measures taken to promote security are sufficiently effective to justify the cost. Unfortunately, as I will discuss in the next section, rarely do discussions about the sacrifice of civil liberties explain the corresponding security benefit, why such a benefit cannot be achieved in other ways, and why such a security measure is the best and most rational one to take.

### C. Assessing the Security Measures

Little scrutiny is given to security measures. They are often just accepted as a given, no matter how ill-conceived or ineffective they might be. Some ineffective security measures are largely symbolic, such as the New York City subway search program. The searches are unlikely to catch or deter terrorists because they involve only a miniscule fraction of the millions of daily passengers. Terrorists can just turn to other targets or simply attempt the bombing on another day or at another train station where searches are not taking place. The vice of symbolic security programs is that they result in needless sacrifices of liberty and drain resources from other, more effective security measures. Nevertheless, these programs have a virtue—they can ameliorate fear because they are highly visible. Ironically, the subway search program's primary benefit was alleviating people's fear (which was probably too high), albeit in a deceptive manner (as the program did not add much in the way of security).

Data mining represents another kind of security measure, one that currently has little proven effectiveness and little symbolic value. Data mining programs are often not visible enough to the public to quell much fear. Instead, their benefits come primarily from their actual effectiveness in reducing terrorist threats, which remains highly speculative. Thus far, data mining is not very accurate in the behav-

ioral predictions it makes. For example, there are approximately 1.8 million airline passengers each day.[42] A data mining program to identify terrorists with a false positive rate of 1 percent (which would be exceedingly low for such a program) would flag eighteen thousand people as false positives. This is quite a large number of innocent people.

Why is the government so interested in data mining if it remains unclear whether it will ever be very accurate or workable? Part of the government's interest in data mining stems from the aggressive marketing efforts of database companies. After September 11, database companies met with government officials and made a persuasive pitch about the virtues of data mining.[43] The technology sounds quite dazzling when presented by skillful marketers, and it can work quite well in the commercial setting. The problem, however, is that just because data mining might be effective for businesses trying to predict customer behavior does not make it effective for the government trying to predict who will engage in terrorism. A high level of accuracy is not necessary when data mining is used by businesses to target marketing to consumers, because the cost of error to individuals is minimal. Amazon.com, for example, engages in data mining to determine which books its customers are likely to find of interest by comparing book-buying patterns among its customers. Although it is far from precise, it need not be because there are few bad consequences if it makes a wrong book recommendation. Conversely, the consequences are vastly greater for government data mining.

Ultimately, I do not believe that the case has been made that data mining is a wise expenditure of security resources. Those who advocate for security should be just as outraged as those on the liberty side of the debate. Although courts should not micromanage which security measures the government chooses, they should examine the effectiveness of any given security measure to weigh it against the liberty costs. Courts should not tell the executive branch to modify a security measure just because they are not convinced it is the best one, but they should tell the executive that a particular security measure is not effective enough to outweigh the liberty costs. The very point of protecting liberty is to demand that sacrifices to liberty are not in vain and that security interests, which compromise civil liberties, are sufficiently effective to warrant the cost.

---

42 See Fred Bayles, *Air-Traveler Screening, Privacy Concerns Collide*, USA Today 6A (Oct 3, 2003).

43 See, for example, Robert O'Harrow, Jr., *No Place to Hide* 56–63 (Free Press 2005) (discussing the lobbying efforts of Acxiom, a data brokerage company, to convince politicians of data mining's potential value for counterterrorism purposes).

### D. The Zero-sum Tradeoff Argument

Those defending the national security side of the balance often view security and liberty as a zero-sum tradeoff. Posner and Vermeule contend, for example, that "[a]t the security-liberty frontier, any increase in security requires a decrease in liberty."[44] It is not clear, however, why security and civil liberties must be mutually exclusive. Not all security measures compromise liberty. Moreover, there is no established correlation between the effectiveness of a security measure and a corresponding decrease in liberty. In other words, the most effective security measures need not be the most detrimental to liberty.

Proponents of security characterize rights as bans and restrictions on surveillance. For example, in justifying the NSA surveillance program, Attorney General Alberto Gonzales stated,

> I cannot help but wonder if [terrorists] are not shaking their heads in amazement at the thought that anyone would imperil such a sensitive program by leaking its existence in the first place, and smiling at the prospect that we might now disclose even more or perhaps even unilaterally disarm ourselves of a key tool in the war on terror.[45]

The balance between security and liberty is often discussed in terms of whether a government surveillance program should or should not occur. In other words, the tradeoff between security and liberty is viewed as all or nothing.

But this is not how most constitutional and statutory protections against surveillance currently work. The Fourth Amendment rarely bans surveillance; it requires judicial oversight of such surveillance and it mandates that the government justify its measures. Under the Fourth Amendment, the government can engage in many searches and seizures that are very invasive of privacy so long as the government can justify its privacy invasions to a neutral judge or magistrate, who will then grant a search warrant. Thus, the cost to security of protecting liberty need not be the scrapping of an entire security measure, but rather, imposing particular kinds of oversight, accountability, and minimization. When security measures are balanced against liberty, what should be weighed is the marginal increase in effectiveness of the security measure from the lack of judicial oversight, minimization, or other procedural protection ordinarily imposed to safeguard liberty.

---

44 Posner and Vermeule, *Terror in the Balance* at 12 (cited in note 16).

45 Wartime Executive Power and the National Security Agency's Surveillance Authority, Hearing before the Senate Committee on the Judiciary, 109th Cong, 2d Sess 10, 15 (2006) (testimony of Alberto Gonzales, Attorney General).

Far too often, however, discussions of security and liberty neglect to assess the balance this way. Polls frequently pose the question as an all-or-nothing tradeoff. A 2002 Pew Research poll asked American citizens: "Should the government be allowed to read e-mails and listen to phone calls to fight terrorism?"[46] A 2005 poll from Rasmussen Reports posed the question: "Should the National Security Agency be allowed to intercept telephone conversations between terrorism suspects in other countries and people living in the United States?"[47] Both these questions, however, neglect to account for warrants and court orders. Few would contend that the government should not be allowed to conduct a wide range of searches when it has a search warrant or court order. So the questions should be posed as: Should the government be allowed to read emails and listen to phone calls *without a search warrant or the appropriate court order required by law* to fight terrorism? Should the National Security Agency be allowed to intercept telephone conversations between terrorism suspects in other countries and people living in the United States *without a court order or judicial oversight*? The choice is not between a security measure and nothing, but between a security measure with oversight and control and a security measure within the sole discretion of executive officials.

## II. THE LIBERTY INTEREST

When it comes to the liberty interest in the security-versus-liberty balance, a host of problems arise in the balancing. One of the primary liberty interests implicated by data mining is privacy, but several proponents of data mining contend that the privacy interest is not particularly strong. Even when the problems of data mining are understood in their full dimensions, and although they implicate many constitutional rights, data mining often falls between the crevices of constitutional doctrine.

The privacy problems with data mining are often defined in narrow ways that neglect to account for the full panoply of problems created by the practice. As Richard Posner argues:

> The collection, mainly through electronic means, of vast amounts of personal data is said to invade privacy. But machine collection and processing of data cannot, as such, invade privacy. Because of their volume, the data are first sifted by computers, which search for names, addresses, phone numbers, etc., that may have intelli-

46 Bob Sullivan, *Have You Been Wiretapped?* MSNBC.com (Jan 10, 2006), online at http://redtape.msnbc.com/2006/01/have_you_been_w.html (visited Jan 12, 2008) (quotation marks omitted).

47 *National Security Agency*, Rasmussen Reports (Dec 28, 2005), online at http://www.rasmussenreports.com/2005/NSA.htm (visited Jan 12, 2008).

> gence value. This initial sifting, far from invading privacy (a computer is not a sentient being), keeps most private data from being read by any intelligence officer.[48]

The potential harm from data mining, Posner argues, is use of the information to blackmail an "administration's critics and political opponents" or to "ridicule or embarrass."[49] Similarly, William Stuntz contends that the privacy problems are created not by collection but by disclosure and use: "The true image of privacy intrusion is not some NSA bureaucrat listening in on phone calls, but rather Kenneth Starr's leaky grand jury investigation, which splashed a young woman's social life across America's newspapers and TV screens."[50]

Posner and Stuntz focus on the problems of disclosure or the threat of disclosure (blackmail). Indeed, data mining is often understood as a problem of information dissemination. Elsewhere, however, I contend that privacy violations involve a range of different kinds of problems.[51] Of the four broad categories of problems I identify, at least three are implicated in data mining programs: information collection, information processing, and information dissemination. Proponents of data mining, therefore, must address all of these types of problems, not just problems with information dissemination.

### A. Problems with the Collection of Information

It is certainly true that disclosure and use of information can create significant privacy problems. But collection can create problems as well. Although the leaks from Kenneth Starr's investigation were a problem, many people were more shocked by the powerful tools used by the prosecutor—for example, Starr calling Monica Lewinsky's mother to testify against her and issuing subpoenas to a bookstore for Lewinsky's book purchases. Such information gathering about First Amendment activities involving people's reading habits and speech might chill the exercise of such rights.[52]

Despite potential problems with the information collection for use in data mining, the Fourth Amendment provides little protection because of doctrinal limitations. One enormous problem is the third-party doctrine. In *United States v Miller*,[53] the Supreme Court held that

---

48 Richard A. Posner, *Our Domestic Intelligence Crisis*, Wash Post A31 (Dec 21, 2005).

49 Posner, *Not a Suicide Pact* at 97 (cited in note 13).

50 Stuntz, *Secret Service*, New Republic at 15 (cited in note 10).

51 See Daniel J. Solove, *A Taxonomy of Privacy*, 154 U Pa L Rev 477, 484–89 (2006).

52 See Daniel J. Solove, *The First Amendment as Criminal Procedure*, 82 NYU L Rev 112, 143–51 (2007) (discussing Supreme Court decisions on chilling effects and the First Amendment, and their implications for government information gathering).

53 425 US 435 (1976).

people lack a reasonable expectation of privacy in their bank records because "[a]ll of the documents obtained, including financial statements and deposit slips, contain only information voluntarily conveyed to the banks and exposed to their employees in the ordinary course of business."[54] Employing analogous reasoning in *Smith v Maryland*,[55] the Supreme Court held that people lack a reasonable expectation of privacy in pen register information (the phone numbers they dial) because users "know that they must convey numerical information to the phone company," and therefore they cannot "harbor any general expectation that the numbers they dial will remain secret."[56]

Data mining is often a collaborative project between the government and businesses. In the Information Age, countless companies maintain detailed records of individuals' personal information: internet service providers, merchants, bookstores, phone companies, cable companies, and many more. The personal data is often gathered by businesses, which then supply it to the government. As so much of our personal information is in the hands of various companies, the third-party doctrine severely limits Fourth Amendment protection.

### B. Problems with the Processing of Information

Some of the most significant problems of data mining involve information processing—the way that previously gathered information is stored, analyzed, and used. The analysis of data to identify people who match certain profiles resembles a dragnet search—casting a giant net to see what it brings in. In many ways, this practice resembles general warrants—broad fishing expeditions for criminal activity—about which the Framers of the Constitution were particularly concerned in drafting the Fourth Amendment. This is why the Fourth Amendment imposes a requirement of particularity; the government must have particularized suspicion that what it is searching will turn up criminal evidence. Although data mining is a form of digital dragnet search, the Fourth Amendment does not regulate it because the searching occurs after the government has obtained the data. Indeed, the Fourth Amendment provides few, if any, limitations on the use, storage, or retention of data after collection.

Another potential threat posed by data mining is that it can target people based on their First Amendment activities. Suspicious profiles might involve information about people's free speech, free asso-

54 Id at 442.
55 442 US 735 (1979).
56 Id at 743.

ciation, or religious activity. Singling people out for extra investigation, for denial of the right to travel by plane, or for inclusion in a suspicious persons blacklist is more troubling if based even in part on protected First Amendment activities. As I argue elsewhere, I believe that there is an argument under existing First Amendment doctrine to challenge such practices, although such challenges have not frequently been made up to this point.[57]

Data mining might also implicate equal protection. A person's race or ethnicity might be used in a profile. To the extent that race or ethnicity is a major factor in singling out people as suspicious and deserving differential treatment, this might implicate the Equal Protection Clause. Some argue, however, that data mining helps to eliminate stereotyping and discrimination. As Tal Zarsky contends, data mining can minimize the human element, thus preventing bias and racism from entering into the process.[58] Whereas some data mining techniques involve a human-created profile of a terrorist and seek to identify people who match the profile, other data mining techniques involve the computer composing the profile by analyzing patterns of behavior from known terrorists. Even this latter technique, however, involves human judgment—about who qualifies as a terrorist and who does not. Profiles can contain pernicious assumptions—hidden in the architecture of computer code and embedded in algorithms so that they appear to be the decision of neutral computers.

On the other hand, one might argue, profiling via data mining might be better than the alternatives. Frederick Schauer aptly notes that there is no escape from profiling, for without data mining, officials will be making their own subjective judgments about who is suspicious.[59] These judgments are based on an implicit profile, though one that is not overt and articulated. "[T]he issue is not about whether to use profiles or not but instead about whether to use (or to prefer) formal written profiles or informal unwritten ones."[60] Although it is true that formal profiles constructed in advance have their virtues over ad hoc profiling by officials, formal profiles contain some disadvantages. They are more systematic than the ad hoc approach, thus compounding the effects of information tied to race, ethnicity, religion, speech, or other factors that might be problematic. Those profiling

---

57 See Solove, 82 NYU L Rev at 143–44 (cited in note 52).

58 See Tal Z. Zarsky, *"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 Yale J L & Tech 1, 27 (2003).

59 See Frederick Schauer, *Profiles, Probabilities, and Stereotypes* 173–74 (Belknap 2003) (discussing the tradeoff between profiles constructed in advance and those made on a case-by-case basis).

60 Id at 173.

informally are subject to scrutiny, as they have to answer in court about why they believed a person was suspicious. Data mining, however, lacks such transparency, a problem I will discuss later on. Formal written profiles cease to have an advantage over informal unwritten ones if they remain hidden and unsupervised.

Data mining also raises due process issues. As Daniel Steinbock notes, "[t]he most striking aspect of virtually all antiterrorist data matching and data mining decisions is the total absence of even the most rudimentary procedures for notice, hearing, or other opportunities for meaningful participation before, or even after, the deprivation [of liberty] is imposed."[61] Will those singled out by data mining programs be able to raise a challenge? Will people have a right to a hearing? How long will it take for people to get hearings? Will people have a right to an attorney? Will people get to correct erroneous data? How?

Suppose a person disagrees with the profile. Can this be addressed at the hearing? Likely not, as the profiles are secret. If the profiles were revealed to the public, the argument goes, the terrorists would be better able to take steps to evade them. But what kind of meaningful challenge can people make if they are not told about the profile that they supposedly matched? How can we evaluate the profiling systems if we are kept in the dark?

Predictive determinations about one's future behavior are much more difficult to contest than investigative determinations about one's past behavior. Wrongful investigative determinations can be addressed in adjudication. But wrongful predictions about whether a person might engage in terrorism at some point in the future are often not ripe for litigation and review. Nevertheless, people may experience negative consequences from such predictive judgments, such as being denied the ability to travel or being subject to extra scrutiny.

## C. Transparency

Another key issue regarding the liberty side of the balance is transparency—the degree of openness by which a particular security measure is carried out. Transparency is essential to promote accountability and to provide the public with a way to ensure that government officials are not engaging in abuse. "Sunlight is said to be the best of disinfectants," Justice Brandeis declared, "electric light the most efficient policeman."[62] As James Madison stated: "A popular Government,

---

61 Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 Ga L Rev 1, 82 (2005).

62 Louis D. Brandeis, *Other People's Money: And How the Bankers Use It* 62 (Natl Home Library 1933).

without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or, perhaps both. Knowledge will forever govern ignorance: And a people who mean to be their own Governors, must arm themselves with the power which knowledge gives."[63]

Stuntz argues that transparency can be detrimental to effective security. Transparency, Stuntz contends, makes it harder for government officials to respond to security threats: "For most officials most of the time, the key choice is not between doing right and doing wrong, but between doing something and doing nothing. Doing nothing is usually easier—less likely to generate bad headlines or critical blog posts."[64]

Government officials, however, very often try to do something—the problem is that the "something" they try to do is not the result of an informed and thoughtful policy analysis but often a cheap gimmicky solution that will grab headlines. The choice for officials is not between doing something or nothing—it is between doing something symbolic versus doing something meaningful but more nuanced and complicated. When it comes to security, the symbolic measures often have high civil liberty costs with very little security payoff. Left unexplored are the many more meaningful alternatives where the benefits might outweigh the costs.

Posner and Vermeule do not contest the value of transparency. Instead, they contend that "[g]overnmental decisionmaking is often more visible during emergencies than during normal times."[65] But it is not at all evident that this is the case. Many of the security measures taken by the Bush Administration following September 11 were done under the cloak of secrecy. The Administration's response to leaks about the existence of the programs was outrage. Responding to reports that revealed that he authorized the NSA to conduct warrantless surveillance within the United States, President Bush criticized the media and public officials who provided the information and declared: "As a result, our enemies have learned information they should not have. And the unauthorized disclosure of this effort damages our national security and puts our citizens at risk."[66]

How, exactly, does revealing the fact that Bush authorized the NSA to conduct surveillance—possibly exceeding the limits of his lawful powers—put "our citizens at risk"? Why is every disclosure about

---

63 Letter from James Madison to W.T. Barry (Aug 4, 1822), in Gaillard Hunt, ed, 9 *The Writings of James Madison* 103, 103 (G.P. Putnam's Sons 1910).

64 Stuntz, *Secret Service*, New Republic at 14 (cited in note 10).

65 Posner and Vermeule, *Terror in the Balance* at 55 (cited in note 16).

66 *Bush on the Patriot Act and Eavesdropping*, NY Times A43 (Dec 18, 2005). See also Peter Baker, *President Says He Ordered NSA Domestic Spying; In Radio Address, He Rebukes Democrats for Opposing Renewal of the Patriot Act*, Wash Post A1 (Dec 18, 2005).

the extent of the government's surveillance somehow assisting the terrorists? Far too often, we trust the government when it claims the need for secrecy, but should we? As Mary-Rose Papandrea notes, courts are often far too willing to defer to government claims of secrecy: "[W]hen information arguably involves national security, courts are too timid to force the executive branch to provide a thorough explanation for continued secrecy."[67]

The problem with many data mining programs is that they lack adequate transparency. The reason for the secrecy of the programs is that exposing the algorithms and patterns that trigger identification as a possible future terrorist will tip off terrorists about what behaviors to avoid. This is indeed a legitimate concern. Our society, however, is one of open government, public accountability, and oversight of government officials—not one of secret blacklists maintained by bureaucracies. Without public accountability, unelected bureaucrats can administer data mining programs in ways often insulated from any scrutiny at all. For example, the information gathered about people for use in data mining might be collected from sources that do not take sufficient steps to maintain its accuracy. Without oversight, it is unclear what level of accuracy the government requires for the information it gathers and uses. If profiles are based on race, speech, or other factors that society might not find desirable to include, how is this to be aired and discussed? If a person is routinely singled out based on a profile and wants to challenge the profile, there appears to be no way to do so unless the profile is revealed.

The lack of transparency in data mining programs makes it nearly impossible to balance the liberty and security interests. Given the significant potential privacy issues and other constitutional concerns, combined with speculative and unproven security benefits as well as many other alternative means of promoting security, should data mining still be on the table as a viable policy option? Of course, one could argue that data mining at least should be investigated and studied. There is nothing wrong with doing so, but the cost must be considered in light of alternative security measures that might already be effective and lack as many potential problems. Data mining might prove lucrative to various database companies and other government contractors; it might also provide government officials in various government agencies with new projects to investigate and explore. But dollars spent for data mining are dollars not spent for other programs.

---

67 Mary-Rose Papandrea, *Under Attack: The Public's Right to Know and the War on Terror*, 25 BC Third World L J 35, 79 (2005) (arguing that the Freedom of Information Act and the First Amendment have proved to be insufficient tools for the public to monitor government counterterrorism efforts).

## CONCLUSION

The current security-liberty debate is deeply flawed, resulting in a balancing between security and liberty that is not very meaningful. The scale is rigged so that security will win out nearly all the time. In an age of consequentialist balancing of rights against government interests, it is imperative that the balancing be done appropriately. Security and liberty often clash, but there need not be a zero-sum tradeoff. Liberty interests are generally not achieved by eliminating particular security programs but by placing them under oversight, limiting future uses of personal data, and ensuring that they are carried out in a balanced and controlled manner. Curtailing ineffective security measures is often not just a victory for liberty but for security as well, since better alternatives might be pursued.

The government is currently seduced by data mining. It is not clear, however, that data mining is an effective security measure. Its lack of transparency serves as a major impediment to any meaningful balancing of its security benefits and liberty costs. By exposing security interests to sunlight and heeding liberty interests, the government could ultimately be more accountable to the people. The result might be not only better protection of liberty but also more thoughtful and effective security.