International Journal of Communications Law & Policy

Issue 9, Special Issue on Cybercrime, Winter 2004/2005
(published in cooperation with the Yale Journal of Law &Technology)

ARTICLE

TECHNOLOGY, SECURITY AND PRIVACY: THE FEAR OF FRANKENSTEIN, THE MYTHOLOGY OF PRIVACY AND THE LESSONS OF KING LUDD

K. A. TAIPALE *

This article was jointly reviewed and edited by YALE JOURNAL OF LAW & TECHNOLOGY and INTERNATIONAL JOURNAL OF COMMUNICATIONS LAW & POLICY.

* Kim Taipale, BA, JD (New York University), MA, EdM, LLM (Columbia University), is the executive director of the Center for Advanced Studies in Science and Technology Policy. Mr. Taipale also serves as director of the Global Information Society Project at the World Policy Institute. *Email: tsp-info@advancedstudies.org*.

The author would like to thank Paul Rosenzweig (Heritage Foundation), Heather Mac Donald (Manhattan Institute), Dan Gallington (Potomac Institute), Jay Stanley (ACLU); Michael Froomkin (Miami), Dan Solove (Geo. Wash.), Orin Kerr (Geo. Wash.); Cass Sunstein (Chi.); Eben Moglen (Colum.); Barry Steinhardt (ACLU), Marc Rotenberg (EPIC); Adm. John Poindexter (formerly DARPA), Spike Bowman (FBI); Tal Zarsky, Nimrod Kozlovski, and Jack Balkin of the Yale Information Society Project; the organizers of CyberCrime 2004; and the Editorial Boards of the YALE J. L. & TECH. and the INTL. J. COMM. L. & POL'Y.

International Journal of Communications Law & Policy

Issue 9, Special Issue on Cybercrime, Winter 2004/2005
(published in cooperation with the Yale Journal of Law &Technology)

TECHNOLOGY, SECURITY AND PRIVACY: THE FEAR OF FRANKENSTEIN, THE MYTHOLOGY OF PRIVACY AND THE LESSONS OF KING LUDD

K. A. TAIPALE

This article suggests that the current public debate that pits security and privacy as dichotomous rivals to be traded one for another in a zero-sum game is based on a general misunderstanding and apprehension of technology on the one hand and a mythology of privacy that conflates secrecy with autonomy on the other. Further, political strategies premised on outlawing particular technologies or techniques or seeking to constrain technology through laws alone are second-best – and ultimately futile – strategies that will result in little security and brittle privacy protection.

This article argues that civil liberties can best be protected by employing value sensitive technology development strategies in conjunction with policy implementations, not by opposing technological developments or seeking to control the use of particular technologies or techniques after the fact through law alone. Value sensitive development strategies that take privacy concerns into account during design and development can build in technical features that can enable existing legal control mechanisms and related due process procedures for the protection of civil liberties to function.

This article examines how identification, data aggregation and data analysis (including data mining), and collection technologies intersect with security and privacy interests and suggests certain technical features and strategies premised on separating knowledge of behavior from knowledge of identity based on the anonymization of data (for data sharing, matching and analysis technologies) and the pseudonymization of identity (for identification and collection technologies). Technical requirements to support such strategies include rule-based processing, selective revelation, and strong credential and audit.

I. PRELUDE

At the turn of the century technological development was occurring at a rate that dizzied the mind. These technological developments were bringing a better standard of living to all, yet the gap between the rich and poor was becoming more pronounced. The government, fearful of foreigners, enacted repressive laws and the intellectual elite suggested that the government was too powerful and that charges of treason were too easily leveled.[1]

It was during this period – the beginning of the nineteenth century – that Lady Mary Wollstonecroft Shelley wrote her novel Frankenstein[2] and the Luddite movement was born.[3] It is claimed that Frankenstein and the monster capture "the complex duality of the Romantic soul, the dark as well as the bright side, the violent as well as the benevolent impulses, the destructive as well as the creative urges"[4] So too with advanced information technology and the duality of our concerns with security and privacy.

II. INTRODUCTION

The current public debate that pits security and privacy as dichotomous rivals to be traded one for another in a zero-sum

1 *See generally* CAROLLY ERICKSON, OUR TEMPESTUOUS DAY: A HISTORY OF REGENCY ENGLAND (1986).

2 MARY WOLLSTONECRAFT SHELLEY, FRANKENSTEIN: THE 1818 TEXT, CONTEXTS, NINETEENTH-CENTURY RESPONSES, MODERN CRITICISM (J. Paul Hunter ed., 1996).

3 *See* MALCOLM I. THOMIS, THE LUDDITES: MACHINE-BREAKING IN REGENCY ENGLAND (1972); KIRKPATRICK SALE, REBELS AGAINST THE FUTURE: THE LUDDITES AND THEIR WAR ON THE INDUSTRIAL REVOLUTION: LESSONS FOR THE COMPUTER AGE (1996). *Luddites* was the name given to groups of workingmen in the industrial centers of England who rioted and began breaking knitting machines (called frames, thus "frame breaking"), and later cotton looms, to the introduction of which they attributed unemployment and low wages. The original Luddite movement occurred between 1811 and 1816 and was harshly suppressed by the government. There was no political aim involved and no real organization to the movement. Later worker movements that took up the Luddite banner with a political agenda were the precursors to the industrial labor union movement. Today, the term *Luddite* is used to describe anyone who is perceived to oppose technological developments or change.

4 PAUL CANTOR, CREATURE AND CREATOR: MYTH-MAKING AND ENGLISH ROMANTICISM 108 (1984).

game is based on a general misunderstanding and apprehension of technology on the one hand and a mythology of privacy that conflates secrecy with autonomy on the other. Further, political strategies premised on outlawing particular technologies or techniques or seeking to constrain technology through laws alone are as doomed to failure as Ned Ludd's[5] swing of the sledgehammer – and will result in little security and brittle privacy protection.

Security and privacy are not a balancing act but rather dual obligations of a liberal democracy[6] that present a *wicked problem* for policy makers. Wicked problems are well known in public policy[7] and are generally problems with no *correct* solution. Wicked problems reveal additional complexity with each attempt at resolution and have infinite potential outcomes and no *stopping rule* – that is, the process ends when you run out of resources not when you reach the correct solution.[8] There is no fulcrum point – as is implicit in the balance metaphor – at which point the correct amount of security and privacy can be achieved. Wicked problems occur in a social context and the wickedness of the problem reflects the diversity of interests among the stakeholders.[9] Resolving wicked problems requires

5 The name Luddite is variously attributed as having its origin from Ned Ludlam, the son of a framework knitter, or the mythical figures Ned Ludd or King Ludd. *Compare, e.g.,* Thomis, *supra* note 3, at 11-12 *with* the entry for 'Luddites' in THE COLUMBIA ENCYCLOPEDIA, (6th ed. 2001) at http://www.bartleby.com/65/lu/luddites/html. *See also,* Erickson, *supra* note 1, at 61 ("General Ludd").

6 "In a liberal republic, liberty presupposes security; the point of security is liberty." Thomas Powers, *Can We Be Secure and Free?* 151 PUBLIC INTEREST 3, 5 (Spring 2003). Powers goes on to argue that the politicization of the civil liberties debate has resulted in a false dichotomy – a choice between liberty and security – that is inconsistent with the liberal political foundation on which this country was founded. *Id.* at 16-20 "From [Madison's] point of view, it is clear that there is not so much a 'tension' between liberty and security as there is a duality of our concern with security, on the one hand, and with liberty, on the other." *Id.* at 21.

7 Horst Rittel & Melvin Webber, *Dilemmas in a General Theory of Planning*, 4 POLICY SCIENCES, 155-159 (1973) and Horst Rittel & Melvin Webber, *Planning Problems are Wicked Problems*, in DEVELOPMENTS IN DESIGN METHODOLOGY (N. Cross ed., 1984) 135-144.

8 Stopping when you have a solution to a complex problem that is "good enough" within your resource constraints has been referred to as "satisficing". HERBERT A. SIMON, SCIENCES OF THE ARTIFICIAL 28-30 (3rd Edition 1996, 1969).

9 Jeff Conklin, *Wicked Problems and Social Complexity*, CogNexus Institute White Paper (2003) *at* http://cognexus.org/wpf/wickedproblems.pdf.

an informed debate in which the nature of the problem is understood in the context of those interests, the technologies at hand for resolution, and the existing resource constraints.[10]

In a technologically mediated information society, civil liberties can only be protected by employing value sensitive technology development strategies in conjunction with policy implementations, not by opposing technological developments or seeking to control the use of particular technologies or techniques after the fact through law alone.[11] Value sensitive development strategies that take privacy concerns into account during design and development[12] can build in technical features that enable existing legal control mechanisms for the protection of civil liberties and due process to function.[13]

Code is not law, but code can bound what law, norms and market forces can achieve.[14] Technology itself is neither the

10 "Because of social complexity, solving a wicked problem is fundamentally a social process" and requires "creating shared understanding about the problem, and shared commitment to the possible solutions." *Id.* at 17.

11 *See* Julie E. Cohen, *Symposium: The Law and Technology of Digital Rights Management: DRM and Privacy*, 18 BERKELEY TECH. L.J. 575, 609-617 (2003) (arguing for building privacy protection into Digital Rights Management (DRM) code (in addition to law) by employing value sensitive design and development strategies. "[B]oth judicial and regulatory sanctions are second-best strategies for ensuring effective [privacy] protection for all users. A far more effective method of ensuring that information users actually enjoy the privacy to which they are entitled would entail building privacy into the design of DRM technologies in the first instance." *Id.* at 609).

12 *See* Ben Shneiderman & Anne Rose, *Social Impact Statements: Engaging Public Participation in Information Technology Design*, in BATYA FRIEDMAN, HUMAN VALUES AND THE DESIGN OF COMPUTER TECHNOLOGY. ("Constructive criticism and guidelines for design could help protect us against the adverse ramifications of technology such as … dissatisfaction with privacy protection." *Id.* at 118); *see generally* Batya Friedman et al., *Value Sensitive Design: Theory and Methods* (Draft of June 2003), *at* http://www.ischool.washington.edu/vsd/vsd-theory-methods-draft-june2003.pdf ("Value Sensitive Design is a theoretical grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process.").

13 *See generally* K. A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 2 (2003) [hereinafter, Taipale, *Data Mining*]; Paul Rosenzweig, *Proposals for Implementing the Terrorism Information Awareness System,* 2 GEO. J. L. & PUB. POL'Y 169 (2004).

14 *See* Lawrence Lessig, CODE AND OTHER LAWS OF CYBERSPACE 3-8 (1999) ("[Code] constitute[s] a set of constraints on how you behave. … The code or … architecture … constrain[s] some behavior by making other behavior possible, or impossible."). *Id.* at 89. Lessig writes that behavior is controlled

problem nor the solution, rather it presents certain opportunities and potentials that enable or constrain public policy choice. Technical features alone cannot eliminate privacy concerns, but by incorporating such features into technological systems familiar privacy protecting due process mechanisms (or their analogues) are enabled.[15]

This article examines how identification, data aggregation and analysis (including data mining), and collection technologies currently being considered for use in the context of domestic security intersect with security and privacy interests and suggests certain technical features and strategies that can help ameliorate these concerns. This article proposes that technical development strategies premised on separating *knowledge of behavior* from *knowledge of identity* based on the *anonymization* of data (for data sharing, matching and analysis technologies) and the *pseudonymization* of identity or authorization (for identification and collection technologies) can help protect individual autonomy while still meeting security needs. Technical requirements to support such strategies include rule-based processing, selective revelation, and strong credential and audit.[16]

III. SOME ASSUMPTIONS

This article focuses on the intersection of technology and domestic and national security in the context of the current 'war on terrorism'[17] but the analysis presented herein is equally

(regulated or constrained) through a dynamic interaction of legal rules, social norms, market forces and architecture (or code). *Id.* at 83-99.

15 *See* Rosenzweig, *supra* note 13 (setting out a proposed legal and procedural framework designed to exploit technical features like those described in this Article).

16 *See also* ISAT 2002 Study, *Security with Privacy*, Dec. 13, 2002 (discussing the purely technical aspects of security with privacy), *available at* http://www.taipale.org/references/isat_study.pdf (formerly at http://www.darpa.mil/iao/secpriv.pdf); James X. Dempsey & Paul Rosenzweig, Heritage Foundation, *Technologies That Can Protect Privacy as Information is Shared to Combat Terrorism* (May 26, 2004), *available at* http://www.heritage.org/Research/ HomelandDefense/lm11.cfm (discussing data anonymization, rules permissioning, and immutable audit trails).

17 I use the phrase 'war on terrorism' throughout this article because it is the prevailing metaphor for the current conflict between organized, but generally stateless actors using asymmetric means, including politically or religiously-motivated violence, against U.S. and other global institutional interests. *But*

applicable to law enforcement more generally – subject, however, to certain caveats. In particular, to the extent that there is a relationship between law enforcement applications and privacy concerns, the lesser the crime targeted the greater the hurdle for any new technology or wider use that implicates those concerns.[18]

It is beyond the scope of this article to attempt to delineate precisely where the line between preemptive and reactive strategies should be drawn by delimiting particular types of crimes that meet particular criteria, or by specifying which government organs or agencies should be permitted particular uses. Rather, this article is primarily concerned with the over-arching issues involved in employing advanced information technologies to help identify and find actors who are hidden among the general population and who have the potential for creating harms of such magnitudes that a consensus of society requires that government adopt a preventative rather than reactive approach.[19]

The events of 9/11 have put to rest any doubts that we face a formidable threat from certain organized but generally state-less forces that are intent on inflicting serious damage on US interests, including the killing of large numbers of innocent civilians.[20] Regardless of one's view of the particular political

---

cf. Terry Jones, *Why Grammar is the First Casualty of War*, LONDON DAILY TELEGRAPH, Dec. 1, 2001. ("How do you wage war on an abstract noun?"). *But see generally* note 113 *infra* (discussing metaphor).

18 *See* Taipale, *supra* note 13, at n.40.

19 In response to the attacks of 9/11, the U.S. Department of Justice and the FBI have undertaken to reorganize their mission from the traditional role of investigating and prosecuting crime that has already occurred to that of preventing future acts of terrorism. *See* U.S. Department of Justice, *Fact Sheet: Shifting from Prosecution to Prevention, Redesigning the Justice Department to Prevent Future Acts of Terrorism* (May 29, 2002), *available at* http://www.fas.org/irp/news/2002/05/fbireorganizationfactsheet.pdf

20 *See, e.g.*, National Commission on Terrorist Attacks Upon the United States, *Overview of the Enemy* (2004) *available at* http://www.9-11commission.gov/hearings/hearing12/staff_statement_15.pdf. In addition to the almost 3,000 civilian deaths, the terrorist attack on the World Trade Center towers has been variously estimated to have caused between $50 billion and $100 billion in direct economic loss. Estimates of indirect losses exceed $500 billion nationwide. General Accounting Office U.S. Congress, GAO-02-700R, *Review of Studies of the Economic Impact of the September 11, 2001 Terrorist Attacks on the World Trade Center* (2002), *available at* http://www.gao.gov/new.items/d02700r.pdf.

strategy being used in response, the current threat is real and, among other things, we need to enlist technology, and reform organizational structures, to help counter this threat.[21] To date we have not taken sufficient advantage of information technology to help secure the nation against these kinds of threats.[22] However, technology cannot provide security by itself, and we also need to adopt new organizational structures and procedures to take advantage of opportunities that information technology can make available.[23]

At the same time, however, we must recognize that the use of these technologies and procedures can be intrusive on certain privacy interests that help protect individual freedom and political autonomy, and are core to our political liberties.[24] These interests must also be protected. It has become cliché, yet remains axiomatic, that every compromise we make to civil liberties in the 'war on terrorism' is itself a victory for those who

21 *See, e.g.*, Markle Foundation, *Protecting America's Freedom in the Information Age: A Report of the Markle Foundation Task Force* at 1-3 (2002), *available at* http://www.markletaskforce.org/ [hereinafter First Markle Report] (the nation must capitalize on its leadership in information technology).

22 *See, e.g.*, JOINT INQUIRY INTO THE INTELLIGENCE COMMUNITY ACTIVITIES BEFORE AND AFTER THE TERRORIST ATTACKS OF SEPTEMBER 11, 2001 HOUSE PERMANENT SELECT COMM. ON INTELLIGENCE & SENATE SELECT COMM. ON INTELLIGENCE, H. REP. NO. 107-792, S. REP. NO. 107- 351 (2002) ("While technology remains one of this nation's greatest advantages, it has not been fully and effectively applied in support of U. S. counter terrorism efforts. Persistent problems in this area include … a reluctance to develop and implement new technical capabilities aggressively.") [hereinafter Joint Inquiry Report] at xvi.

23 *See, e.g.*, First Markle Report, *supra* note 21, at 9 ("Though we need technology to secure our nation, a successful domestic intelligence and information strategy should start with the way we organize our people to take advantage of innovation."); Markle Foundation, *Creating a Trusted Network for Homeland Security: Second Report of the Markle Foundation Task Force* at 8-9 (2003) *available at* http://www.markletaskforce.org/ [hereinafter Second Markle Report] ("[building a networked community for homeland security] requires changes in policies, procedures, and the use of technology."); *See also* Committee on Science and Technology for Countering Terrorism, National Research Council, *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism* (2002), *available at* http://www.nap.edu/html/stct/index.html.

24 *See generally* ALAN WESTIN, PRIVACY AND FREEDOM (1967) and Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609 (1999).

would like to destroy our way of life.[25] Terrorism itself is a complex problem. Eliminating the current terrorist threat involves a mix of three essential strategies.[26]

First, we must eliminate political preconditions to terrorism. We must solve unresolved conflict throughout the world, end lack of economic and political opportunity, and generally make the world safe for democratic processes and civil society. It is beyond the scope of this article to discuss these issues fully.[27] Second, we must harden targets.[28] Of course, target hardening generally only influences an adversary's target-selection process – when preferred targets are hardened terrorists will seek softer targets like any rational enemy. More importantly, we cannot harden all potential targets – not even all high-value targets.[29] Thus, discussing locking cockpit doors

25 "Those who would give up Essential Liberty to purchase a little temporary Safety, deserve neither Liberty nor Safety." Attributed to Benjamin Franklin. Benjamin Franklin, Pennsylvania Assembly: Reply to the Governor, November 11, 1755. 6 THE PAPERS OF BENJAMIN FRANKLIN 242 (Leonard W. Labaree, ed., 1963).

There are three ways that we as a society can 'lose' to terrorism – first, we can fail to provide security and future successful terrorist attacks could undermine the confidence and optimism required to maintain our economic and political system, second, we can bankrupt our economy by incurring defense costs not appropriately apportioned to actual risk or by imposing security burdens that undermine its competitiveness, *see* note 29 *infra*, or, third, we can create a totalitarian society no longer worth maintaining. *See generally* K. A. Taipale, *Losing the War on Terror*, Center for Advanced Studies (forthcoming Winter 2005), on file with the author [hereinafter, Taipale, *Losing the War*].

26 *Id.*

27 But see *id.*, arguing in part that better managing the effects of globalization on local economies, investing foreign aid in human rights, women's equality, secular education and family planning, and developing a rational and sustainable energy policy are fundamental steps to achieving a long-term solution. See also JOSEPH S. NYE, SOFT POWER: THE MEANS TO SUCCESS IN WORLD POLITICS (2004); ZBIGNIEW BRZEZINSKI, THE CHOICE: GLOBAL DOMINATION OR GLOBAL LEADERSHIP (2004); THE BATTLE FOR HEARTS AND MINDS: USING SOFT POWER TO UNDERMINE TERRORIST NETWORKS (Alexander T. J. Lennon, ed. 2003).

28 Target hardening refers to defensive strategies such as employing guards or physical barriers to make it more difficult for terrorists to act against a specific target. *See* White House, National Strategy for Physical Protection of Critical Infrastructures and Key Assets (2003) ("protecting our critical infrastructures and key assets from physical attack") *available at* http://www.whitehouse.gov/pcipb/physical.html.

29 "The nation could never sufficiently harden all potential targets against attack." Second Markle Report, *supra* note 23, at 1. Indeed, we stand a good chance of bankrupting our economy by engaging in a vulnerability-based,

is not an "alternative strategy" to employing information technology as some have implied,[30] rather physical defense is a discrete strategy that needs to be considered on its own merits. Which brings us to the third strategy – that is, we must identify terrorists and preempt terrorist acts.[31] To do this requires in part the better use of information and the better use of advanced information technology to share relevant information and to help sort relevant from irrelevant information.[32]

This article concerns itself with the use of advanced information technologies in support of this third strategy. Thus, this article assumes that there is some category of malicious actor – terrorist, if you will – for which there exists a political consensus for proactive investigative strategies intended to

---

rather than threat-based, defensive strategy in which all possible targets are protected for political reasons rather than concentrating resources on the most likely targets or threats. *See* Taipale, *Losing the War*, *supra* note 25. *Cf. Bin Laden: Goal is to Bankrupt U.S.*, CNN.COM (Nov. 1, 2004) *available at* http://www.cnn.com/2004/WORLD/meast/11/01/ binladen.tape/index.html ("We are continuing this policy in bleeding America to the point of bankruptcy," statement attributed to Osama bin Laden).
Further, the issue of cost is particularly relevant in homeland security (as contrasted with national security generally) given that approximately 85% of critical infrastructure to be protected is in the private sector. *See* Richard Rector, *Infotech and the Law: Homeland security -- Who pays for protecting infrastructure?* 17 WASH. TECH. (Mar. 10, 2003) *available at* http://www.washingtontechnology.com/news/17_23/federal/20234-1.html.
The cost of infrastructure protection is essentially a "security tax" burden imposed throughout the economy. *See* Taipale, *Losing the War*, *supra* note 25; *see also* Kenneth Rogoff, *The Cost of Living Dangerously: Can the Global Economy Absorb the Expenses of Fighting Terrorism?* FOREIGN POLICY at 70 (Nov./Dec. 2004).
Further, there is a significant cost to functionality of the system itself – that is, the friction imposed by various security measures on the free flow of commerce, capital, talent, and ideas, for example, impeding air travel or the shipment of goods through the use of physical searches, impeding the free flow of capital through anti-money laundering requirements, making the bureaucratic cost of obtaining a student visa prohibitive resulting in a dearth of graduate student research assistance, or removing from the public domain essential scientific information, etc. – all of which can undermine our long-term competitiveness.
30 *See, e.g.*, Laura W. Murphy, ACLU, Remarks at the National Press Club, Washington, DC (Aug. 25, 2003) *available at* http://www.aclu.org/SafeandFree/ SafeandFree.cfm?ID=13355.
31 *See, e.g.*, note 19 *supra*.
32 *See* Second Markle Report, *supra* note 23, at 11 (need to enhance the government's "ability to discern indicators of terrorist activity amid overwhelming amounts of information").

prevent future acts of terrorism. The one conclusion that seems clear from the report of the Congressional Joint Committee looking into 9/11 is "that terrorism cannot be treated as a reactive law enforcement issue, in which we wait until after the bad guys pull the trigger before we stop them."[33] But, reconciling the need for preventative strategies with traditional notions of due process and individual freedom is a complex task.

It is also important to recognize that technology alone cannot provide security; at best – and even then only if used within appropriately designed security systems – it can help better allocate scarce security resources towards more effective uses.[34] There is no technological silver bullet that will provide absolute security nor is there any technical solution that will absolutely protect privacy.[35] Technology alone is not a solution to either problem; but neither are simple laws prohibiting the use of specific technologies or particular techniques the answer in themselves.[36] Instead, some complex system – a social construction[37] – combining organizational structures, rules and

33 *See* Editorial, *The Limits of Hindsight*, WALL ST. J., July 28, 2003, at A10 (responding to the release of the Joint Inquiry Report, *supra* note 22).

34 *Cf.* Taipale, *supra* note 13, at 21.

35 Recognizing that no system – technical or not – can provide absolute security or absolute privacy also means that no technical system or technology ought to be burdened with meeting an impossible standard for perfection, especially prior to research and development. Technology is a tool and as such it should be evaluated by its ability to either improve a process over existing or alternative means or not. Opposition to research programs on the basis that the technologies "might not work" is an example of what has been called the "zero defect" culture of punishing failure, a policy that stifles bold and creative ideas. At least one commentator has characterized such opposition to risk-taking as "downright un-American." *See, e.g.,* David Ignatius, *Back in the Safe Zone*, WASH. POST, Aug. 1, 2003, at A:19 (discussing the knee-jerk opposition to a "terrorist futures market"). *See also* discussion of 'confidence intervals', *infra.*

36 *See* Taipale, *supra* note 13, at n.32 (arguing that privacy protection based on law alone is "brittle" in an engineering sense, that is, any breach results in catastrophic failure of protections. "If technologies are developed without privacy protecting features built in but outlawed for law enforcement or domestic security purposes and then the laws are changed in the future, for example, in response to a new terrorist attack, the then existing technologies will not be capable of supporting implementation policies that provide any privacy protection." *Id.*); *see also, id.,* at n.28, describing various recent legislative attempts to outlaw the development or use of certain technologies, techniques or programs.

37 *See generally* Wiebe E. Bijker, OF BICYCLES, BAKELITES, AND BULBS: TOWARD A THEORY OF SOCIOTECHNICAL CHANGE (1997); Trevor J. Pinch & Wiebe E. Bijker, *The Social Construction of Facts and Artifacts*, *in* THE

procedures, and technologies must be developed (or must evolve) together to ensure that we achieve better security while protecting privacy and civil liberties.[38]

This article examines the conflict between security and privacy in the context of advanced digital information systems and their related technical characteristics in order to achieve some better understanding of potential solutions – organizational, procedural and technical – to achieving security while protecting privacy.

IV. FRANKEN-TECH: THE FEAR OF TECHNOLOGY

Cass Sunstein, among others, has written much about the notion that people act apparently irrationally with regard to certain risk trade-offs.[39] For example, during the recent DC sniper episode, citizens of one state would drive to another to get gas rather than use a local gas station for fear of the sniper – thus exposing themselves to greater statistical risk of death from a traffic fatality than from an actual sniper attack. So too, people who fear flying and prefer to drive may actually expose themselves to a much greater risk of injury or death on the highway.[40]

SOCIAL CONSTRUCTION OF TECHNOLOGICAL SYSTEMS (Wiebe E. Bijker et al. eds., 1994) (describing technological development as social construction); SHAPING TECHNOLOGY/BUILDING SOCIETY (Wiebe E. Bijker & John Law eds., 1992).

38 *See, e.g.,* Second Markle Report, *supra* note 23 at 8-9 ("Building a networked community for Homeland Security.").

39 Cass R. Sunstein, *Terrorism and Probability Neglect*, 26 JOURNAL OF RISK AND UNCERTAINTY 121 (2003), reprinted in THE RISKS OF TERRORISM (W. Kip Viscusi ed. 2003) [hereinafter, Sunstein, *Terrorism*], and Sunstein, *Probability Neglect: Emotions, Worst Cases, and Law*, U. CHICAGO LAW & ECONOMICS, Olin Working Paper No. 138. (November 2001) [hereinafter, Sunstein, *Emotions*] *available* at http://ssrn.com/abstract=292149. *See also*, Sunstein, RISK AND REASON (2002).

Much of Sunstein's work in this area builds on that of Amos Tversky and Daniel Kahneman. *See generally* Amos Tversky & Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974) [hereinafter Tversky, *Judgment under Uncertainty*]; JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, (Daniel Kahneman, Paul Slovac & Amos Tversky, eds., 1982) [hereinafter, KAHNEMAN, JUDGMENT UNDER UNCERTAINTY].

40 *Cf.* Jane Brody, *Don't Lose Sight of Real, Everyday Risks*, N. Y. TIMES, October 9, 2001, at F6.

Sunstein identifies three noteworthy points about how fear impacts risk analysis.[41] The first is that without actual knowledge of a particular risk, people rely on the *availability heuristic,* through which they assess a risk by reference to whether a readily available example of the outcome can be recalled,[42] that is, people exhibit a greater fear of a risk the more they are reminded of actual or similar outcomes. The second is that people generally show a disproportionate fear of risks that are either unfamiliar or appear hard to control,[43] that is, people exhibit a greater fear of a risk from an unfamiliar or novel source even if its probability is slight. And, the third is that people are prone to what Sunstein calls *probability neglect* – that is, in the face of risks with high emotional content, emotion plays a significant role in obscuring 'rational' choice.[44]

These three impacts are also observed in policy choice. Sunstein has documented many instances in which media attention to a particular environmental issue, for example, Love Canal, Alar, or asbestos in schools, has resulted in 'irrational' policy choices not grounded in objective assessments of relative risk.[45] In these instances the media focus essentially determines the emotional state of the polity.[46] Further, the media attention itself is often manipulated by what Sunstein calls "availability entrepreneurs" who take advantage of a particular event to publicize (and thus elevate) a relatively unlikely risk in order to further their own particular agenda.[47]

Thus, the public debate on policy issues – particularly on complex issues or novel problems with unknown consequences – is often dominated by these information entrepreneurs, including activists and the media itself, who attempt to engender information cascades to further their own particular

41 Sunstein, *Terrorism, supra* note 39 at 121-122.
42 *Id.* The availability heuristic was first described by Kahneman and Tversky, *supra* note 39.
43 Sunstein, *Terrorism, supra* note 39 at 121, citing P. SLOVAC, THE PERCEPTION OF RISK (2000).
44 "Sometimes people focus on the worst possible case, which triggers strong emotions. When this is so, people fail to inquire into the probability that the worst case will occur. In such cases, emotions lead to what I will call probability neglect." Sunstein, *Emotions*, *supra* note 39, at 4.
45 *Id.* at 18-21.
46 *See generally* Timur Kuran and Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683, 691–98 (1999).
47 *See generally* SUNSTEIN, RISK AND REASON, *supra* note 39, at 78-98 ("Chapter 4: This Month's Risk").

agenda.[48] "An [information] cascade is a self-reinforcing process of collective belief formation by which an expressed perception triggers a chain reaction that gives the perception increasing plausibility through its rising availability in public discourse."[49] The result is often that relatively minor risks can be overblown causing a high level of social anxiety, the expenditure or misallocation of significant resources, and the imposition of costly regulation in situations where other risks, of greater magnitude, are ignored.[50]

This same phenomenon skews the public debate on technology, security and privacy. The availability of information privacy horror stories (in particular, the prevalence of identity theft, spam and hacker stories in the media),[51] and the general mistrust in government agencies to handle personal information appropriately,[52] combined with a general apprehension about technology[53] and how it works,[54] and the natural anxiety relating to disclosure of personal, particularly intimate, information – all spurred on by the privacy lobby[55] – has created

48 I do not mean to imply that these actors are not justified in their concerns, only that their particular focus comes to dominate the information flow and their rhetoric sets the terms of the debate.
49 Kuran and Sunstein, *supra* note 46 at 684.
50 Compare, for example, Alar with tobacco. *See generally*, Kuran and Sunstein, *id.* at 683-768; SUNSTEIN, RISK AND REASON, *supra* note 39 at 78-98.
51 Although these risks have little to do directly with policies relating to government access to data, their prevalence in the public conscience tends to add to the general apprehension about control of personal information in a networked environment.
52 *See, e.g.*, Eric J. Sinrod, *Do you trust Big Brother with your personal information?* USA TODAY, Feb. 5, 2004, *available at* http://www.usatoday.com/tech/columnist/ericjsinrod/2004-02-05-sinrod_x.htm.
53 *See generally* LEWIS MUMFORD, MYTH OF THE MACHINE: TECHNICS AND HUMAN DEVELOPMENT (1963, 1934); JACQUES ELLUL, THE TECHNOLOGICAL SOCIETY (1964); NEIL POSTMAN, TECHNOPOLY: THE SURRENDER OF CULTURE TO TECHNOLOGY (1993); TECHNOLOGY, PESSIMISM, AND POSTMODERNISM (Yaron Ezrahi, *et al.*, eds. 1994); but *cf., e.g.,* GEORGE GILDER, TELECOSM (2000) (exhibiting an exuberant optimism in a technology determined future).
54 *Cf.* Rosenzweig, *supra* note 13, at n.6 ("Even among computer professionals there is substantial misunderstanding … [but] those with the seeming greater familiarity with the technology are less apocalyptic in their reactions.").
55 By privacy lobby I mean those individuals or institutions whose political *raison d'etre* (and fundraising) is, at least in part, shaped, driven or determined by the privacy issue. The privacy lobby includes both civil libertarians on the left and libertarians on the right. *See, e.g.*, *Barr to join*

a public anxiety about electronic privacy[56] out of proportion to the actual privacy risks and has obscured discussion of the very real threats posed by either failing to provide security or by misallocating security resources.[57]

Anecdotal support for the notion that there is an unreasonable fear based on unfamiliarity with the technology underlying the public debate on privacy can be found by drawing an analogy with early concerns about the use of credit cards online. While people do not think twice now about using their credit cards online, there was much consternation in the late 1990s when even the long-term success of online commerce was questioned based on the unwillingness of consumers to use credit cards online – a fear wholly out of proportion to the actual risk and one that never entered their minds when they handed over their card to a minimum wage busboy or threw their credit card receipt in a public trash receptacle. Some would argue that the overblown concern for electronic privacy may be the 'risk of the moment' based in part on a lack of awareness or understanding of the nature and consequences of current technology developments[58] and the novelty of the threat.

ACLU, FOX NEWS, Nov. 27, 2002, *available at* http://www.foxnews.com/story/0,2933,71553,00.html.

56 *See, e.g., America's Number One Fear In The 21st Century Is Loss Of Personal Privacy*, ST. PETERSBURG TIMES, Nov. 3, 1999 at 18A.

57 *See* Heather MacDonald, *Total Misrepresentation*, WEEKLY STANDARD, Jan. 27, 2003, *available at* http://www.weeklystandard.com/Content/Public/Articles/000/000/002/137dvufs.asp?pg=2 ("[critics of initiatives to improve intelligence] barely mention the motivation for the initiative, if at all. [They write] … without once referring to terrorism or the 9/11 strikes.")

58 *Cf.* comments by Kevin Ryan, CEO of DoubleClick, at the Reuters Technology Media and Telecommunications Summit (Feb. 24, 2004) as reported by Reuters (Feb. 25, 2004):

> Ryan suggested that privacy concerns have eased over the years, similar to how many people have relaxed about using their credit cards online. While people don't think twice now about using their credit cards for online purchases, polls showed that Internet users in the late 1990s were more afraid of fraud, he said.
>
> "I said the same thing many, many years ago, that I thought privacy concerns would follow the credit card fraud concerns," he said. "What happened was the actual risk wasn't that great. In fact, people started to realize that nothing is 100 percent safe ever."

While some might argue that the government has used the fear of terrorism (the actual threat to any particular individual from terrorism, even in 2001, was a relatively low probability risk) to push policies without adequate public debate,[59] so too, others could argue that the "privacy lobby" has used fear of electronic privacy intrusion – wholly disproportionate to its actual injury or risk to civil liberty – to oppose technological developments and further their own agenda.[60]

V. THE PRIVACY NORM PROSELYTIZERS: A FETISH FOR SECRECY

A significant problem in determining policy in this area is that privacy means different things to different people.[61] It is beyond the scope of this article to definitely define privacy or reconcile competing views.[62] However, much of the public debate about the use of technology seems to take place within an unexamined mythology of privacy – a mythology that conflates privacy with absolute secrecy on the one hand and the maintenance of absolute secrecy with liberty on the other. But, this deified notion of privacy based on absolute secrecy – that is, keeping others from knowing what we are doing by emphasizing concealment[63] – confounds two simpler ideas: knowing what someone *does* (behavior) and knowing who someone *is* (identity). Further, it is based on a presumed privacy entitlement for

59 *See, e.g.*, Albert Gore, Remarks to moveon.org (November 9, 2003) at http://www.moveon.org/gore/speech.html

60 *See, e.g.*, Heather Mac Donald, *What We Don't Know Can Hurt Us*, 14 CITY JOURNAL (Spring 2004) *available at* http://www.city-journal.org/html/14_2_what_we_dont_know.html.

61 "Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all." Robert C. Post, *Three Concepts of Privacy*, 89 GEO. L.J. 2087, 2087 (2001); "Perhaps the most striking thing about the right to privacy is that nobody seems to have any very clear idea what it is." Judith Jarvis Thomson, *The Right to Privacy*, 4 PHIL. & PUB. AFF. 295-314 (1975).

62 *See* Taipale, *supra* note 13, at 50-57 (for an overview of competing views).

63 *Cf.* Daniel J. Solove, *Digital Dossiers and the dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1086 (critiquing the Supreme Court's conceptualization of privacy premised on "a form of total secrecy" and safeguarding only "intimate information that individuals carefully conceal.")

electronic data that exceeds that afforded paper-based records or real-world experience.[64]

This perception of a privacy entitlement arose not by accident or necessity, but from the intentional action of what Steven Hetcher calls norm proselytizers (what I refer to as the privacy lobby and Sunstein might call availability entrepreneurs) who have an interest in promoting online privacy.[65] Nevertheless, it is not my intention to minimize the privacy interests at stake here.[66] Quite the contrary, I argue that we should insist on value sensitive development strategies that build in technical constraints; that we subject the development and use of these technologies to strict authorization, oversight, and judicial review; and that we use advanced technical means to "watch the watchers" to prevent abuse or misuse.

However, we face one of two inevitable futures – one in which technologies are developed with privacy protecting values and functions built into the design or one in which we rely solely on legal mechanisms and sanctions to control the use of technologies that have been developed without regard to such protections.[67] In my view, it is the fetish for absolute secrecy

64 See, for example, the recent opposition on 'privacy' grounds by many of the leading self-styled 'privacy groups' to Gmail, a free, web-based email service offered by Google in which users *consent* to having their email scanned automatically so that topic-relevant ads can be served. "Privacy fundamentalists … insist that new services they believe to be harmful should be banned, even if consumers are clamoring for them." Declan McCullagh, *Gmail and Its Discontents*, NEWS.COM (April 26, 2004) *available at* http://news.com.com/2010-1032-5199224.html.

65 Steven Hetcher, *Norm Proselytizers Create a Privacy Entitlement in Cyberspace*, 16 BERKELEY TECH. L. REV. 877 (2001).

66 Indeed, I agree with Marc Rotenberg, "Privacy will be to the information economy of the next century what consumer protection and environmental concerns have been to the industrial society of the 20th century." James Gleick, *Big Brother is Us*, N. Y. TIMES MAGAZINE (September 29, 1996). Marc Rotenberg is the executive director of the Electronic Privacy Information Center (EPIC), a leader of the anti-Gmail lobby, *see* note 64 *supra*.

67 Taipale, *supra* note 13, at 12. There is no realistic scenario under which the development of these technologies is simply halted. The emergence of digital technology has changed certain underlying base conditions to information management. First, the cost of data retention is less than the cost of selective deletion, and, second, the cost of indiscriminate data collection is less than the cost of selective acquisition. Therefore, in general, more data will be collected and retained throughout the information economy. To manage these vast data volumes with the same or fewer

promulgated by the privacy lobby that precludes or delays the development of appropriate technologies to improve security while also protecting civil liberties, and leaves us with little security and brittle privacy protection.

Thus, for example, I have previously argued that last year's defunding by Congress of DARPA's Information Awareness Office (IAO) and its Terrorism Information Awareness (TIA) program and related projects[68] was a pyrrhic victory for civil liberties as that program provided a focused opportunity around which to publicly debate the rules and procedures for the future use of these technologies and, most importantly, to oversee the development of the appropriate technical features required to support any concurred upon implementation or oversight policies to protect privacy.[69]

In any case, privacy (particularly any legal or moral claim for the protection of privacy) should be based on the need to protect individual political and personal autonomy, not simply as a characteristic of data for its own sake.[70] Thus, a fetish for

---

analytical resources means that technologies to automate or augment these processes will be developed. Thus, the question is under what circumstances and by whom they will be developed and used. In my view, the choice is between open government research and deployment according to established norms of due process, or classified government and proprietary commercial programs not subject to traditional controls.

68 DARPA is the Defense Advanced Research Project Agency. IAO was the program office for TIA and related projects. TIA was the systems level project to integrate various advanced information technologies, including language translation, data aggregation and data analysis, and others into a "counterterrorism information architecture" in order "to better detect, classify, and identify potential foreign terrorists." *See IAO Report to Congress regarding the Terrorism Information Awareness Program* at 3 (May 20, 2003) in response to Consolidated Appropriations Resolution, 2003, No. 108-7, Division M, §111(b) [signed Feb. 20, 2003] [hereinafter IAO Report]. For a more detailed discussion, including a description of the various IAO projects, see Taipale, *supra* note 13, at 35-50. The IAO and TIA program were defunded by Congress in October 2003. *See id.* at n.28; *see also* Carl Hulse, *Congress Shuts Pentagon Unit Over Privacy*, N.Y. TIMES, Sept. 26, 2003, at A20.

69 Taipale, *supra* note 13, at 48-50 (defunding TIA has resulted in research moving into classified or commercial programs not subject to public oversight).

70 *See* Mark Alfino & G. Randolph Mayes, *Reconstructing the Right to Privacy*, 29 SOC. THEORY & PRACTICE 1–18 (2003) (arguing privacy is a moral right of the individual to protect autonomy and distinguishing theories based on maintaining *informational* privacy).

absolute secrecy of innocuous data (or voluntarily produced data) that results in alternative intrusions or harms – say a physical search at the airport (or physical harm from lack of security) – is suspect and should be questioned.[71]

Additionally, the brittle nature of privacy protection based solely on law needs to be considered.[72] If technologies are developed without privacy protecting features built in but outlawed for law enforcement or domestic security purposes, those laws can be changed in the future in response to a new terrorist attack, and the then existing technologies will not be capable of supporting implementation policies that provide any privacy protection at all.[73]

Post hoc analyses of the events of 9/11 have revealed that much relevant information existed but intelligence agencies and law enforcement were unable to "connect the dots."[74] It would be an unusual polity that now demanded accountability from its representatives for being unable to connect the dots from existing datasets to prevent terrorist acts[75] yet denied them the available tools to do so, particularly if there were to be another catastrophic event.

Thus, simple opposition to government research projects or outlawing the use of particular technologies or techniques seems a second-best – and ultimately futile – strategy; one that leaves us dependent on classified programs or proprietary

71 *Cf.* Paul Rosenzweig, *Civil Liberties and the Response to Terrorism*, 42 DUQ. L. REV. 663, 715 (2004) (discussing the trade offs between electronic data disclosure and physical body searches). *See also* Maureen Dowd, *Hiding Breast Bombs*, N. Y. TIMES Op-Ed, Nov. 25, 2004, at 35; Joe Sharkey, *Another Shoe Drops on the Subject of Airport Security*, N. Y. TIMES, Nov. 30, 2004, at 9. (both discussing physical abuse by TSA airport screeners).
Also, note the opposition by the Electronic Privacy Information Center (EPIC) to the voluntary participation in the "registered traveler" program being tested by the Transportation Security Administration. *See* EPIC Alert 11:13, July 12, 2004 at http://www.epic.org/alert/EPIC_Alert_11.13.html; Privacy Act Notice, Transportation Safety Administration, Docket No. TSA-2004-17982, June 1, 2004.
72 Here I mean privacy protection that is brittle in an engineering sense – that is, any breach results in catastrophic failure.
73 Taipale, *supra* note 13, at n.32.
74 *See, e.g.*, First Markle Report, *supra* note 21 at 28, Illustration 2; *see also* Taipale, *supra* note 13, at n.3; see *generally* the National Commission on Terrorist Attacks Upon the United States, *Final Report* (July 2004).
75 *See id.*

commercial interests to develop security technologies[76] and laws alone to protect privacy. A more effective strategy for the protection of privacy and civil liberties while improving security is to build in technical features that support those values in the first place.

The early Luddites resisted the introduction of technology by smashing frames, and they were imprisoned or shipped off to Australia accomplishing little; later movements in their name adapted to the introduction of new technologies by forming organizational structures – the precursors to the modern labor union – and procedures – collective bargaining – to control the terms under which new technology was to be developed and deployed. Perhaps there is a lesson for privacy advocates to be learned from King Ludd.[77]

VI. PRIVACY INTERESTS AT STAKE

There can be no doubt that vital privacy interests are at stake. We must preserve the general culture of freedom in America[78] and do everything in our power to maintain, improve

76 Another problem with defunding government research, particularly DARPA projects, is that government research in these areas tends to be customer- or solution-driven (that is, specifically developed to solve intelligence and law enforcement needs), whereas commercial research tends to be vendor-driven (that is, product is developed that meets vendor needs). *See* Taipale, *Losing the War*, *supra* note 25.
Further, unrealistic restrictions on government access to information or use of technologies for legitimate needs will result in the further "outsourcing" of government information needs to the private sector with the result of less public oversight or controls. *See, e.g.,* Robert O'Harrow, Jr., *Bahamas Firm Screens Personal Data to Assess Risk: Operation Avoids U.S. Privacy Rules*, WASH. POST A:01 (Oct. 16, 2004); Eric Lichtblau, *Homeland Security Department Experiments with New Tool to Track Financial Crime,* N. Y. TIMES A:48 (Dec. 12, 2004).
77 *See generally* Thomis, *supra* note 3; Sale, *supra* note 3.
78 *See generally* Olmstead v. U.S., 277 U.S. 438, 478 (1928) (J. Brandeis, dissenting):

> The makers of our Constitution undertook to secure conditions favorable to the pursuit of happiness. They recognized the significance of man's spiritual nature, of his feelings, and of his intellect . . . .They sought to protect Americans in their beliefs, their thoughts, their emotions, and their sensations. They conferred, as against the government, the right to be let alone . . . . To protect that right, every unjustifiable intrusion by the government upon the privacy of

and protect it.[79] Individual freedom is the basis on which our country was founded and its incorporated values stand at the core of our Constitution and Bill of Rights.[80] Thus, we must stand ever vigilant to potential dangers to our civil liberties.[81]

Nevertheless, rights incur responsibilities.[82] Security and liberty are dual obligations and we cannot slight one for the other.[83] It should be remembered that the Fourth Amendment implicitly recognizes this duality because – in the words of Amitai Etzioni – the "prohibition on unreasonable searches is not accorded more weight than the permission to conduct reasonable searches."[84] In past crises, particularly when they have threatened national security, many have been willing to sacrifice civil liberties in the short-term in order to meet the particular emergency or challenge.[85] In many cases, we as a

---

> the individual, whatever the means employed, must be deemed a violation of the Fourth Amendment.

See also Florida v. Riley, 488 U.S. 445, 456 (1989) (J. Brennan, dissenting) (in considering whether government surveillance is reasonable, the Court needs to consider whether "the amount of freedom remaining to citizens would be diminished to a compass inconsistent with a free and open society.")

79 *See generally* ALEXANDER HAMILTON, ET AL., THE FEDERALIST (Benjamin Wright ed., 1961). *But cf.* FAREED ZAKARIA, THE FUTURE OF FREEDOM (2003) (arguing that democracy and freedom are in tension and that the American model of democracy may not have universal applicability.)

80 Freedom must be preserved for practical reasons as well as to maintain these noble traditions because it is the foundation of our national economic and political strength. Freedom for individuals and ideas to compete with as little governmental interference is what makes our system a powerful magnet for human development. Imposing unwarranted civil liberty burdens will make us less competitive in attracting the trade, capital, and talent that we need to maintain global economic leadership. *See* Taipale, *Losing the War*, *supra* note 25.

81 *See generally* THE WAR ON OUR FREEDOMS: CIVIL LIBERTIES IN AN AGE OF TERRORISM (Richard C. Leone & Greg Anrig, Jr. eds., 2003); and note 80 *supra*.

82 *See* AMITAI ETZIONI, THE SPIRIT OF COMMUNITY: RIGHTS, RESPONSIBILITIES AND THE COMMUNITARIAN AGENDA (1993).

83 Powers, *supra* note 6, at 21.

84 AMITAI ETZIONI, THE LIMITS OF PRIVACY 206 n.77 (1999).

85 *See* Rosenzweig, *supra* note 71, at 667-670 ("The Lessons of History"). For a more detailed history of these events, *see* Geoffrey Stone, *Civil Liberties in Wartime*, 28 J.S. CT. HIST. 215 (2003); s*ee also* WILLIAM H. REHNQUEST, ALL THE LAWS BUT ONE (1998).

nation later came to regret those actions as having gone too far.[86]

In meeting the current challenge of international terrorism we are confronted with two additional complexities. First, the 'war on terrorism' may be one with no definable end, thus we need to develop doctrine and procedures that serve to protect important values from the outset and that can be maintained indefinitely. We cannot sustain "emergency procedures" for any length of time.[87] Second, we face a threat from actors who as part of their strategic and tactical doctrine move among the general population and take advantage of our open society to mask their own organization and activities.[88] The task therefore is not to defend against outsiders but to identify and investigate potentially malicious actors from within the population without undermining or compromising the freedom and autonomy of the vast majority of innocent people.

Therefore, neither demonizing a minority nor engendering suspicion of everyone is a viable or acceptable outcome – however, neither is undermining legitimate security needs by deifying absolute secrecy as the only means of protecting individual autonomy. The particular privacy concerns most implicated by employing advanced information technologies for proactive law enforcement activities are primarily three: first, the *chilling effect* that information access and data sharing by government might have on innocent behavior, second, the *slippery slope* that may result when powerful tools are used for increasingly pettier needs until finally we find ourselves smothered under a veil of constant surveillance, and, third, the potential for *abuse or misuse*.

86 Stone, *supra* note 85, at 215 ("In time of war – or, more precisely, in time of national crisis – we respond too harshly in our restrictions of civil liberties, and then, later regret our behavior.")
87 *See, e.g.,* Rosenzweig, *supra* note 71 at 684 ("The war on terror, uniquely, is one with no immediate foreseeable end. Thus, excessive intrusions may not be justified as emergency measures that will lapse upon the termination of hostilities.")
88 *See* Staff Statement, *supra* note 20. Ted Senator, DARPA, has referred to this as looking for in-liers, rather than out-liers. Center for Democracy and Technology and Heritage Foundation Roundtable on Data Mining (Dec. 2003).

A. THE CHILLING EFFECT

The chilling effect primarily involves the concern that potential lawful behavior, particularly constitutionally protected activity, would be inhibited due to the potential for a kind of post hoc surveillance (often referred to as "dataveillance") that is said by many to result from the increased sharing of information among currently discrete sources.[89]

"Potential knowledge is present power," and awareness that government may analyze activity is likely to alter behavior, "people act differently if they know their conduct *could be* observed."[90] The risk is that protected rights of expression, protest, association, and political participation may be affected by encouraging "conformity with a perceived norm, discouraging political dissent, or otherwise altering participation in political life."[91]

Maintaining individual privacy, however, is not synonymous with being able to commit or plan terrorist acts in secret without being discovered. Thus, chilling effects-based arguments against technologies or procedures that can potentially protect against catastrophic terrorist acts must show a real privacy impact on legitimate and innocent activity not just exhibit a fetish for absolute secrecy premised on vague referrals to potentially inhibited acts.

The Supreme Court requires that chilling-effects based challenges present more than allegations of a subjective chill; it requires that such challenges show both actual harm and a significant effect on protected activities not outweighed by legitimate government interest. Thus, in *Laird v. Tatum*,[92] the Court wrote:

---

89 Roger Clark, *Information Technology and Dataveillance*, 31 COMM. OF THE ACM 498-512 (1988) (coining the term "dataveillance" to describe how database stores of personal information have facilitated new surveillance practices); see *also* Solove, *supra* note 63, at 1084 (government access to digital dossiers can chill activities).

90 Safeguarding Privacy in the Fight Against Terrorism, The Report of the [Department of Defense] Technology and Privacy Advisory Committee at 35 (March 2004) *available at* http://www.sainc.com/tapac/finalReport.htm [hereinafter, TAPAC Report].

91 *Id.* at 35-36.

92 408 U.S. 1 (1972).

> In none of these cases, however, did the chilling effect arise merely from the individual's knowledge that a governmental agency was engaged in certain activities or from the individual's concomitant fear that, armed with the fruits of those activities, the agency might in the future take some other and additional action detrimental to that individual.[93]

Although the Court went on to note that it was not opining on the "propriety or desirability, from a policy standpoint, of the challenged activities" but merely its adjudicability,[94] it nevertheless seems appropriate for the policy debate likewise to require articulation or identification of some specific harm not outweighed by the compelling government interest. A vague claim of enforced conformity ought not in itself, *ipso facto*, win the argument.

Further, the mere existence of a chilling effect is not alone sufficient to hold governmental action unconstitutional:

> [T]he existence of a "chilling effect," … has never been considered a sufficient basis, in and of itself, for prohibiting state action. Where [the state action] does not directly abridge free speech, but – while regulating a subject within the State's power – tends to have the incident effect of inhibiting First Amendment rights, it is well settled that the [state action] can be upheld if the effect on speech is minor in relation to the need for control of the conduct and the lack of alternative means for doing so.[95]

Thus, chilling effects arguments against the use of technology should require determining *confidence intervals* – that is, the acceptable error rate – for a particular application in a particular use (i.e., its *reasonableness*). In the context of information processing for preemptive law enforcement, the confidence interval is the net result of false positives and false negatives, each adjusted for its related consequence and

93 *Laird,* 408 U.S. at 11.
94 *Id.* at 15.
95 Younger v. Harris, 401 U.S. 37, 51 (1971).

resource consumption.[96] To analogize to the Court's analysis, the adjusted false positive rate equates to the potential for actual harm to the individual and the adjusted false negative rate equates to the government interest, in this case, security.

Thus, determining confidence intervals for policy purposes can be viewed as a function of these two competing relationships – the number of false positives (innocents identified) adjusted by the severity of the consequences to the individual on the one hand and the number of the false negatives (terrorists not identified) adjusted by the consequences to security on the other.[97] If the consequences of a false positive are relatively low, for example, a bag search at the airport, and the consequences of a false negative are high, for example, the plane crashes into the Pentagon, the acceptable confidence interval for policy purposes should be adjusted (either technically or by procedures) to bias towards false positives and reduce false negatives. If, on the other hand, the consequences to the individual from a false positive are severe, for example incarceration, and the consequences of false negatives are slight, for example, a parking ticket scoff-law slips through, then the confidence interval should be adjusted (either technically or by policy) to reduce false positives at the risk of increasing false negatives.

This is not to suggest that there is some perfect correlation to be calculated among relative risks (which risks cannot be precisely quantified) but rather to suggest that when it comes to setting policy, recognizing that appropriate controls for a particular use will depend on the totality of the circumstance at the point and time of use – including (as discussed below) the scope and method of inquiry, the sensitivity of data, and the particular security interest or threat as well as the nature of the privacy intrusion – and cannot be rigidly proscribed or even anticipated. Thus, a perfect system design would incorporate flexibility in both its policy and technical controls to allow for changes in circumstances at the point of

96 For a discussion of the relationship between Type 1 errors (false positives, that is, innocents falsely identified as suspicious) and Type 2 errors (false negatives, that is, terrorists not identified) in the context of the 'war on terrorism', see Rosenzweig, *supra* note 71, at 677-683.

97 The consequences to security also include the costs associated with the misallocation of resources resulting from having to investigate false positives.

use, and its reasonableness would be judged on its use in such circumstances.

B. THE SLIPPERY SLOPE

The slippery slope argument[98] is that measures that might be adopted now for perfectly legitimate national security concerns might eventually be used in the ordinary course of law enforcement to investigate and apprehend lesser law breakers resulting in extraordinary procedures developed to counter a specific threat becoming the norm – in this case leading to a permanent and complete surveillance society (a world in which Michael Froomkin notes "it should be possible to achieve perfect law enforcement"[99]).

This fear is particularly relevant when one recognizes that there will always be an insatiable need for more security[100] and there will always exist a bureaucratic imperative for additional control.[101] There is also the practical consequence of making tools available – they will be used. For the law enforcement professional seeking to accomplish their mission we could expect no less than that they try to take advantage of every tool or opportunity that is available for each and every task that they are responsible for.[102] When these three factors – the need for more security, the imperial bureaucratic drive, and the practical availability of tools – are combined, the threat of the slippery slope is real and potentially significant.

Structural implementation options can help ameliorate these concerns. For example, the data analysis (intelligence)

98 *See generally* Eugene Volokh, *The Mechanisms of the Slippery Slope*, 116 HARV. L. REV. 1026 (2003).
99 Michael Froomkin, *The Death of Privacy?* 52 STAN. L. REV. 1461, 1471 (2000). *See also infra* note 138.
100 One can never be completely safe thus there is always more that could be done.
101 *See generally* Matthew Holden, Jr., *'Imperialism' in Bureaucracy*, 70 AM. POL. SCI. REV. 943 (December 1966).
102 In a similar vein, among the criticisms of the USA PATRIOT Act, Pub. L. No. 107-52, 115 Stat. 272 (2001), is that although it was passed as an anti-terrorism measure it has been used for non-terrorism related purposes. *See, e.g.,* Bryan Bender, *AG Touts Patriot Act; Opponents Unconvinced*, THE BOSTON GLOBE, July 14, 2004 ("[some Democrats in Congress] expressed concern that many of the crimes that have been uncovered via the new powers were not associated with terrorism").

function could be operationally separated from the law enforcement function as the Markle Taskforce has suggested.[103] The Gilmore Commission has recommended that the Terrorist Threat Integration Center be spun off as an independent agency to coordinate the domestic intelligence function,[104] and I (and others) have argued that a separate agency with a narrow charter to process intelligence for domestic security, no independent law enforcement powers, and subject to strict oversight should be considered.[105] While these organizational structures do not eliminate concern they can help. Further, technical architectures to counter the slippery slope also exist. A distributed architecture with local responsibility and accountability for data and access, together with strong credential and audit functions to track usage,[106] can provide protection from a centralized expansion of power or use.[107]

C. ABUSE AND MISUSE

Information systems are also open to abuse or misuse. There are many examples of such misuse – from IRS agents looking up their neighbor's tax returns[108] to law enforcement officials sharing information with criminal suspects.[109] Even examples of institutionalized abuse, such as the FBI

103 First Markle Report, *supra* note 21 at 2, 22-24

104 ADVISORY PANEL TO ASSESS DOMESTIC RESPONSE CAPABILITIES FOR TERRORISM INVOLVING WEAPONS OF MASS DESTRUCTION [the "Gilmore Commission"] FIFTH ANNUAL REPORT, Dec. 15, 2003, *available at* http://www.rand.org/nsrd/terrpanel.

105 *See, e.g.*, Chloe Albanesius, *Officials Defend Idea of Data Mining; Experts Weigh Options*, NAT'L J.'S TECH. DAILY (Dec. 2, 2003) (quoting Kim Taipale, "perhaps there is a 'need for a specific intelligence agency to go after terrorists' with a limited charter"). Compare, however, the Intelligence Reform and Prevention of Terror Act of 2004, approved by the House on Dec. 7, 2004 and the Senate on Dec. 8, 2004, which centralizes certain intelligence functions under a new Director of National Intelligence and waters down certain privacy and civil liberties oversight provisions that were in the earlier Senate version of the bill, S.2845.

106 *See, e.g.*, the Second Markle Report, *supra* note 23, at 1 (explaining need for a distributed network and arguing "against a centralized … system.").

107 Taipale, *supra* note 13, at 42-44.

108 *See* WILLIAM V. ROTH, JR. & WILLIAM H. NIXON, THE POWER TO DESTROY (1999).

109 *See, e.g.*, Press Release, U.S. Department of Justice, *FBI Legal Technician Pleads Guilty To Unlawfully Accessing The FBI's Computer System*, (Feb. 26, 2004) *available* at http://www.usdoj.gov:80/opa/pr/2004/February/04_crm_120.htm.

COINTELPRO, are recent enough to evoke concern.[110] For purposes of policy and technical design, however, the substance of the concerns need not be resolved – that is, we do not need to debate whether, for example, the government or its employees should be trusted to do what is right and not abuse its citizens. Instead, organizational structures, procedures, and technical features that function together to limit the potential for abuse can (and should) be designed and implemented to address these concerns.

Often neglected in this part of the debate is acknowledgment that the same characteristics of these technologies that give rise to some of the privacy concerns in the first place – the existence of "electronic footprints" in dataspace – also provide opportunities for resolution or mitigation – that is, these systems can be turned on themselves to "watch the watchers." Immutable logging together with strong credentialing and audit can provide significant deterrent to abuse making "abuse difficult to achieve and easy to uncover" by providing secure access control and tamper-resistant evidence of where data goes and who has had access to it.[111]

Additionally, real-time automated monitoring of system usage and post usage analysis and review, together with oversight of systems logs, can provide significant checks on both abuse and misuse.[112] Organizational structures to ensure such results should also be devised as part of systems implementations. Thus, for example, determining whether log files are to be kept locally (and, if so, under whose authority, for example, by the technical systems administrators, or the agency's inspector general, general counsel, or privacy officers, etc.) or externally by oversight bodies is not just a technical question but also one with substantive policy implications.

110 *See* COINTELPRO (Cathy Perkus, ed. 1976). COINTELPRO is an acronym for a series of FBI counterintelligence programs between 1956-1971 through which the FBI carried out domestic intelligence activities against political dissidents.

111 Rosenzweig, *supra* note 13, at 196-197.

112 As a technical matter, it is in these kinds of monitoring activity that automated analysis has shown the most success. *See* Taipale, *supra* note 13, at n.312.

D. JOSEPH K. AND THE SEPARATION OF SELF

It may well be that existing metaphors[113] and doctrines based on outdated notions of defining the relationship between an individual and their 'personal' information based on place or expectation are inadequate to address compelling new challenges brought by emerging technology to civil liberties.[114] Dan Solove[115] has suggested that a more appropriate metaphor for the problem of dataveillance[116] than Orwell's Big Brother[117] is Kafka's *The Trial*.[118] The concern is of a "more thoughtless process of bureaucratic indifference, arbitrary error, and dehumanization, a world where people feel powerless and vulnerable, without meaningful form of participation in the collection and use of their information" rather than the more

113 *See* GEORGE LAKOFF & MARK JOHNSON, METAPHORS WE LIVE BY 3-6 (2003) discussing how metaphors not only affect how we communicate but actually structure our perceptions and understandings from the outset. To paraphrase Marshall McLuhan, the "metaphor is the message." *See* MARSHALL MCLUHAN, UNDERSTANDING MEDIA: THE EXTENSIONS OF MAN 23-35 (1964) (McLuhan's iconic phrase "the medium is the message" suggests that the form of mediation itself gives substantive meaning to the mediated communication independent of content. So too, then, the metaphor.).
Choice of metaphor has suasive power because metaphor brings to a new subject an expectation imbued with all the old constraints and formal bounds that attend that which is being used as metaphor without requiring rigorous independent justification in the new case. *See* ANTHONY WILDEN, THE RULES ARE NO GAME 196-221 (1987), and ROMAN JACOBSON AND MORRIS HALLE, FUNDAMENTALS OF LANGUAGE 90-96 (Reprint second edition 2002, 1956). Metaphor, particularly in legal analysis, can presuppose the outcome, that is, by saying that *this* is metaphorically *that*, old legal doctrines can be applied to new situations without regard to differences in circumstance. In argument by analogy, the victory often goes to those who get the audience (or court) to accept their proffered metaphor. *See* K. A. Taipale, *Free Speech, Semiosis, and Cyberspace*, Center for Advanced Studies in Sci. & Tech. Pol'y Comment Draft at 6 (Jan. 2003) ("Part II. Metaphors: Is Cyberspace a Place or Social Condition?") *available at* http://www.taipale.org/papers/CyberSemiosis.pdf.
114 *See generally* DANIEL J. SOLOVE AND MARC ROTENBERG, INFORMATION PRIVACY LAW 275-322 (2003).
115 Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1398 (2001).
116 *See* note 89 *supra.*
117 GEORGE ORWELL, NINETEEN EIGHTY-FOUR (1949) depicts a totalitarian society of the future, ruled by an omnipotent dictator called Big Brother. In this society, called Oceania, people's thoughts and actions are continuously monitored. The term *Big Brother* has subsequently been used to refer to any ruler or government that invades the privacy of its citizens.
118 FRANZ KAFKA, THE TRIAL: A NEW TRANSLATION BASED ON THE RESTORED TEXT (Brion Mitchell, tr. 1999) (Joseph K. is arrested, tried and executed for an unspecified crime).

traditional concern of secrecy or surveillance.[119] In suggesting Kafka's *The Trial* as metaphor for the inchoate sense of lost control that comes from unknown uses of personal information in vast technological systems, Solove has illustrated an interesting divide in the theoretical underpinnings of information privacy premised on "control" of private information and distinct from the issues raised by traditional concerns of surveillance.[120]

Michael Froomkin has stated that "information privacy [is used] as shorthand for the ability to control the acquisition or release of information about oneself."[121] Jeffrey Rosen has written that a "central value" of such control is to protect individuals "from being misidentified and judged out of context in a world of short attention spans, a world in which information can easily be confused with knowledge."[122] And, Paul Schwartz has concluded that this process can lead to the "autonomy trap ... a reduced sense of the possible."[123]

The underlying concern that emerges here seems to be not so much that government will observe individual behavior (even after the fact) (i.e., Big Brother) but that it will come to the wrong conclusion with subsequent unpleasant consequence to the individual (i.e., *The Trial*) – that is, the fear is that data relating to an individual will be mismanaged or misinterpreted with real-world consequences to that individual. In another era, this might have been expressed as "do not fold, spindle or mutilate me (or my data)."[124] In *The Trial*, when the examining

119 Solove, *supra* note 115, at 1398.
120 *See also* Julie Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373 (2000).
121 Michael Froomkin, *supra* note 99, at 1463.
122 JEFFREY ROSEN, THE UNWANTED GAZE: THE DESTRUCTION OF PRIVACY IN AMERICA 8 (2000).
123 Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815, 825 (2000).
124 The phrase "do not fold, spindle or mutilate," which has almost disappeared from popular usage but was a cultural icon in the 1960s, was coined to prevent people from disabling the process of feeding machine readable computer punch cards into information systems. The phrase itself, as well as the punch cards to which it referred, became symbols of the computer, of alienation, and of anxiety about technology generally. *See, e.g.,* Steven Luber, Smithsonian Institute, *"Do not fold, spindle or mutilate": A cultural history of the punch card* (1991) at http://ccat.sas.upenn.edu/slubar/fsm.html.

magistrate consults his notes and asks Joseph K. (the protagonist and a bank clerk) whether he is a house painter, K. rises up to address the proceedings arguing that the very fact that such a question could be asked undermines the legitimacy of the proceedings.[125] A more light-hearted but still illustrative example of this is the now quasi-famous "My TiVo thinks I'm Gay" article[126] in which the subject finds that because he recorded a Steve Reeves gladiator movie his TiVo begins suggesting that he might like shows with a gay theme.[127]

A fundamental issue, as yet not fully resolved to everyone's satisfaction in the context of emerging technologies, is whether data *about* an individual (whether disclosed by that individual or otherwise obtained) should "belong" to that individual in any kind of sense that would invoke legal mechanisms of ongoing control – i.e., some notion of property[128] – or perhaps even a renewal of "expectations" of privacy for secondary uses[129] – after it shared or otherwise becomes known.

---

125 Kafka, *supra* note 118, at 44-47. ("Your question, your honor, about me being a house painter – and you weren't really asking at all, you were telling me outright – is characteristic of the way these entire proceedings against me are being conducted." *Id.* at 45.)

126 Jeffrey Zaslow, *If TiVo Thinks You are Gay, Here's How to Set it Straight*, WALL ST. J., Nov. 26, 2002 at 1.

127 TiVo is a network enabled digital recording service through which subscribers can digitally record television programming from multiple delivery sources (satellite, cable, broadcast, etc.). Among the TiVo features is one that suggests additional programming that the user might be interested in based on an analysis of past recordings by that individual. *Cf.* the amazon.com service described in note 202 *infra*.

128 *See*, e.g., Pamela Samuelson, *Privacy as Intellectual Property?* 52 STAN. L. REV. 1125, 1127 (2000) ("some American commentators have proposed that the law should grant individuals a property right in their personal data"); Jessica Littman, *Information Privacy/Information Property*, 52 STAN. L. REV. 1283, 1287 (2000) ("People should own information about themselves and, as owners of property, should be entitled to control what is done with it. [citing Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193, 1246-94 (1998)]. This essay explores that proposal."); *cf.* Julie E. Cohen, *supra* note 120, at 1377-1391 ("Conventional understandings of ownership, liberty, and expression do not easily stretch to accommodate informational privacy rights." *Id.* at 1375); *see also* Lawrence Lessig, *Privacy as Property*, 69 SOC. RES. 247, 247 and n.1 (2002) ("In my view, we would better support privacy within American society if we spoke of privacy as a kind of property. Property talk, in other words, would strengthen the rhetorical force behind privacy").

129 *Cf., e.g.*, the "Fair Information Practices" (as first set forth in U.S. Department of Health, Education and Welfare, *Records, Computers and the Rights of Citizens* (1973) [hereinafter *HEW Report*]; *see also OECD*

Although the general legal rule is well established that the Fourth Amendment does not prohibit the government from obtaining information that was voluntarily given to a third party and then conveyed by that party to government authorities because there can be no reasonable "expectation of privacy" for such already shared information[130] this issue continues to be subject to ongoing legal, policy and philosophical debate.[131] In particular, the question has been raised whether this 'third party rule' continues to be appropriate in the Information Age in which vast amounts of personal information is maintained by third parties in private sector databases, and

---

Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980) [hereinafter OECD *Guidelines*]; Information Infrastructure Task Force, Information Policy Committee, Privacy Working Group, *Privacy and the National Information Infrastructure: Principles for Providing and Using Personal Information* (1995) [hereinafter *IITF Report*]; U.S. Dept. of Commerce, *Privacy and the NII: Safeguarding Telecommunications-Related Personal Information* (1995); *The European Union Directive on the Protection of Personal Data* (1995) [hereinafter *EU Directive*];) that explicitly state that the corollary to identifying the purposes for data collection (i.e., the notice requirement) is that the data not be used for other or subsequent purposes without the data subject's consent. *See HEW Report* at 61-62; *OECD Guidelines*, Use Limitation Principle and para. 10; *IITF Report* § II.D; *EU Directive* arts. 6-7.

130 United States v. Miller, 425 U.S. 435, 441-443 (1976) (financial records); Smith v. Maryland, 442 U.S. 735, 743 (1979) (pen register).

131 In particular, the notion whether privacy should include a legal right to ongoing control of information about oneself after others know it, that is, to manage one's own reputation, is controverted. *Compare*, e.g., Jeffrey Rosen, *supra* note 122, at 8 (where Rosen argues that a "central value of privacy" is to protect individuals "from being misidentified and judged out of context in a world of short attention spans, a world in which information can easily be confused with knowledge") *and* Andrew E. Taslitz, *The Fourth Amendment in the Twenty-First Century: Technology, Privacy, and Human Emotions*, 65 L. & CONTEMP. PROBS. 125, 128 (2002) ("we want to choose the masks that we show to others") *with* Richard A. Posner, THE ECONOMICS OF JUSTICE 232-42 (1983) (arguing from an economic perspective that the individual want for privacy stems from a desire "to manipulate the world . . . by selective disclosure of facts . . . [in order] to mislead those with whom [the individual] transacts" and is therefore economically and socially inefficient.) *See also* Eugene Volokh, *Freedom of Speech and Information Privacy: The troubling implications of a right to stop people from speaking about you.* 52 STAN. L. REV. 1049 (2000); *cf.* the references in note 128 *supra.*

Whether "public" information is entitled to expectations of privacy, *see* Helen Nissenbaum, *Protecting Privacy in an Information Age: The Problem of Privacy in Public*, 17 LAW & PHIL. 559 (1998); Taslitz, *supra*, at text accompanying n.289 ("We do not shed all privacy expectations simply because we walk on a public street, or enter a classroom, or attend a ball game.")

the very nature of the medium requires that data be shared with, maintained by, or exposed to such third parties.[132] This question does not need to be resolved here.

Instead, to pick up on Solove's *Trial* metaphor, the concern as it relates to individual autonomy seems not so much about control of data but, rather, the consequences (or potential consequences) of error. Thus, to some extent this Kafkaesque concern is ameliorable through the same mechanisms as the chilling effect argument – that is, there is a direct relationship between the reasonableness of the action to the efficacy or confidence interval (error rate) of the process and its error correction procedures. It is, after all, the need for error correction through due process that is the underlying theme of *The Trial*, not secrecy of data.

According to Theodore Ziolkowski,[133] Kafka's subtle critique of the clash between traditional and modern law[134] in *The Trial* is exemplified by a system in which accusation is the same as guilt and leads to execution without the ability of the accused to get a hearing in a higher court:

> [A] system in which the preliminary investigation has displaced the other stages of the procedure with its guaranteed protection for the rights of the individual. The preliminary investigation, in turn, goes to the extreme in ignoring the objective facts of the case and

132 *See* Solove, *supra* note 63, at 1085-1086:

> The Court, however, has held that there is no reasonable expectation of privacy in records maintained by third parties. … Thus, there is a profoundly inadequate legal response to the emerging problem of government access to aggregations of data, "digital dossiers" [held by third parties] that are increasingly becoming digital biographies.

Solove goes on to conclude that "A new architecture of power must be constructed, one that effectively regulates the government's collection and use of third party records." *Id.* at 1167. *See also* discussion of 'scope of access' *infra.*

133 THEODORE ZIOLKOWSKI, THE MIRROR OF JUSTICE 233-240 (1997).

134 According to Ziolkowski, *id.* at 236-238, Kafka is in part motivated by the perceived clash between traditional notions of law – based on guilt – and modern notions – based on violation of rules – as exemplified in differences between the German and Austrian penal codes with which Kafka was familiar. *Cf.* STEPHEN M. FELDMAN, AMERICAN LEGAL THOUGHT FROM PREMODERNISM TO POSTMODERNISM (2000) (tracing the evolution of American legal thought).

> focusing on the guilt of the accused. ... [T]he system moves from inquiry to execution without defense, trial, or notification.[135]

The lesson that I draw from Solove's metaphor is not that a fetish for absolute secrecy is needed to address these concerns but rather an insistence on procedural protections – that is, requiring a system that is designed with organizational, procedural, and technical features that allow due process mechanisms to function, and that recognizes the potential for error and provides mechanisms for its correction.[136]

Thus, technological systems should conform to existing (or evolving) notions of due process and technical features and implementations should be designed to support those procedures.[137] As additional protection, information processing technologies should be used only as investigative tools – that is, to allocate law enforcement resources – not for evidentiary purposes.[138] Further, the result of automated processing should

---

135 Ziolkowski, *supra* note 133, at 239-240.

136 The importance of error correction procedures as fundamental to privacy protection is also highlighted by its explicit incorporation in the Fair Information Practices in the *HEW Report, supra* note 129, at 41, 59, 63, *OECD Guidelines*, *supra* note 129, at para. 13; and *EU Directive*, *supra* note 129, at art. 12. *See also* Fair Credit Reporting Act ("FCRA") §§ 609-11, 15 U.S.C. §§ 1681g-1681i (providing for consumer access to, and the right to correct inaccuracies in, consumer credit reports).

137 With respect to system design, note that Solove calls for a new "architecture of power" to address inadequacies in current doctrine. Solove, *supra* note 63, at 1151-1167. This Article argues that technical system architecture itself can and should be developed concurrently with the development of any such policy or legal structure. *See also* Neal Kumar Katyal, *Architecture as Crime Control*, 111 YALE L.J. 1039, 1047 (2002) [hereinafter, Katyal, *Architecture*] (discussing the use of physical architecture – structural and space design – as an effective alternative form of crime control; design mechanisms discussed include: (1) creating opportunities for surveillance, (2) instilling a sense of territoriality, (3) building community and avoiding isolation, and (4) protecting targets); Neal Kumar Katyal, *Digital Architecture as Crime Control*, 112 YALE L.J. 2261 (2003) [hereinafter, Katyal, *Digital Architecture*] (applying these four principles of realspace architecture design to the problem of security in cyberspace); K. A. Taipale, *Internet and Computer Crime: System Architecture as Crime Control*, Center for Advanced Studies (Feb. 2003).

138 *See* Taipale, *supra* note 13, at n.56. *Cf.* the increasing use of automated traffic-surveillance camera systems to issue traffic tickets without human review, *see, e.g.,* William Matthews, *Battle Lines Form over Red Light*

be subject to human review before triggering adverse individual consequences.[139] Finally, large-scale technical systems or applications with significant privacy implications should be subject to programmatic authorization prior to

---

Cameras, FCW.com, Sep. 3, 2001, at http://www.fcw.com/geb/articles/2001/sep/geb-comm2-09-01.asp.
Note also the potential outcome if automated analysis of ubiquitous data becomes the norm and is taken to its logical extreme:

> Ultimately, if data is collected on everyone's location and on all transactions, it should be possible to achieve perfect law enforcement, a world in which no transgression goes undetected and, perhaps, unpunished. At that point, the assumptions of imperfect detection, the need for deterrence, and the reliance on police and prosecutorial discretion on which our legal system is based will come under sever strain.

Froomkin, *supra* note 99, at 1470-1471.
It is beyond the scope of this Article to address this issue in any depth, however, I have argued elsewhere, *see, e.g.* K. A. Taipale, *Technology, Security and Privacy: Rethinking the Problem Statement*, Presentation at In Search of J. Doe: Can Anonymity Survive in Post-911 Society Conference, Woodrow Wilson International Center for Scholars, at slide 21 (May 4, 2004) *available at* http://www.taipale.org/presentations/CAS-WWICS.htm, that such a system, based on perfect law enforcement through ubiquitous control technologies, is more akin to a Pigovian tax system, *see generally* ARTHUR PIGOU, THE ECONOMICS OF WELFARE (2002, 1920), for social control than a traditional Beccarian criminal justice model, *see generally*, CESARE BECCARIA: ON CRIME AND PUNISHMENTS (Henry Paolucci trans. 1963, 1764), and will require rethinking what is criminalized.
By analogy, for purposes of this Article, privacy – if we define it as autonomy – can also be protected by changing the consequences of disclosure (i.e., by lessening its affect on autonomy). *See generally* DAVID BRIN, THE TRANSPARENT SOCIETY (1998).
On the general issue of social control through systems design, *see generally* JEREMY BENTHAM, PANOPTICON (1971); JACQUES ELLUL, THE TECHNOLOGICAL SOCIETY (1964); DAVID BURNHAM, THE RISE OF THE COMPUTER STATE (1983); MICHEL FOUCAULT, DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON (1995); OSCAR H. GANDY, THE PANOPTICON SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION (1993); GARY T. MARX, UNDERCOVER: POLICE SURVEILLANCE IN AMERICA (1988); *see also* Katyal, *Architecture,* s*upra* note 137; Katyal, *Digital Architecture*, *supra* note 137; Taipale, *supra* note 137.
139 The European Union proposes a right to have human checking of adverse computer, generated results. *See EU Directive, supra* note 129, at art. 15:

> Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him ….

implementation,[140] oversight during use, and judicial review in accordance with existing (or evolving) due process doctrines and practices after the fact of use.[141]

The purpose of this article is not to minimize these privacy concerns, rather it is to illustrate where they have applicability in practice and where appropriate procedures or technological development strategies can help ameliorate concerns by allowing familiar mechanisms to operate.

VII. THE TECHNOLOGIES

The kinds of technologies with which this article is concerned can be classified broadly as three types:

- technologies of identification,
- technologies of data aggregation and automated analysis,[142] and
- technologies of collection.

The purpose of this article is not to detail technical developments in each of these areas in great depth, but rather to identify certain thematic characteristics that illustrate where technical solutions and system design can provide intervention points in the application of these technologies in order for familiar due process procedures and related mechanisms to function. Thus, rather than a discussion of particular technologies this section describes their functional application and their relationship to security needs.[143] Throughout this

140 For example, as recommended by the TAPAC Report, *supra* note 90, at x-xii, for data mining applications.
141 *See generally* Rosenzweig, *supra* note 13 (setting out a proposed legal and procedural framework for implementation of TIA).
142 Including "data mining". *See generally* Taipale, *supra* note 13.
143 Aligning information technology requirements and capabilities with *business process* needs is core to most current models of *enterprise architecture* in the private sector. Enterprise architecture recognizes that most system design problems are not technology problems but business process problems and seeks to align information and technical architecture to support business needs. *See generally* MELISSA COOK, BUILDING ENTERPRISE INFORMATION ARCHITECTURE: REENGINEERING INFORMATION SYSTEMS (1996); Federal Enterprise Architecture Management Office, *What is Federal Enterprise Architecture (FEA)* at http://www.feapmo.gov/fea.asp.

article, but particularly in this section, I use the terms *system* and *security* broadly and generically. I use system to mean any bounded system – i.e., national borders, air transportation, a particular physical location, or a computer network – within which one wants to provide security. And, I use security to mean any effort to ensure that 'users' of a system comport to rules for behavior in that system.

Before discussing the functional aspects of identification systems it is important to reiterate that the use of any intrusive technology in any particular application has to be measured not only against its privacy impact but also its efficacy for meeting security needs and countering designated threats. If a particular technology or application is not effective at improving security it should not be considered for use in the first place.[144] Thus, for example, very high false positive rates for any screening system are not only intolerable because of their impact on privacy but are not useful for security as they misallocate or waste security resources. Also, the need to collect, maintain or process any particular type of data within a specific security application needs to be weighed against its salience for that particular security need.[145]

Both the security and privacy effects from any identification system are derived from the security and privacy features and design of the overall system in which they are to be

---

It is beyond the scope of this Article to fully explore enterprise architecture. For purposes of this Article it is sufficient to recognize that information management and technical systems architectures (that is, systems design) needs to be developed to support the relevant business processes, in this case, both security and privacy. *See also,* K. A. Taipale, Presentation at Counterterrorism Technology and Privacy Conference, McCormick Tribune Foundation Cantigny Conference Series, ABA Standing Committee on Law and National Security (June 24-25, 2004) *available at* http://www.taipale.org/presentations/Cantigny-062504.pdf.

144 *But see* note 161 *infra* and accompanying text (discussing collateral benefits from "security theater").

145 Admittedly, determining salience for certain types of data prior to analysis may not be possible (i.e., without attempting to "connect the dots" one cannot know which dots may or may not be relevant), however, understanding the relationship of the data need for the particular security application itself is possible. For example, in any given security context, what is the purpose for requiring "identification" – to prove that the person has an ID, to prove that the person is who they say they are, or to prove that the person is authorized to do something? Each requires a different information management procedure and each has different implications for security and privacy.

used and are not inherent in the identification technologies themselves. Obviously, different systems (and different threats) require different levels of security and will require different trade-offs to be made between privacy, security and functionality,[146] however, for analytic purposes in this section, we discuss functional aspects of security systems generally.

A. TECHNOLOGIES OF IDENTIFICATION

Identification technologies or systems serve to *authenticate* data attribution – that is, they provide confidence that a particular piece of data (an attribute) or collection of data (an identity) correlates with a specified entity (an individual or other object).[147]

Authentication generally serves as the first step in one or both of two kinds of security applications or strategies – *authorization* and/or *accountability*.[148] Authorization (or permission) is the process of deciding what an identified individual is permitted (or not permitted) to do within a system (including whether they are allowed access in the first place). For example, an individual may be authorized to enter a secure zone, may be denied access to board a plane, or may be given access to a computer system but constrained from accessing certain services or information. Accountability, on the other hand, is the process of associating a consequence to the individual for any actions that they may take within the system, for example, by recording identifying information prior to entry into a system, or by monitoring, recording or logging activity within the system, to allow for subsequent tracking or sanction. Both authorization and accountability serve to ensure that rules governing behavior within a system are obeyed.[149]

---

146 *See generally* Michael Froomkin, *The Uneasy Case for National ID Cards*, [YISP CyberCrime 2004] (2004) (discussing the trade-offs involved in implementing a national ID card); *But cf.* Richard Sobel, *The Demeaning of Identity and Personhood in National Identification Systems*, 15 HARV. J. LAW & TECH. 319 (2002).

147 *See* NATIONAL RESEARCH COUNCIL, WHO GOES THERE? [hereinafter NRC] 16-32 (Stephen T. Kent and Lynette I. Millett, eds., 2003). Much of this section on identification follows the analysis set forth in this reference.

148 *Id.* at 337-38.

149 Unfortunately, accountability strategies are generally ineffective against users unconstrained by after-the-fact punishment, for example, suicide

From a political point of view, authorization based on access control strategies are generally associated with totalitarian systems (i.e., the default state of these systems is that all users are suspect, the chosen receive permission) and accountability strategies are associated with freedom (i.e., the default state is that all users are presumed innocent, only those who have already done something wrong are sanctioned).[150] Unfortunately, accountability strategies are not very effective against suicidal attackers or situations with catastrophic outcomes, and, thus, give rise to the difficult policy choices facing a free society in taking a preemptive, rather than traditional reactive law enforcement, approach to terrorism.

In any identification system, there are generally three forms of authentication that can occur:

- *Entity authentication* is the process of establishing confidence that an *identifier*, for example, a name, number or symbol, refers to a specific *entity* (an individual, place or thing),[151]

- *Identity authentication* is the process of establishing confidence that an identifier refers to an *identity* (a collection of data related to an entity),[152] and

- *Attribute authentication* is the process of establishing confidence that an attribute (a property associated with an entity, for example, a

attackers without accountable patrons or other support infrastructure subject to sanction.

150 Dan Greer, Keynote Address at the Yale Information Society Project CyberCrime and Digital Law Enforcement Conference (Mar. 27, 2004).

151 NRC, *supra* note 147, at 19. Entity authentication usually happens in two phases; first, some identifier is selected (or offered), and second the identifier is authenticated. In computer security, entity authentication is generally referred to as "user authentication" and in biometrics it is generally referred to as "verification". *Id.*

152 NRC, *supra* note 147, at 19. Like with entity identification, first some identifier is selected (or offered), then authenticated, however, the identity authenticated is not necessarily linkable to a particular individual. For example, I may allow you to use my AOL account by sharing my password. You offer the password for access and it is authenticated as belonging to that account or identity.

> physical descriptor or a role, etc.) applies to a specific entity.[153]

Individuals (or other entities) may have multiple *identifiers*, even within the same systems, for example, name, social security number, driver's license number, etc. and may have one or more *aliases* for each identifier.[154] Note that the greater the *uniqueness* of any particular identifier, the greater the confidence that it applies to a particular individual or entity.[155]

In addition, individuals or entities may also have multiple *identities* – that is, multiple discrete sets of related data defining, for example, a particular role. Any particular individual might have identities relating to different roles, for example, as a family member, as a work professional, or as a community participant, each of which may or may not share attributes or identifiers.

Entity resolution is the technical process whereby different identifiers or different identities are resolved (attributed) to the same entity or individual usually through analysis of shared attributes. Technical methods for entity resolution of individuals – that is, confirming that multiple discrete sets of related data (i.e., different "identities") actually belong to the same individual – have achieved high success rates

153 NRC, *supra* note 147, at 20.

154 Aliases may include derivatives of a related identifier, for example, Robert, Bob or Bobby; can be related to a particular physical characteristic, for example, Shorty or Stretch; or can be unrelated, for example, Spike, Plubius or Lenin. Aliases can be adopted for social (for example, nicknames), nefarious (for example, criminal aliases) or autonomy-protecting (for example, a political *nom de plume* or *nom de guerre*) reasons. Aliases can be known to link to a particular identity or not.

155 But note that even a particular identity consisting of a set of attributes may apply to more than one individual. For example, the identity: George Bush, U.S. President, Yale graduate, Texas resident – applies to at least two different individuals. The corollary, of course, is that by combining independent variable attributes, entity resolution occurs – for example, 43 people fit the US President attribute, hundreds the George Bush attribute, thousands the Yale graduate attribute, and millions the Texas resident attribute – yet in combination, these attributes together give a high confidence that only one of two individuals out of a population of billions is "identified". *See* Taipale, *supra* note 13, at n.128, discussing the use of *ensemble classifiers* – that is, multiple independent models – to increase confidence intervals in pattern-matching applications.

and for some applications are a "solved problem."[156] However, entity resolution of places (or objects) has not been satisfactorily automated as yet.[157] Some form of entity resolution (or other data normalization) is generally required for automated analysis, particularly in systems based on anonymization and pseudonymization described below.[158]

Identity verification can be achieved through tokens (something you have), passwords (something you know), or a data match (something you are).[159] The highest level of confidence combines all three, for example, a token (ID card), requiring a password (PIN), and that contains a data match (for example, a biometric identifier).[160]

---

156 Jeff Jonas, Presentation at the Center for Strategic and International Studies, *Data-mining in the Private Sector* (July 23, 2003) (the "question of resolving identity – that is, ensuring that data all refer to a single unique individual – is a 'solved problem'.") cited in Rosenzweig, *supra* note 13, at n.14; *see also* Gang Wang, Hsinchun Chen & Homa Atabakhsh, Automatically *Detecting Deceptive Criminal Identities*, 47 COMM. OF THE ACM 71 (March 2004).

157 Resolving the many ways of referencing geographic location in text – for example, determining that '123 Main Street', 'the corner of Broadway and Main', and 'the location of the Fist Federal Bank' are identifiers all describing the same location – is required to bridge the gap between Geographic Information Systems (GIS) (in which data is tied to specific coordinates in space) and plain text data, which may refer to location in any manner. For a discussion of legal issues relating to GIS, *see* Jeremy Speich, *Comment: The Legal Implications of Geographical Information Systems (GIS)*, 11 ALB. L. J. SCI. & TECH. 359 (2001).

158 In order to do anonymous data matching (see discussion *infra*) some method of ensuring that variants in input (for example, different spellings of names) still yield outputs that can be matched. Either all related data needs to be resolved to one entity prior to processing (entity resolution) or some form of data normalization needs to be built into the process. Data normalization uses rules to reduce variants to one input (e.g., Robert, Bob, Bobby all reduced to ROBERT) or links the output from related inputs (for example, recognizing the hashes of Robert, Bob, or Bobby as matches); *see also* Dempsey, *supra* note 16, at 7-8 (discussing data standardization). Indeed, doing any automated processing on data requires some form of data normalization or transformation to account for errors on input or other 'dirty data'. *See* Taipale, *supra* note 13, at 27.

159 D. E. Raphael & J. R. Young, *Automated Personal Identification*, SRI International (1974); National Bureau of Standards, Evaluation Techniques for Human Identification, FIPSPUB-48 (Apr. 1977), cited in NRC, *supra* note 147, at 46.

160 Often called *three-factor authentication*. *See, e.g.*, James McGuire, *The Enterprise Authentication Game*, NEWSFACTOR NETWORK, January 13, 2003, at http://www.newsfactor.com/perl/story/20444.html.

Confidence in identification depends not only on the technologies of identification but on the integrity of the process of *enrollment* (the issuing and maintaining of tokens, passwords and the data to be matched), as well as the process of *verification* (confirming or verifying identity).[161] Even technologies of identification with very low error rates for matching (for example, certain *biometrics*) can be compromised if the enrolment process is corrupted or if the measurement process is fooled.[162]

Identification technologies can also be classified as *participatory*, where the person to be identified either cooperates or engages with the system knowingly, or *passive*, where the individual is not required to actively participate in the identification process. Examples of the former are the use of ID cards, fingerprint or iris scanners, and passwords, examples of the latter are face and gait recognition (and other so-called *recognition-at-a-distance* technologies), DNA sniffers, and the like. Passive identification can be either overt or surreptitious.[163] Each of these characteristics has obvious security and privacy implications.[164]

Authentication (that is, identification) in a security system is only the first step and does not provide security against a particular threat on its own. After identity is authenticated it must be used for some security purpose – either by authorizing the individual to do or not do something,[165] or by

161 *See, e.g.*, Second Markle Report, *supra* note 23, at app. A: Reliable Identification for Homeland Protection and Collateral Gains.
162 *See, e.g.*, John Leyden, *Gummi Bears Defeat Fingerprint Sensors*, THE REGISTER, May 16, 2002 *available at* http://www.theregister.co.uk/2002/05/16/gummi_bears_defeat_fingerprint_sensors/. Obviously, systems need to be designed to ensure that the appropriate biometric is being measured at verification (as well as to ensure the integrity of the data against which it is being matched).
163 *See* David E. Steinberg, *Making Sense of Sense-enhanced Searches*, 74 MINN. L. REV. 563, 569-574 (1990) (discussing issues involved in surreptitious or secret identification or search). *Cf. also* the American Bar Association Standards on Technologically-Assisted Surveillance §2-9.1 discussing relevant factors to consider in regulating surveillance, including whether a particular implementation is overt, that is, surveillance of which a reasonable person should be aware, *see also* §2-9.3, or covert.) *available at* http://www.abanet.org/crimjust/standards/taps_blk.html.
164 *Id.*; *see also*, NRC, *supra* note 147, at 55-79.
165 Another complexity in designing identification based systems for security is that a particular authorization may "belong" to an entity, an identifier, or

logging or tracking identifying data in some fashion to provide for later accountability. Thus, any identification system is only as good as the watch list or other criteria against which the authenticated identity is compared for authorization[166] or the deterrent effectiveness of the sanction for accountability.[167]

---

an identity – and in each case may vary by context. For example, an individual may be authorized to do something by virtue of their individual relationship (e.g., as spouse or parent), by virtue of a token (e.g., possession of bearer bonds, hall pass, bathroom key, etc.), or by virtue of their identity in context (e.g., the right of a police officer to carry a firearm while on duty or for a baggage handler to enter a secure area). That authorizations may relate to context creates potential weaknesses in access systems that do not distinguish, for example, whether an individual is on-duty or off-duty upon identification.

166 Problems with government "watch lists" in the war on terrorism are well documented. These problems include the difficulty of integrating multiple lists, *see, e.g.,* John Mintz, *DHS Blamed for Failure To Combine Watch Lists*, WASH. POST A02 (Oct. 2, 2004); Dibya Sarkar, *Inspector general finds watch list leadership lacking*, FCW.COM (Oct. 4, 2004); Department of Homeland Security Inspector General, *DHS Challenges In Consolidating Terrorist Watch List Information* OIG-04-31 (Aug. 2004); as well as the problems associated with using non-unique identifiers – i.e., common names – for screening purposes, *see, e.g.*, Sara Kehaulani Goo, *Hundreds Report Watch-List Trials, Some Ended Hassles at Airports by Making Slight Change to Name,* WASH. POST (Aug. 21, 2004, at A08) (also highlighting the ease with which such systems are defeated, for example, by using a middle initial). These problems are especially concerning because the use of watch lists is spreading. *See, e.g., Sept. 11 Commission Wants 'No-Fly' List for Trains, Ships*, FOXNEWS.COM, Sept. 08, 2004 *available at* http://www.foxnews.com/story/0,2933,131820,00.html.

167 Clearly, showing ID to prove that you have an ID to gain entrance (as is the case in many commercial buildings, for example) is more "security theater" as some security consultants have called it, than real security. BRUCE SCHNEIER, BEYOND FEAR 38-40 (2003). Nevertheless, while security theater is a clever phrase sure to garner press quotes, such measures – that is, security measures that provide a feeling of security regardless of whether they actually reduce risks or counter specific threats – can also serve a beneficial function if they maintain confidence in systems and allow for normal functioning.

Thus, policy makers must also consider whether making passengers *feel* safer is important for maintaining the viability of the economic, transportation or other systems regardless of whether it actually increases security against a specific threat. For analytical purposes, policy-makers need to take an expansive view of security. Security in this broad sense encompasses maintaining viability of economic, transportation or other systems and is not the same as mere physical security against a specific terrorist threat. Focusing only on the latter – physical security – is in my view myopic and best left for security consultants selling books. Policy analysts should consider security in its broader sense.

1. IDENTIFICATION SYSTEMS AND SECURITY

Identification based security is always somewhat vulnerable because of what is known as the *trusted systems* problem.[168] With few exceptions, "secure" systems need to be penetrated – under authorized circumstances by trusted people.[169] Unfortunately, there is inherently no way to prove trust, the best that any identification system can do is confirm not-yet-proven-untrustworthy status, i.e. confirm that a particular individual is not on a watch list for example.[170]

This essentially creates three classes of users of any system based only on identification, those confirmed as untrustworthy and denied access (and there may be false positives), those deemed not-untrustworthy who are in-fact trustworthy and are allowed access (good guys), and those deemed not-untrustworthy who are in-fact untrustworthy but have not yet been identified as such and may be mistakenly allowed access (false negatives).

Therefore, any system of identification needs to be part of a larger security system that recognizes, and compensates for, this problem. So, for example, a system for screening passengers (like CAPPS II or its successor, "Secure Flight")[171]

On the other hand, too much security theater can result in complacency and a false sense of security if such "feel good" measures are not also accompanied by real security strategies. Also, their cost – in terms of resource allocation or friction – needs to be considered in the context of their overall benefit. *See* Taipale, *Losing the War*, *supra* note 25.

168 *See generally* National Computer Security Center, A Guide to Understanding Discretionary Access Control in Trusted Systems, NCSC-TG-003-87 Library No. S-228, 576(Sep. 30, 1987), *available at* http://www.radium.ncsc.mil/tpep/library/rainbow/ NCSC-TG-003.html.

169 Schneier, *supra* note 167, at 181.

170 *See supra* note 166 (discussing problems with watch lists).

171 The Computer Assisted Passenger Pre-Screening ("CAPPS II") was a Transportation Safety Administration project designed to pre-screen passengers to assess threat levels to aviation security. For a detailed discussion of CAPPS II, *see* Taipale, *supra* note 13, at 37-39. The CAPPS II program has been "scrapped" but a new program is being developed. *See, e.g.*, Chris Stromh, *DHS scraps computer pre-screening system, starts over,* GOVEXEC.COM, July 15, 2004 at http://www.govexec.com/dailyfed/0704/071504c1.htm; Larry Greenemeie, CAPPS II Is Dead, Says Ridge, *But Door Is Open For CAPPS III*, INFORMATION WEEK, July 15, 2004 at http://www.informationweek.com/story/showArticle.jhtml?articleID=23901115. *But see* note 166 *supra* (discussing problems with watch lists for screening); and

should be combined with random searching of non-flagged passengers to provide layered security.[172]

Another general problem in security systems is balancing security with *usability* or *functionality*.[173] Authentication imposes friction or overhead on a system and can interfere with its usefulness. In the context of the 'war on terrorism' security systems based on authentication that impose too high a cost on functionality risk undermining the very system for which protection is sought.[174] Thus, for example, too high a burden in terms of physical intrusion or time spent in airport security screening lines can undermine the air transportation system. Inefficient port security can fail in two ways – terrorists can gain entrance or legitimate commerce can be impeded to the point that it interferes with trade. Denying access to immigrants or visa-applicants deprives the economy of needed talent. This issue is beyond the scope of this article except, however, as it relates to privacy concerns. To the extent that any security system imposes privacy costs on users out of proportion to the perceived threat, it risks undermining the confidence and support that is required from existing users for systems to function or for systems to attract new users – i.e., the capital and talent it needs for proper functioning or further development.

Laura Murphy & Barry Steinhardt, ACLU Comments to Department of Homeland Security on the "Passenger and Aviation Security Screening Records (Sep. 30, 2003) *available at* http://www.aclu.org/SafeandFree/SafeandFree.cfm?ID=13847&c=206. On September 21, 2004, TSA announced the filing and release of a Privacy Impact Statement and a Systems of Records Notice for the testing phase of "Secure Flight" the follow-on program to CAPPS II. *See* Press Release, U. S. Department of Homeland Security, Transportation Safety Administration (Sep. 21, 2004) *available at* http://www.tsa.gov/public/display?theme=44&content=09000519800cf2f3.

172 *See* Taipale, *supra* note 13, at n.285. Combining a screening system with random checks also protects against counter-programming (that is, adaptive behavior on the part of terrorists). *Id.* (discussing vulnerability of CAPPS II to "Carnival Booth" attacks).

173 NRC, *supra* note 147, at 80-103.

174 Taipale, *Losing the War*, *supra* note 25. Note that as a general rule, access control mechanisms impose a higher cost on functionality than accountability systems. This cost tends to increase as systems scale, thus, trade-offs between system security and functionality need to be considered especially if systems are to be widely deployed. *See* Geer, *supra* note150.

2. PRIVACY CONCERNS

Identification systems can either enhance or intrude upon privacy depending on their use and context.[175] Identification systems can enhance privacy when they are used to secure data or to protect identity, for example, by ensuring that an individual is indeed the authorized user of a credit card or a particular computer network, or is permitted access to certain information. Identification systems can also provide convenience, for example, by allowing personalized services to be delivered.[176]

On the other hand, identification systems can be intrusive of privacy and their use can be self-proliferating. Proliferation occurs when the prevalence of a security paradigm premised on fully mediated access becomes the norm.[177] For example, once ID checks are common for boarding airplanes or entering government buildings, they become acceptable (or required) for lesser uses – for example, prior to boarding trains or buses, or entering stores, etc.

Additionally, identification systems themselves tend to increase the collection of personal data, for example, by creating additional transaction records at the time and place of authentication, and may also expose personal information to additional disclosure at multiple points during the operation of the system or subsequently.[178] Availability of these transaction records may also allow for linkages and profiling, and the ability to create digital dossiers, not otherwise possible.[179]

Also, as noted above, the use of identification or authentication systems in conjunction with access control

175 NRC, *supra* note 147, at 55-56.
176 Consumers seem willing to trade personal information for personalized service as they increasingly perceive the value of such service, *see* Joshua Weinberger, *The Price of Personalization*, destinationCRM.com (July 28, 2004) *at* http://www.destinationcrm.com/articles/default.asp?ArticleID=4312 ("A new survey shows that consumers are willing to part with personal data in return for personalized service"). However, *cf.* Froomkin, *supra* note 99, at 1501-1505 (discussing the economics of *privacy myopia*, that is, the problems arising from individuals valuing privacy at its marginal cost and aggregators valuing it at its average cost.).
177 NRC, *supra* note 147, at 55.
178 NRC, *supra* note 147, at 56-57.
179 NRC, *supra* note 147, at 57; *see also* Solove, *supra* note 63, at 1084.

strategies may challenge traditional notions of freedom. In particular, access control strategies may impact on individual autonomy, including freedom of speech (denying access to information or communication systems),[180] freedom to travel or peaceably assemble (by denying access to particular modes of transport),[181] and freedom to petition the government (by denying access to government buildings or other resources).

Certain privacy impacts cannot be eliminated as they are inherent in the act of authentication, which requires the revelation and confirmation of some 'identifying' information to function, however, identification and authentication systems can be designed to minimize these privacy impacts and maximize security gains.[182] Further, identification should not be required where it does not provide a security gain. Thus, for example, a distinction should be drawn between systems or occasions when an identifier is required for security and situations where only authentication is required .[183]

In addition, even where identification or authentication strategies are appropriate, they should be designed so as to

180 *See, e.g.*, Matt Richtel, *Barring Web Use after Web Crime*, N. Y. TIMES, Jan. 21, 2003 at A:1 (discussing whether persons who commit a computer crime can be barred from use of the Internet after their release).

181 *See e.g.*, Gilmore v. Ashcroft, No. C02-No. C02-3444 SI (N.D. Cal.), challenging the requirement to show identification prior to boarding a commercial aircraft. On March 23, 2004 the district court dismissed plaintiff's complaint holding that the requirement for identification did not constitute a search for Fourth Amendment purposes, or, if it did, it was reasonable, and that such requirement did not infringe on plaintiff's right to travel, associate or petition government. *See* Nixon Peabody LP, Aviation Law Alert, April 2004, *available at* http://www.nixonpeabody.com/linked_media/publications/ ALA_04082004.pdf

182 Government should not build a massive identification surveillance system *unless* such features are built in. *See* K. A. Taipale, Statement at a Meeting of the Program on Law Enforcement and National Security in the Information Age (Oct. 29, 2004), *in* Press Release, Center for Advanced Studies in Science and Technology Policy (Oct. 29, 2004) *available at* http://releases.usnewswire.com/GetRelease.asp?id=39202 ("Congress should not rush to legislate a massive government identity surveillance system under the press of a politically expedient deadline without considering alternatives that can meet legitimate law enforcement and national security needs while still protecting privacy").

183 *See, e.g.*, Dan Farmer and Charles C. Mann, Surveillance Nation-Part Two, MIT TECH. REV. (May 2003) (describing how personal data on Malaysia's smart card chips – designed to replace driver's licenses – are stored in isolated files, each accessible only to authorized readers for that particular data).

neither require more personal information than is necessary for the particular security application (and even then in proportion to the threat) nor generate additional transaction records beyond what is required for the particular security purpose.[184]

Pseudonymization strategies, discussed *infra*, based on certificated authorizations from trusted third parties, selective disclosure of identifying information, and escrowed identity, can be designed to protect identity privacy but still meet legitimate law enforcement and security needs.

B. TECHNOLOGIES OF DATA AGGREGATION AND ANALYSIS

For an in-depth analysis of data aggregation and data analysis technologies, particularly data mining, see *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*.[185]

1. DATA AGGREGATION, DATA ANALYSIS, AND SECURITY

Recent reports by the U.S. Congress,[186] the National Research Council,[187] the Markle Foundation[188] and others have

184 *See* "The Relationship between Authentication and Identification," NRC, *supra* note 147, at 51-54. *Cf.* Hiibel v. Nevada, No. 03-5554 (S. Ct. June 21, 2004) (holding that requiring a suspect to disclose their name during *Terry* stop is permissible).

185 Taipale, *supra* note 13; see *also* Mary DeRosa, *Data Mining and Data Analysis for Counterterrorism*, Center for Strategic and International Studies (March 2004).

186 Joint Inquiry Report, *supra* note 22, at 6.

187 NATIONAL RESEARCH COUNCIL, MAKING THE NATION SAFER: THE ROLE OF SCIENCE AND TECHNOLOGY IN COUNTERING TERRORISM (Committee on Science and Technology for Countering Terrorism, ed., 2002), *available at* http://www.nap.edu/ html/stct/index.html:

> Currently one of [intelligence agencies'] significant problems is managing a flood of data that may be relevant to their efforts to track suspected terrorists and their activities. … There are well-known examples in which planned terrorist activities went undetected despite the fact that evidence was available to spot it – the relevant evidence was just one needle in a huge haystack.

188 First Markle Report, *supra* note 21.

highlighted that the amount of available data to be analyzed for domestic security purposes exceeds the capacity to analyze it. Further, these reports identify a failure to use information technology to effectively address this problem.[189] Among the recommendations are the increased use of data aggregation (information sharing) and automated analysis (in particular data mining) technologies.[190]

a) DATA AGGREGATION

Data itself can become more meaningful through aggregation or sharing.[191] First, data may be meaningless in any particular location but becomes increasingly useful as people perceive it to be useful within their local context. Second, data may become more valuable in proximity to other data when previously unknown relationships may become evident.

Data aggregation (including data integration and data sharing) is intended to overcome the "stovepipe" nature of existing datasets.[192] Research here is focused on making information available to analysts regardless of where it is located or how it is structured.[193]

A threshold systems design issue that has technical, security and privacy implications is whether to aggregate data in a centralized data warehouse or to access information locally in distributed databases.[194] An architecture based on distributed data sources provides additional privacy

189 *See id.*
190 *See, e.g.,* Joint Inquiry Report, *supra* note 22, at 6-7.
191 *See* K. A. Taipale, Presentation at Heritage Foundation, *A Critique of the Markle Report on Trusted Information Networks for Homeland Security*, at slides 5-7, December 11, 2003, at http://www.taipale.org/presentations/CAS-Markle-121103.htm; s*ee also* Hal Varian, *Pricing Information Goods*, in PROCEEDINGS OF SCHOLARSHIP IN THE NEW INFORMATION ENVIRONMENT SYMPOSIUM, HARV. LAW SCHOOL (May 1995).
192 Organizational stovepipes may exist between agencies or within agencies and technical stovepipes between different databases that are technically incompatible or use different data base structures to store information.
193 *See* Taipale, *supra* note 13, at 45-46 (discussing DARPA IAO's project Genisys for virtual data aggregation).
194 Taipale, *supra* note 13, at 42-44 (discussing the technical, security and privacy implications between aggregating data from disparate sources into one central database ("warehousing") or integrating data sources by accessing individual distributed databases ("federated access")).

protection.[195] First, it provides additional technical intervention points for rules-based interventions, for example, for the insertion of "privacy appliance" middleware.[196] Second, a distributed architecture provides additional organizational checks and balances against abuse by maintaining local (or distributed) control over data access, user authentication, and logs.[197]

b) AUTOMATED ANALYSIS

Automated data analysis (including data-mining) is intended to turn low-level data, usually too voluminous to understand, into higher forms (information or knowledge) that might be more compact (for example, a summary), more abstract (for example, a descriptive model), or more useful (for example, a predictive model).[198] "A key problem [for using data mining for counter-terrorism] is to identify high-level things – organizations and activities – based on low-level data – people, places, things and events."[199]

Domestic security or law enforcement needs for data mining differ from commercial data mining applications in significant ways.[200] Commercial data mining techniques are

195 A distributed architecture also has implications for efficiency in information management as well as for system security, *see id.*
196 "Privacy appliances," as envisioned in the TIA program, would act as gateways between databases and analysts and enforce access rules, due process procedures or accounting policies as discussed throughout this article. *See* Matthew Fordhal, *Researchers Seek to Safeguard Privacy in Anti-terrorism Plan*, SEATTLE TIMES, July 14, 2003, *available at* http://seattletimes.nwsource.com/cgibin/PrintStory.pl?document_id=135262838&zsection_id=268448455&slug=btprivacy14&date=20030714; see *also* IAO Report, *supra* note 68, at A-13 ("DARPA is examining the feasibility of a privacy appliance … to enforce access rules and accounting policy.") See discussion of privacy appliances *infra* VII.A.
197 *See* David Jensen, *Data Mining in Networks*, Presentation to the Roundtable on Social and Behavior Sciences and Terrorism of the National Research Council, Division of Behavioral and Social Sciences and Education, Committee on Law and Justice, at slide 18, (Dec. 1, 2002), *available at* http://kdl.cs.umass.edu/people/jensen/papers/nrcdbsse02.html ("This approach keeps institutional control of databases distributed, providing a bulwark against both outside intruders and widespread institutional misuse.")
198 Taipale, *supra* note 13, at 22.
199 Jensen, *supra* note 197, at slide 22.
200 Taipale, *supra* note 13, at 33-35, 47.

generally applied against large transaction databases in order to classify people according to transaction characteristics and extract patterns of widespread applicability. The problem in counter-terrorism is to focus on a smaller number of subjects within a large background population and identify links and relationships from a far wider variety of activities.[201]

Commercial data mining is focused on classifying propositional data from homogeneous databases (of like transactions, for example, book sales) while domestic security applications seek to detect rare but significant relational links between heterogeneous data (people, places, things, activities, associations, etc.). In general, commercial users have been concerned with identifying patterns among unrelated subjects based on their transaction patterns in order to make predictions about other unrelated subjects doing the same.[202] Intelligence analysts are generally interested in identifying patterns that evidence organizations or activities among related subjects in order to expose additional related subjects or activities.[203]

The application of data aggregation and automated analysis technologies to domestic security is the attempt to "make sense of data" by automating certain analytic tasks.[204] Automating such tasks can allow for better and more timely analysis of existing datasets by identifying and cataloging various threads and pieces of information that may already exist but remain unnoticed using traditional means. In addition, these tools can help develop predictive models based on known or unknown patterns to identify additional people, objects or actions that are deserving of further resource commitment or law enforcement attention.

Compounding the problem for data analysis in domestic security applications is that relevant data (that is, information about terrorist organizations and activities) is hidden within vast amounts of irrelevant data and appears innocuous (or at least ambivalent) when viewed in isolation. Individual data

201 IAO Report, *supra* note 68, at A-14.
202 A simple example of commercial data mining techniques that should be familiar to most readers can be experienced at amazon.com, which uses "association rules" to suggest books, CDs and other products that a user might be interested in purchasing on return visits based on correlations between that users purchases and purchases by other users.
203 Taipale, *supra* note 13, at 47.
204 Taipale, *supra* note 13, at 21-22.

items – relating to people, places and events, even if identified as relevant – are essentially meaningless unless viewed in context of their relation to other data points. It is the network or pattern itself that must be identified, analyzed and acted upon.[205]

There are three distinct applications for automated analysis in the context of domestic security:[206]

- first, subject-oriented link analysis, that is, automated analysis to learn more about a particular data subject, their relationships, associations and actions;

- second, pattern-analysis (or data mining in the narrow sense), that is, automated analysis to develop a descriptive or predictive model based on discovered patterns; and,

- third, pattern-matching, that is, automated analysis using a descriptive or predictive model (whether the model itself is developed through automated analysis or not) against additional datasets to identify other related (or "like") data subjects (people, places, things, relationships, etc.).

---

205 See Jensen, *supra* note 197, at slides 21, 22 (identifying the key challenge for counter-terrorism as "analyzing relational data"). An example of how relational data analysis can be useful for counter terrorism can be seen in the analysis of *betweeness* in email traffic. "By looking for patterns in email traffic, a new technique can quickly identify online communities and the key people in them. The approach could mean terrorists or criminal gangs give themselves away, even if they are communicating in code or only discussing the weather." Hazel Muir, *Email Traffic Patterns can Reveal Ringleaders*, NEW SCIENTIST, at http://www.newscientist.com/news/news.jsp?id=ns99993550 (Mar. 27, 2003). *See also* Philip Vos Fellman & Roxana Wright, *Modeling Terrorist Networks: Complex Systems at the Mid-Range, presented at* Complexity, Ethics and Creativity Conference, LSE, September 17-18, 2003; H. Brinton Milward & Jorg Raab, *Dark Networks: The Structure, Operation, and Performance of International Drug, Terror, and Arms Trafficking Networks,* presented at the International Conference on Empirical Study of Governance, Management and Performance, Barcelona, Spain, October 4-5, 2002; Matthew Dombroski et al, Estimating the Shape of Covert Networks, PROCEEDINGS OF THE 8TH INTERNATIONAL COMMAND AND CONTROL RESEARCH AND TECHNOLOGY SYMPOSIUM (2003); D. B. Skillicorn, *Applying Matrix Decomposition to Counterterrorism*, Queen's University, Canada, Technical Report 2004-484 (May 19, 2004).

206 Taipale, *supra* note 13, at 34.

Because spectacular terrorist events may be too rare or infrequent for automated analysis to extract useful patterns, the focus of these techniques in counter terrorism is generally to identify lower level, frequently repeated events (for example, illegal immigration, money transfers, front businesses and recruiting activity), or to identify communication patterns evidencing covert organization, that together may warrant further attention or resource commitment.[207]

Data aggregation and automated analysis are not substitutes for human analytic decision-making, rather, they are tools that can help manage vast data volumes and potentially identify relational networks or other patterns that may remain hidden to traditional analysis.[208] If successful, these technologies can help allocate available domestic security resources to more likely targets.

2. PRIVACY CONCERNS

Because data aggregation and automated analysis technologies can cast suspicion based on recognizing relationships between individually innocuous data, they raise legitimate privacy concerns. However, much of the public debate regarding the potential use of these technologies is overshadowed by simplifications, misunderstandings and misrepresentations about what the technologies can do, how

207 Jensen, *supra* note 197, at slide 25. *See generally* Vladis E. Krebs, *Uncloaking Terrorist Networks*, FIRST MONDAY (mapping and analyzing the relational network among the September 11 hijackers), at http://www.firstmonday.dk/issues/issue7_4/krebs/; see *also* the references in note 205 *supra.*

208 However, the use of probabilistic models developed through data mining can substantially improve human decision-making in some contexts. *See* Jensen, *supra* note 197, at slide 39. Using probabilistic models can focus human attention and resources, can outperform humans in certain limited contexts (for example, in certain clinical medical diagnostic applications), and can encourage an institutional culture of hypothesis testing and probability assessment. *Id. See generally* Tversky, *Judgment under Uncertainty*, *supra* note 39; KAHNEMAN, JUDGMENT UNDER UNCERTAINTY, *supra* note 39 (both describing biased heuristics used in human judgment); Robyn M. Dawes et al., *Clinical versus Actuarial Judgment*, 243 SCIENCE 1668 (describing how statistical/actuarial methods often outperform human judgment in certain diagnostic contexts); Tal Z. Zarsky, *Mine Your Own Business*, 1 YALE J. L. & TECH. 8, 47-48 (2003) ("There is no convincing reason to suppose that decisions made by software are inferior to the ones made by humans (and . . . there are several occasions where the opposite is true).").

they are likely to be employed, and what actual affects their employ may have on privacy and security.[209] Further, arguments premised on finding pattern-matching (or any automated analysis) constitutionally objectionable on their face are based more on ideological rhetoric than legal analysis.

The significant privacy concerns relating to these technologies are primarily of two kinds: those that arise from the aggregation (or integration) of data itself and those that arise from the automated analysis of data that may not be based on any individualized suspicion – the former might be called the *database* problem and the latter the *mining* problem.[210]

The database problem is implicated in subject-based inquiries that aggregate data or access distributed databases to find more information about a particular subject. To the extent that maintaining certain government inefficiencies helps protect individual rights from centralized state power, the primary privacy question involved in aggregation is one of increased government efficiency[211] and the resulting demise of "practical obscurity."[212]

The mining problem is implicated in the use of pattern-matching inquiries, in which profiles or models are run against

209 Even within the technical community there is significant divergence in the understanding what these technologies can do, what particular government research programs entail, and the potential impact on privacy and civil liberties of these technologies and programs. *Compare* Letter from Public Policy Committee of the Association for Computing Machinery to Senators John Warner and Carl Levin (Jan. 23, 2003) (expressing reservations about the TIA program) *available at* http://www.acm.org/usacm/Letters/tia_final.html, *with* Executive Committee SIGKDD of the ACM, *Data Mining is NOT Against Civil Liberties* (June 30, rev'd July 28, 2003) *available at* http://www.acm.org/sigkdd/civil-liberties.pdf (defending data mining technology and expressing concern that the public debate has been ill-informed and misleading); *see also* Rosenzweig, *supra* note 13, at n.6 ("Even among computer professionals there is substantial misunderstanding … [but] those with the seeming greater familiarity with the technologies are less apocalyptic in their reactions.").

210 Taipale, *supra* note 13, at 57-67 (for a more detailed discussion of these issues).

211 *See* Rosenzweig, *supra* note 13, at 181.

212 *See, e.g.,* U.S. Dept. of Justice v. Reporters Committee, 489 U.S. 749, 780 (1989) explicitly recognizing that the aggregation of public records in one place negated the "practical obscurity" that protected those records in the world of distributed paper records. See discussion in Taipale, *supra* note 13, at 58-60.

data to identify unknown individuals. To some, pattern-matching inherently raises privacy issues relating to non-particularized suspicion in violation of the Fourth Amendment.[213] I disagree.

For a particular method to be categorically unreasonable or suspect in the context of the Fourth Amendment, its efficacy – that is, the confidence interval for its particular use – must first be determined and considered. Thus, for example, racial or ethnic profiling for law enforcement purposes may be inherently unreasonable because it is not a reliable predictor of criminality and thus cannot be the sole basis for a reasonable suspicion.[214] (On the other hand, ethnic profiling might be appropriate for medical screening where there is a proven link between a particular condition and ethnic background.)[215]

However, to assert that automated pattern analysis based on behavior or data profiles is inherently unreasonable or suspect without determining its efficacy – that is, without determining the relationship between the pattern and its results in a particular application – seems not only inappropriate but also analytically unsound.[216]

---

213 *See* Taipale, *supra* note 13, at 60-67. I use the phrase *non-particularized suspicion* in discussing the charge by some privacy advocates that these technologies or techniques using pattern-based queries may be constitutionally suspect because they do not meet the general requirement that "some quantum of individualized suspicion is usually a prerequisite to a constitutional search or seizure," United States v. Martinez-Fuerte, 428 U.S. 543, 561 (1976). However, it should be noted that "the Fourth Amendment imposes no irreducible requirement of such [individualized] suspicion," *id.*, and the appropriate test requires balancing the potential intrusion with the state interest in the context of the circumstances.

214 *See* United States v. Brignoni-Ponce, 422 U.S. 873, 886 (1975). Note, however, the Supreme Court has never ruled explicitly on whether race can be a *relevant* factor for reasonable suspicion under the Fourth Amendment. *See id.* at 885-887 (implying that race could be a relevant, but not sole, factor). *See also* Whren v. United States, 517 U.S. 806, 813 (1996).

215 For example, Tay-Sachs disease, *see* http://www.marchofdimes.com/professionals/681_1227.asp, last visited Dec. 6, 2004 or sickle cell anemia, *see* http://www.ascaa.org/comm.htm, last visited Dec. 6, 2004.

216 And, the asserted equivalence of behavior profiling with racial profiling (and therefore inherently intrusive of privacy or otherwise abhorrent) is a rhetorical tactic used to misdirect the public debate and such assertion does not hold up under more rigorous analysis. Behavior is, of course, the very foundation for assessing suspicion under any constitutional test.

For example, a pattern-based analysis (automated or not) that was 100% accurate (gave no false positives and no false negatives) in identifying terrorists and only terrorists before they acted could not be constitutionally unreasonable. Such accuracy would far exceed even a stringent requirement of probable cause – indeed, absolute accuracy (if it were possible) would prove guilt beyond a reasonable doubt. Thus, the policy issue with regard to pattern-matching ought to be deciding what accuracy rate is appropriate or required under what circumstances (or what error rate is acceptable as reasonable) and what consequences appropriately flow from its use, not demonizing a technology or technique.

Indeed, the Supreme Court has specifically held that the determination of whether particular criteria are sufficient to meet the reasonable suspicion or probable cause standard does not turn on the probabilistic nature of the criteria but on their probative weight. In *United States v. Cortez*[217] the Court opined:

> The process [of determining reasonable suspicion] does not deal with hard certainties, but with probabilities. Long before the law of probabilities was articulated as such, practical people formulated certain common-sense conclusions about human behavior; jurors as factfinders are permitted to do the same - and so are law enforcement officers.[218]

In *United States v. Sokolow*[219] the Court specifically rejected the approach of the Court of Appeals, which had divided criteria into evidence of "ongoing criminal behavior," on the one hand, and "probabilistic" evidence, on the other. Further, with respect to the use of patterns to establish suspicion,[220] the Court held:

> Any one of these factors is not by itself proof of any illegal conduct and is quite consistent with innocent travel. But we think taken together they amount to reasonable suspicion. See *Florida v.*

217 449 U.S. 411 (1981).
218 *Id* at 418.
219 490 U.S. 1, 8 (1989).
220 Sokolow involved the use of DEA drug courier profiles.

> *Royer*. We said in *Reid v. Georgia*, "there could, of course, be circumstances in which wholly lawful conduct might justify the suspicion that criminal activity was afoot." Indeed, *Terry* [*v. Ohio*] itself involved "a series of acts, each of them perhaps innocent" if viewed separately, "but which taken together warranted further investigation." See also [*United States v.*] *Cortez*. We noted in [*Illinois v.*] *Gates* that "innocent behavior will frequently provide the basis for a showing of probable cause," and that "[i]n making a determination of probable cause the relevant inquiry is not whether particular conduct is 'innocent' or 'guilty,' but the degree of suspicion that attaches to particular types of noncriminal acts." That principle applies equally well to the reasonable suspicion inquiry. [citations omitted][221]

The fact that patterns of relevant criteria may be generated by automated analysis (data-mined) or matched through automated means (computerized matching) should not change the analysis – the reasonableness of the basis for suspicion should be judged on its probative value (efficacy in evidencing reasonable suspicion) in the particular circumstances of its use[222] – not on its probabilistic nature or whether it is technically mediated.

But, it is further argued, applying automated pattern-based analysis to data may still be qualitatively different from existing methods because it is initiated not on the trail of a particular suspect individual but on a non-particularized suspicion that a given data population might contain evidence of "terrorists." For example, Solove writes that pattern-matching "alters the way that government investigations typically occur"[223] and Priscilla Regan contends that pattern-matching investigates everyone, and most people who are investigated are innocent.[224]

---

221 490 U.S. at 9-10.
222 *See* Cortez, 449 U.S. at 417 ("But the essence of all that has been written is that the totality of the circumstances - the whole picture - must be taken into account.").
223 Solove, *supra* note 63, at 1109.
224 PRISCILLA M. REGAN, LEGISLATING PRIVACY 90 (1995).

But, how is pattern-matching *qualitatively* different (for that is the claim) then existing methods? While there is certainly a legitimate debate to be had about whether a particular technique (see discussion of method of inquiry below) should be applied to a particular data space (see discussion of scope of inquiry and sensitivity of data below), that debate is not premised on the novelty of automated behavior or data profiling, only the expectation of (or desire for) privacy for certain data sets (assuming comparable circumstances).

The point is not that there is no privacy issue involved with the electronic equivalent of observing suspicious behavior in a public space based on a general concern about terrorism, but only that the issue to be resolved is the traditional concern of expectations of privacy for the particular space being observed under those circumstances, and the reasonableness of the government action based on the observation – not a categorical finding of "non-particularized suspicion" based on technique.[225] The doctrine of non-particularized suspicion[226] ought to only have applicability in cases where the pattern matching is unreasonable because the pattern itself is found not to be an effective predictor of criminality (for example, profiling based *solely* on race) thus cannot be said to provide reasonable or probable cause to take further action.[227]

Thus, from a policy point of view, the issue to be determined regarding automated analysis and pattern-matching

---

225 *Cf. Martinez-Fuerte*, 428 U.S. at 559-67 (holding routine traffic checkpoints and selective referral for secondary inspections reasonable under the Fourth Amendment).

226 If there even is a constitutional requirement for individualized suspicion. As noted above, the Court has explicitly recognized that the requirement for individualized suspicion is not irreducible. *See supra* note 213. In *Treasury Employees v. Von Raab*, 489 U.S. 656, 665-66 (1989), the Court also explicitly noted that where the government interest – as here with terrorism – serves a need beyond routine law enforcement, the *practicality* of requiring individualized suspicion is also a relevant factor:

> Our cases establish that where a Fourth Amendment intrusion serves special governmental needs, beyond the normal need for law enforcement, it is necessary to balance the individual's privacy expectations against the Government's interests to determine whether it is impractical to require a warrant or some level of individualized suspicion in the particular context.

227 *See* discussion *supra; Brignoni-Ponce*, 422 U.S. at 886.

is what predicate procedures and/or oversight, and what judicial review, should be applied to particular pattern-based applications or programs to ensure that they are effective and reasonable for their intended purpose within the totality of the circumstances of their use,[228] and how that efficacy relates to what data they should access. Pattern-based queries, even those based on behavior or data profiling, are reasonable or unreasonable only in the context of their intended application[229] – not because they are automated or not. Establishing procedures for implementation and making a determination of reasonableness either prior to employ (authorization), during their general use (oversight) or after their use in a particular case (judicial review) requires analyzing the *calculus of reasonableness* described below.

To argue that the use of a technique – pattern-matching – is inherently unreasonable under the Fourth Amendment without actually determining the reasonableness of its use (based on its efficacy in context and the circumstances of its use) would mean that the Fourth Amendment prohibits all preemptive policing strategies. That is, any investigation or allocation of resources based on the observation of suspicious behavior prior to the actual commission of a crime (or based on the analysis of historical experience with criminal activity) would be unreasonable absent a prior showing that the specific observed behavior relates to that particular crime. Taking this to its logical extreme would mean that any proactive policing strategy to allocate resources, including assigning a police officer to a high crime location, would be unconstitutional if it was not in response to a specific observed behavior relating to a particular crime. Such a conclusion seems unwarranted as a matter of both policy and law.[230]

Nevertheless, because they are statistical based techniques, it is important to emphasize again that these technologies, including pattern-based queries, should generally

228 *Cf.* TAPAC Report, *supra* note 90, at 45-59 (recommending programmatic approval for data mining applications).

229 And, determining such reasonableness requires judging, among other things, the severity of the consequences to the individual from a potential false positive match. A false positive with limited consequences – for example, a non-intrusive follow-up investigation – must be weighed against the legitimate state interest in security.

230 *See, e.g.,* Martinez-Fuerte, 428 U.S. at 559-67; *see generally* ANTHONY A. BRAGA, PROBLEM-ORIENTED POLICING AND CRIME PREVENTION (2002).

not to be used to determine guilt or innocence but rather to allocate security resources exactly like any other proactive policing strategy. Thus, their reasonableness for general use in allocating resources or attention should be judged according to the policy calculus described below and their constitutional reasonableness in a particular case can be subject to later judicial review just like any other investigative procedure.

There are legitimate privacy concerns relating to the use of these technologies (just like with any other policing technique) – but there is not a presumptive Fourth Amendment non-particularized suspicion problem *inherent in the technology or the technique* even in the case of automated behavioral or data pattern-matching. The privacy issue in both subject- and pattern-based queries is determining reasonableness in the context of a particular use – and that determination hinges on the calculus set forth below not a predetermination that a technology or technique is categorically suspect.

C. TECHNOLOGIES OF COLLECTION

The technologies of *collection* are those that extend the edge of the network or add information to the data sets.[231] Broadly, collection technologies may include the identification technologies and aggregation and analysis technologies already discussed above (to the extent that they add new data), as well as what are commonly referred to as surveillance or search technologies, including image and signal collection technologies.[232]

For example, facial recognition technology can serve as both an identification technology – authenticating a particular attribute at the point of verification – or as a collection technology – by recording an identity match as a transaction record. So too, data aggregation can be viewed in terms of either

231 This Article is generally concerned with the issues arising from digital information systems and networks. These systems significantly affect the five processes that determine information management strategies in social activity, including law enforcement – their collection; storage; transmission; selection; and intelligent processing. Existing policies, including privacy policies and related regulatory and legal structures, premised on old models for assessing and controlling these processes are under significant stress.

232 *Cf.* Froomkin, *supra* note 99, at 1475-1500 (describing various "ubiquitous surveillance" technologies).

aggregating data sets or collecting additional data about a subject. Also analysis, to the extent that it creates new data (for example, identifying a pattern or link), can be viewed as a collection technology because it creates new information about (or that can be linked to) existing data.[233]

However, this section is generally concerned with those collection technologies more commonly associated with surveillance or search and often referred to as *sense-enhancing* technologies.[234]

1. SENSE-ENHANCING TECHNOLOGIES AND SECURITY

Sense-enhancing technologies are simply those technologies used to enhance the human ability to observe or recognize physical characteristics or activities.[235] Generally, these technologies can be further classified as those that *amplify* the existing human senses, or those that *extend* these senses by making previously undetectable phenomena observable. Examples of technologies that amplify human senses include such 'low-tech' devices as binoculars, telescopes and cameras as well as 'no-tech' devices such as drug sniffing dogs, but also include 'high-tech' devices, for example, sensors that can hear through walls. Examples of technologies that extend the senses include various devices that measure wavelengths not usually detectable, such as infrared or ultra-violet sensors, radar and

233 An interesting question that has not yet been addressed is whether queries themselves (or new information, such as links, generated through query) become part of the information about the individual and subject to some new legal interest on the part of that individual – *i.e.*, does the fact of processing itself become part of the digital dossier. If so, other questions arise: will queries be subject to, for example, the Freedom of Information Act, 5 U.S.C. § 552, *amended by* Public Law No. 104-231, 110 Stat. 3048 (1996) (assuming for a moment that the intelligence or law enforcement exceptions are not applicable)? Should they be as part of error correction policies? *Cf.* how credit report queries become part of the report and are themselves then subject to the Fair Credit Reporting Act, 15 U.S.C. § 1681 et seq.

234 *See* Froomkin, *supra* note 99, at 1496 ("Sense enhanced searches rely on one or more technologies to detect that which ordinarily could not be detected with un-aided human senses"); David. E. Steinberg, *Making Sense of Sense-enhanced Searches*, 74 MINN. L. REV. 563, 563 n.1 ("this article uses the term "sense-enhanced search" to describe any [search] through the use of some method that provides information not available to the unaided sensory perceptions.").

235 *Cf.* Steinberg, *supra* note 234.

even radio receivers. Additionally, sense-enhancing technologies could be categorized as those that provide a new perspective from which to observe, for example, aircraft over-flight or satellite remote sensing[236] and those that interpose human senses (*e.g.*, hearing) within mediated information flows, for example, the classic wiretap intercepting electronic representations of human speech from the telephone system.

It is beyond the scope of this article to discuss specific sensing technology in any detail. Rather, we seek here to understand functionally how these collection technologies interact with system design and privacy concerns, and where technical intervention or design strategies might be applicable to provide for due process protections for privacy policy.[237]

2. PRIVACY CONCERNS

Privacy issues involving remote or enhanced sensing generally tend to revolve around the appropriateness of place and manner of collection and turn on whether there exists a *reasonable expectation of privacy*.[238] In general, the courts have struggled to determine whether the use of any particular new technology is a *search* under the Fourth Amendment, and, if so, was it *reasonable*.[239]

Historically, the Supreme Court based its analysis on the existence (or absence) of a physical trespass or handling of property in deciding whether a challenged government action

236 These categorizations are not exclusive and some technologies provide multiple functions, for example, a low light camera with telephoto lens mounted on an aircraft amplifies human vision, extends it into low light situations, and allows for direct overhead observation.
237 For an overview of policy considerations to be considered with sense-enhanced surveillance systems, see, for example, ABA Standards for Technology-enhanced Surveillance, *supra* note 163.
238 Katz v. United States, 389 U.S. 347, 361 (1967) (holding that the use of a wiretap requires a warrant under the Fourth Amendment).
239 U. S. CONST. AMEND. IV:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

amounted to a search requiring a warrant under the Fourth Amendment.[240] Despite the rejection of this physicality test in *Katz v. United States*,[241] the courts have continued to proceed "from the premise that [sense-enhanced] searches are less intrusive than physical searches" in analyzing new technologies,[242] and "remain likely to invalidate only those warrantless searches that involve a physical trespass into a constitutionally protected area."[243] Thus, for example, the Court has upheld the warrantless use of a beeper to monitor a container transported by car,[244] but found that monitoring such a beeper in the home violates the Fourth Amendment.[245]

In rejecting the physicality test in *Katz*, the Court set out the two-part *reasonable expectation of privacy* test, which requires finding both an actual *subjective* expectation of privacy and a reasonable *objective* one:

> My understanding of the rule that has emerged from prior decisions is that there is a twofold requirement, first that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as "reasonable."[246]

Although the Court has never explicitly articulated the following distinction, I would generalize the Court's approach to new technologies since Katz as follows:

Where a new technology merely amplifies the existing human senses the Court has generally upheld sense-enhanced searches by inferring no reasonable *subjective* expectation of privacy since the technology in question merely allowed the observation of activity that could have been viewed anyway and

240 Olmstead v. United States, 277 U.S. 438, 464 (1928) (upholding warrantless wiretaps because such searches did not involve a physical trespass) (overruled in *Katz*).
241 *Katz*, 389 U.S. at 361.
242 Steinberg, *supra* note 234, at 568.
243 *Id.* at 585.
244 United States v. Knotts, 460 U.S. 276, 281 (1983) ("A person traveling in an automobile on public thoroughfares has no reasonable expectation of privacy in his movements from one place to another." *Id.* at 281)
245 United States v. Karo, 486 U.S. 705, 706 (1984) ("monitoring of a beeper in a private residence, a location not opened to visual surveillance, violates the Fourth Amendment.")
246 *Katz*, 389 U.S. at 361 (Harlan, J., concurring).

"what a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection."[247] Thus, under an analogy to the plain view exception to the warrant requirement,[248] the Court has sanctioned the use of airplanes to conduct an aerial search without a warrant,[249] and the use of an electronic beeper to track movements on the public roads.[250] Following this same reasoning, the courts have also generally upheld the use of telescopes and binoculars, even in some cases when used to view a constitutionally protected area, for example, a home or office, from a public space.[251]

However, where a new technology extends the ability for human senses to observe something previously commonly held to be unobservable, the Court has extended Fourth Amendment protection, in essence arguing that there is a presumptive *objective* expectation of privacy in such cases. Thus, in *Kyllo v. United States*, the Court held "where, as here, the Government uses a[n infrared scanning] device that is *not in general public use*, to explore details of the home that would *previously have*

247 *Katz*, 389 U.S. at 351.
248 Harris v. United States, 390 U.S. 234, 236 (1968); *see also* Steinberg, *supra* note 234, at 596-601.
249 California v. Ciraolo, 476 U.S. 207, 213-14 (1986) ("Any member of the public flying in this airspace who glanced down could have seen everything that these officers observed."); Dow Chemical v. United States, 476 U.S. 227, 238 (1986) (upholding use of sophisticated cameras during overflight, "the mere fact that human vision is enhanced somewhat, at least to the degree here, does not give rise to constitutional problems"); Florida v. Riley, 488 U.S. 445, 451 (1989) (upholding warrantless use of helicopter at 400 feet, "[a]ny member of the public could legally have been flying over Riley's property in a helicopter at the altitude of 400 feet and could have observed Riley's greenhouse.")
250 Knotts, *supra* note 244. Under what could be called the "plain smell" doctrine, the Court has also upheld the use of dogs to enhance smell. United States v. Place, 462 U.S. 696, 699 (1983). Following this reasoning, it has been argued that passive alcohol sensors are not a search within the meaning of the Fourth Amendment, *see, e.g.,* SHENEQUA L. GREY, PASSIVE ALCOHOL SENSORS AND THE FOURTH AMENDMENT, CIVIC RESEARCH INSTITUTE (Spring 2001) *available at* http://www.ndaa-apri.org/apri/programs/traffic/passive_alcohol_sensors_fourth_amendment_2.html. The ACLU has opposed such passive sensors, see, for example, *'Sniffer' Device a Violation of Privacy?,* ABOUT.COM *at* http://alcoholism.about.com/library/weekly/aa000823a.htm
251 *See* Steinberg, *supra* note 234, at 605-609.

been unknowable without physical intrusion, the surveillance is presumptively unreasonable without a warrant."[252]

In analyzing new technology-enhanced searches, the Court has also drawn a distinction between aural and visual enhancements, in general applying a stricter standard to enhanced hearing than vision.[253] For example, in *Dow Chemical v. United States* the Court upheld the warrantless use of sophisticated camera equipment used in an aerial search but concluding that "an electronic device used to penetrate walls or windows so as to hear and record confidential discussions … would raise very different and far more serious questions."[254] Commentators have attributed this distinction to a particular concern for protecting communication, implicating First Amendment concerns.[255] Whether this heightened concern for communication would extend to email or other forms of electronic communication has not been resolved.[256]

In any case, as Solove has pointed out, the Court's current conception of Fourth Amendment protected privacy is based primarily on maintaining *secrecy* of information: "The Court's new conception of privacy is one of total secrecy. If any information is exposed to the public or if law enforcement officials can view something from any public vantage point, then the Court has refused to recognize a reasonable expectation of privacy."[257]

252 Kyllo v. United States, 533 U.S. 27, 40 (2001) (emphasis added).
253 Steinberg, *supra* note 234, at 592.
254 *Dow*, 476 U.S. at 238-239.
255 *See, e.g.*, Steinberg, *supra* note 234, at 595.
256 However, the courts of appeals have held that the interception of communications from radio frequencies that are accessible to the general public does not constitute a Fourth Amendment search, even though radio waves cannot be perceived by natural senses (and certainly involve human communication). *See, e.g.,* United States v. Rose, 669 F.2d 23, 26 (1st Cir. 1982); Edwards v. Bardwell, 632 F. Supp. 584, 589 (M.D. La. 1986), *aff'd*, 808 F.2d 54 (5th Cir. 1986). And, with respect to email, *cf.* United States v. Councilman 373 F.3d 197(1st Cir. 2004) (holding that private party interception of email while in temporary storage during transit is not a violation of the Wiretap Act, 18 U.S.C. § 2510 *et seq.*). Although this case turns specifically on questions of statutory interpretation it appears that email may not fit easily within historical protections against wiretapping verbal communications.
257 Solove, *supra* note 63 at 1133; *see also id* at 1122 ("after Katz, the Court shifted to viewing privacy as a form of total secrecy"); *Id* at 1131, 1136, 1152.

Earlier in this Article, I argued that the privacy lobby has a fetish for secrecy of data over autonomy of the individual – here then is the source of such fetish: by rooting its conception of privacy narrowly in total secrecy based on concealment, the Supreme Court has constructed an artificial all-or-nothing standard at odds with the implicit balancing of interests required by the Fourth Amendment's demand for reasonableness.[258]

An additional problem with the Court's emphasis on secrecy and analysis of expectation based on whether a technology is in common usage, is that it is an uncertain standard sure to shrink the realm of privacy since the more common a particular intrusive technology becomes, the less the Fourth Amendment will protect.[259]

Solove has argued that this limited conception of privacy as total secrecy is not in keeping with the architecture of power that the Fourth Amendment was meant to provide.[260] Other commentators have also argued for a new conception of privacy based more on protecting autonomy of the individual rather than secrecy of information.[261]

Reconceptualizing privacy to favor protection of autonomy (over secrecy of data for its own sake) could both extend and restrict the domain of privacy.[262] For example, any attempt to gain information to interfere with an individual's rationale choice, *i.e.*, their autonomy, could be held to implicate their privacy (whether or not the data was secret) and thus trigger whatever appropriate procedural or legal protections might

258 *See generally* Solove, *supra* note 63; Taslitz, *supra* note 131.
259 *Cf.* Froomkin, *supra* note 99, at 1523 ("The more commonplace that ubiquitous surveillance becomes, the less the Fourth Amendment will be able to protect the average citizen."). Additionally, could expectation, for purposes of the *Katz* test, then be subject to overt manipulation? Could, for example, the government widely disseminate information on a new intrusive technology prior to its employ, thus undermining any credible claim to either subjective or objective expectation of privacy thereafter?
260 *See* Solove, *supra* note 63, at 1117-38.
261 *Cf., e.g.,* Alfino, *supra* note 70 at 6, ("Privacy plays a fundamental and ineliminable role in constructing personal autonomy."); Cohen, *supra* note 120 at 1377, 1425 (arguing for "zones of personal autonomy," within which the individual can develop and make autonomous "decisions about speech, belief, and political and intellectual association").
262 Alfino, *supra* note 70.

exist.[263] On the other hand, clandestine surveillance (or data analysis) of which the individual is unaware and through which they are not subject to any additional consequence, might not be considered to implicate privacy at all since such surveillance has no affect on autonomy.[264] It is beyond the scope of this article to fully explore these issues.

Nevertheless, the salient point is that a more sophisticated or nuanced view of privacy – for example, one based on protecting individual autonomy, not data secrecy for its own sake – provides opportunities to build technical information systems designed to protect core privacy interests while still improving security.[265] Autonomy can be protected by separating knowledge (or observation) of behavior from knowledge (or discovery) of identity – in effect, using a form of procedurally protected anonymity to protect autonomy, nee privacy. Thus, strategies based on protecting what I have called the *privacy divide* (that is, the point where data attribution occurs) by building in procedural interventions through anonymization or pseudonymization can help protect these core privacy interests in privacy while still meeting legitimate law enforcement and national security needs.[266]

VIII. THE PRIVACY DIVIDE

Reconciling competing interests in the privacy-security debate requires determining under what circumstances (and following what procedures) identity is to be associated with behavior or behavior with identity.[267] This section explores the

263 That is, whatever combination of administrative, legislative and judicial protections that develops. *Cf. e.g.,* Solove, *supra* note 63, at 1151-67 (describing reconstructing a new "architecture of power" based on a fusion of constitutional and statutory protections).

264 *Cf.* the approach of the *EU Directive, supra* note 129, where processing itself may be considered a privacy violation.

265 *Cf. generally* Daniel J. Solove, *Conceptualizing Privacy*, 90 CAL. L. REV. 1087, 1126-1155 (2002) (discussing the complexity of conceptualizing privacy but arguing for a more contextual based approach).

266 *See* K. A. Taipale, Presentation at the Potomac Institute for Policy Studies, The Politics and Law of Identity and Identification in the Context of the War on Terror, Arlington, VA, at slides 13-16 (Jan. 28, 2004), *available at* http://www.taipale.org/presentations/CAS-IDsystems-012804.pdf.

267 In this section, I use *identity* in its more common form to mean the identification of a particular individual. *Cf.* the discussion in Part VII.A. *supra*.

privacy divide – that point within systems and security processes where attribution of behavior and identity occurs, and argues that the key to developing new information technologies and systems that can provide improved security while still protecting privacy is to design systems that allow for procedural intervention and control at that point.

The question for both policy and information systems development is when and under what circumstances certain data attribution is to occur – that is, when is an entity or identifier to be attributed to a data collection (or its related attributes) or, conversely, when is a data collection (or its related attributes) to be attributed to an identifier or entity.[268] Simply put, when is an individual to be associated with data representing their behavior or, conversely, when is behavior (whether observed by physical surveillance or within data) to be attributed to a specific individual.

Traditionally, the security purpose for using identification technologies and systems is to attribute an individual to (or associate an individual with) an identifier or identity in order to grant authority or provide accountability.[269] However, technical means exist to prove authorization (for example, through third party *certification*) and to provide accountability (for example, through *identity escrow*) without necessarily disclosing individual identity at the point of verification while still meeting either (or both) security needs.[270]

Likewise, data analysis, including pattern-analysis and pattern-matching, is used to attribute (or associate) behavior or transaction records (suspicious links or patterns) to an individual for further investigation.[271] Similarly, traditional surveillance technologies, for example video cameras in public places, are used to attribute observed behavior to an individual for follow-up investigation (or to record behavior for later accountability).[272] In both of these cases as well, system design can provide intervention points for legal or policy procedures to control the circumstance under which such behavioral analysis or observation is attributed to a particular individual. Some

268 *See generally supra* Part VII.A.
269 *See* text accompanying *supra* note 147.
270 *See* discussion of *certification* and *identity escrow* in Part VIII.C.1. *infra*
271 *See generally, supra* Part VII.B.
272 *See* Froomkin, *supra* note 99, at 1476-79.

form of rule-based processing, either automated or procedural (or both), can be interposed at the privacy divide, that is, before identity is attributed to behavior.

A. CONTROLLING THE PRIVACY DIVIDE: THE PRIVACY APPLIANCE AS METAPHOR

Conceptualizing and designing the appropriate mechanisms to exert control over the privacy divide is the key issue for protecting privacy in networked information systems. The policy challenge is to determine the rules and procedures governing the divide, and the technical challenge is to build in technical features to execute or enforce those rules and to manage accountability.[273] The overall architecture must include organizational, procedural, and technical features in a framework that integrates these control requirements within business process needs.[274]

The notion of a *privacy appliance* – that is, a technical systems component sitting between the point of access and the data itself to enforce rules and provide accountability – has been suggested.[275] Here I consider the privacy appliance as a metaphor – an analytic device representing the need for policy intervention in technical systems to enforce rules – rather than as a particular technical device or application. As a technical matter, the privacy appliance might be instantiated as any one of several different approaches,[276] and its actual form of implementation is secondary to understanding what business needs are to be supported.

---

273 *See infra* Part IX (policy calculus) and Part X (technical features).

274 *See supra* note 143 (discussing business process and enterprise architecture).

275 *See* Taipale, *supra* note 13, at 78 n.328, 80. *See also* IAO REPORT, *supra* note 68, at A-13 ("DARPA is examining the feasibility of a privacy appliance … to enforce access rules and accounting policy"). For purposes of this Article I continue to use the term 'privacy' appliance although in other work I have begun to refer to 'policy' appliances in order to encompass the broader notion that these mechanisms can be used to enforce policy rules more generally, including rules for operational security and information assurance, for example, in addition to meeting privacy needs. *See, e.g.,* K. A. Taipale, *Designing Technical Systems to Support Policy (Enterprise Architecture, Policy Appliances, and Civil Liberties),* in ROBERT POPP AND JOHN YEN, 21ST CENTURY ENABLING TECHNOLOGIES AND POLICIES FOR COUNTERTERRORISM (forthcoming 2005).

276 *See* Taipale, *supra* note 13, at 75-78.

From a policy perspective, it is not necessary *a priori* to decide if the privacy appliance should be a specific piece of hardware, for example, a firewall, or an application, such as an analytic filter. The point for policy makers to understand is that intervention can happen at many different points in the technical architecture,[277] and can be subject to varying methods of control.[278]

For the policy maker, the privacy appliance represent the technical objects that enforce policy in systems, thus who controls them (for example, the party using the data, the party supplying the data, a trusted or untrusted third party) and how (through direct technical control, automated monitoring, control of audit or logs) and subject to what general oversight and review (for example, executive, legislative or judicial) are the pertinent policy questions.

For the technologist, understanding the policy needs forms the basis for determining technical requirements. Together, policy needs and system design form an enterprise architecture in which the information management, data, systems, and technology architecture all support the overall business process needs[279] – in this case, enabling certain security processes while still protecting privacy interests.[280]

This article suggests that technical design strategies that emphasize *anonymization* of data for analysis and *pseudonymization* of identity for identification and surveillance systems can provide intervention points where due process procedures can function. Together with strong user authentication and audit controls, these strategies can be built into privacy appliances that mediate between distributed data

277 Intervention can occur both between nodes in a topographical layout of the network as well as at each layer, for example, at the interface, authentication, application, transport, or database level, *see* Taipale, *supra* note 143, at slides 28-29; K. A. Taipale, *Technology, Security, and Privacy: Designing Technical Features to Support Policy*, Presentation at the NSF Science and Technology Center for Discrete & Theoretical Computer Science (DIMACS)(April 15, 2004) at slides 69-70, *available at* http://www.taipale.org/presentations/CAS-DIMACS.pdf.

278 Control can be located with the user, the data owner, or externally.

279 HOWARD SMITH & PETER FINGAR, IT DOESN'T MATTER - BUSINESS PROCESSES DO (2003).

280 *Cf.* Cohen, *supra* note 120, at 1436-38 (discussing informational privacy and technical design choice).

sources and the analyst on one hand (for data aggregation and analysis) and the individual and the data collection on the other (for identification and collection). Such procedures are dependant on organizational, structural and technical design features functioning together to meet articulated policy needs.[281]

B. ANONYMIZATION OF DATA

An anonymous record or transaction is one in which the data cannot be associated with a particular individual, either from the data itself, or by combining the transaction with other data.[282] Here traditional encryption strategies need to be distinguished from potentially truly anonymous procedures such as *one-way hashing*. With traditional encryption strategies data is encrypted but can be decrypted with the use of a key. An analogy would be handing over data in a locked box. Theoretically, encrypted data is not truly anonymous as the underlying data can be accessed by combining it with the key (and the key might be acquired by applying brute computational force or otherwise).

Using one-way hashes, on the other hand, allows data to be shared "with your worst enemy," as the original data cannot be reconstructed from the hash.[283] Hashing is a process of passing data through a one-way algorithm that returns a digital signature in place of the original data.[284] This digital signature is unique to the underlying data but cannot be turned back into the original data – much like a sausage can be identified as pork but cannot be turned back into a pig.[285] One-way hash technologies allow for anonymous data processing to occur in a

281 *Cf.* Second Markle Report, *supra* note 23, at (discussing SHARE network).
282 Roger Clarke, *Identified, Anonymous and Pseudonymous Transactions: The Spectrum of Choice* § 3.3, Presented at User Identification & Privacy Protection Conference (June 14-15, 1999), *available at* http://www.auu.edu/au/ people/Roger.Clarke/0v/UIPP99.html.
283 *See* Don Clark, *Entrepreneur Offers Solution For Security-Privacy Clash*, WALL ST. J., Mar. 11, 2004, at B:1 (quoting Kim Taipale, "with data hashing, 'you can hand your data over to your worst enemy and they don't have anything.") Note, however, the discussion of the use of *salt* to prevent "dictionary attacks" *infra.*
284 Dempsey, *supra* note 16, at 7.
285 Jeff Jonas, SRD, *quoted in* Steve Mollman, *Betting on Private Data Search*, WIRED MAG., Mar. 5, 2003, *available at* http://www.wired.com/ news/technology/0,1282,57903,00.html.

shared environment since the only thing exchanged is the hash.[286] If a match occurs, the processing party still needs to come back to the original data holder to access the underlying data. This allows organizational and procedural structures to be imposed between data matching and revelation of identity (or other sensitive data).

Theoretically, hashing is vulnerable to what is known as a "dictionary attack" in which an attacker compiles a list of all potential inputs and computes a hash function for each, then compares the hashed output from the target data set against their own list of hashes computed from all possible inputs to determine if there is a match.[287] To counter the dictionary attack, *salt* is used. Salt is a random string that is concatenated to the original data before it is operated on by the hash function.[288] In order to match the hashed outputs you need to share the salt key. Salt keys can be encoded in hardware or software, can be used to control the domain across which sharing occurs, and can even be used to control expiration of data.[289]

1. ANONYMIZATION AND SECURITY

Many security needs for data analysis, including watch list matching and pattern-matching, can be accomplished within an anonymized data framework. Data analysis, including some forms of data mining, may also be possible.[290]

---

286 This also allows the insertion of a trusted or untrusted third party to provide additional protections. For example, two parties wishing to data-match a particular data set, for example, a list of names, can give each give their hashed lists to an independent (trusted or not) third party for processing rather than exchanging data directly. This allows organizational structures in which even the identity of the counterparty is unknown.

287 For example, to determine whether a particular hashed value compares to a specific name, the hacker would compute hashes for all possible inputs, in this case, all names in the data universe, and then compare the output with the original hash. If there were a match, the hacker would look at their own list of inputs and determine the original data. In addition to using salt to overcome vulnerability to dictionary attacks, systems can also be designed to fragment data sets for discrete processing.

288 See Dempsey, *supra* note 16, at 8-9.

289 *Id.*

290 It should be noted that there is a constant tension between analyzing de-identified data and re-identifying the data, that is, the more effective a

A simple example of this is illustrated by the following.[291] Suppose a primary dataset contains traveler data and a second dataset contains suspected terrorists. Before the data was analyzed, both datasets are subject to the same one-way hash algorithm. Now, identifiers for "Joseph K." in the first dataset (for example, name, birth date, telephone number, etc.) are represented by encrypted digital signatures that do not reveal personal data but can still be exchanged or matched against the corresponding data in the second set since the matching name or other identifiers in that database would have a matching "hash" (digital signature). Should a match occur, the agency would be required to follow the appropriate administrative or judicial procedures for that particular use prior to being granted access to the raw data corresponding to the match held by the original owner who maintains exclusive control of the actual data throughout.

As noted above, by controlling the sharing of *salt keys* additional policy restrictions can be enforced. Not only can hashed data not be turned back into the original data, it cannot be matched or used for any other purpose without a matching salt key. Thus, control of salt variables allows searches to be restricted to certain data sets or domains.[292]

Beyond simple list matching, more sophisticated link analysis may also be possible using anonymized data.[293] Whether full-scale data analysis, including data mining for unknown patterns, is possible using (anonymized or otherwise de-identified) data is a research question.[294]

2. DEVELOPMENT IMPERATIVES

While anonymizing technologies provide privacy protection they are not dependant on privacy concerns alone for

technology is at analyzing de-identified data, the more it is able to re-identify data without resort to traditional identifiers. Data analysis is, by definition, the process of making sense of data. Thus, the goal in designing technical systems is not to maintain absolute secrecy of data but to provide sufficient layers of abstraction at which due process intervention can occur.

291 *Cf.* Dempsey, *supra* note 16, at 9-10.

292 *Cf. id.*

293 *See* Mollman, *supra* note 285; Clark, *supra* note 283.

294 *See, e.g.,* Dawn Song et al., *Practical Techniques for Searches of Encrypted Data*, Proc. of IEEE SRSP, May 2000.

development. Intelligence agencies themselves have a need to develop methods for anonymous data transfer and processing in order to enable the "sharing" of confidential data with untrusted sources. For example, domestic intelligence or law enforcement agencies may need to match data with corporate databases (for example, employee records) without revealing the watch list names. So too, even among government agencies there is a need to protect sources and methods (as well as avoiding potential liability) that precludes sharing raw data. Thus, data anonymization strategies are not at odds with security – rather, they serve both privacy and security needs.

C. PSEUDONYMITY

A pseudonymous record or transaction is one that cannot – in the ordinary course of events – be associated with a particular individual.[295] Pseudonymity is a form of "traceable anonymity" and requires legal, organizational or technical procedures so that the association (that is, the data attribution) can only be accomplished under specified and controlled circumstances. Pseudonymity is also referred to as *identity escrow*.[296]

The use of pseudonyms allows for anonymous but traceable (or otherwise accountable) identities to be used.[297] A pseudonym can be either *transient* (used once or for a limited time) or *persistent* (used over a period of time). Persistent

295 Clarke, *supra* note 282.
296 *See, e.g.,* Joe Kilian and Erez Petran, *Identity Escrow*, presentation at Advances in Cryptology - CRYPTO'98: 18th Annual International Cryptology Conference (H. Krawczyk ed., Aug. 1998) *available at* http://www.springerlink.com; Camenisch, *infra* note 301.
297 There are also technical methods to provide for untraceable pseudonyms, *see, e.g.,* David Chaum, *A Cryptographic Invention Known as a Blind Signature Permits Numbers to Serve as Electronic Cash or to Replace Conventional Identification*, SCI. AM. 96-101 (Aug. 1992) *available at* http://www.eff.org/Privacy/ Digital_money/?f=chaum_privacy_id.article.txt.
A traceable pseudonym allows for both authorization and accountability applications. An untraceable pseudonym can provide authorization (for example, in digital cash applications) but could not be used for direct accountability. Traceability can be maintained by designing the pseudonym to resolve to their issuers for subsequent retrieval pursuant to approved procedures of the underlying identity.

pseudonym's develop their own reputational attributes and can be tracked over time or across systems.[298]

Pseudonymity allows for the disclosure of only the particular attribute or data relevant (and appropriate) for the particular transaction in which an exchange of data is required. For example, in the use of credit cards, the merchant does not actually require the purchaser's name to complete the transaction – only the authorization that American Express will pay the amount of purchase is relevant to the transaction. (Whether the cardholder is the authorized user is an authentication issue unrelated to the transaction specifically and also does not require revealing a name).[299]

1. PSEUDONYMITY AND SECURITY

Identification systems are generally used for either authorization or accountability (or both). Technical means exist to prove authorization without revealing actually identity by using *third party certification* in which a trusted third party certifies an authorization. The holder of the certificate (digital or otherwise) then presents the certificate to the second party (who may still authenticate that the individual is the authorized user), however, the original party does not have to reveal additional identifiers (or identity) in order for the second party to grant the level of authorization certified by the third party.[300]

Technical means also exist to provide accountability without disclosing identity at the point of verification. *Escrowed identity* is a form of third party certification, in which the trusted third party certifies that they know the "true" identity of the user.[301]

---

298 Clarke, *supra* note 282. Persistence is also sometimes called *linkability* (allowing individual transactions to be linked).

299 *See supra* Part VII.A. (discussing how whether a user is the authorized user could be verified, for example, using a biometric match with the card without revealing individual identity).

300 Secure third party certification can be accomplished through public key infrastructure (PKI) systems. *See generally Public Key Infrastructure*, WIKIPEDIA, *available at* http://en.wikipedia.org/wiki/Pki.

301 *See, e.g.,* Jan Camenisch & Anna Lysyanskaya, *An Identity Escrow Scheme with Appointed Verifiers*, in *Advances in Cryptology – CRYPTO2001*, 2139 *Lecture Notes in Computer Science* 388-407. International Association for Cryptologic Research (Joe Kilian, ed., 2001), *available at*

For example, a pseudonymous driver's license based on smart card technology[302] could be designed to only reveal during a traffic stop that the driver is authorized to drive (for example, by producing a digital certificate from the DMV certifying the holder's authorized status) without revealing a common identifier. The police officer could still run a data match against, for example, a wanted-felon or terrorist watch list (also without revealing name) by reading a hashed identifier keyed (through shared salt) to the felony or watch list database hashing algorithm. If there were an initial data match then additional procedures may or may not be called for, however, without a match, the purpose of the traffic stop could be accomplished without revealing identity.[303] The same card could be designed to only exchange, for example, age information with a bartender's card reader, or health information with a medical worker, etc.

An important policy consideration in any such system is determining whether such pseudonymous encounters are logged – that is, do they generate their own transaction records. If so, do such queries become part of the data record subject to whatever policy controls are envisioned.[304] As noted earlier, who controls the logs has policy implications.[305]

In the example above, in addition to whether there should be a law enforcement database logging the transaction, there is also the question of whether the card itself should be designed to record the encounter. Arguments in favor would emphasize the empowerment to the individual from an immutable record in their possession of their encounter with law enforcement and a specific record of what queries were run. Arguments against might include that the card record itself becomes a vulnerability

http://springerlink.metapress.com/ openurl.asp?genre=article &issn=0302-9743&volume=2139&spage=388.

302 A smart card is any pocket-sized device that contains a processor or microchip that can interact with a reader. *See, e.g.,* Farmer, *supra* note 183 (describing how personal data on Malaysia's smart card chips – designed to replace driver's licenses – are stored in isolated files, each accessible only to authorized readers for that particular data).

303 The summons could be issued against another certificate from a third party certifying that they had the needed identifying information if the driver subsequently needed to be traced.

304 *Cf.* the Fair Credit Reporting Act where queries to a credit report become part of the report and are themselves subject to the Act. *See* note 233 *supra.*

305 *See supra* Part VI.C..

point for privacy – that is, recovery of transaction data from the card itself could be used against the individual at a later date.[306]

Identity management, including specifically the use of anonymous or pseudonymous strategies, is a well-developed research field.[307] It is beyond the scope of this article to fully explore technical issues involved in developing these systems. Nevertheless, the point to be recognized for policy purposes is that these issues are not unique to intelligence or law enforcement use in counterterrorism but ubiquitous to resolving identification issues throughout emerging information infrastructure and systems. The need to provide authentication and accountability without disclosing identity (as we traditionally know it) is fundamental to further development of an effective and efficient networked information-based society.

Pseudonymity gives policy makers an additional method to control disclosure of identity. For example, in *Hiibel v. Nevada*,,[308] the issue was whether a suspect could be compelled to give their name during a *Terry* stop.[309] Among the arguments put forward against disclosure of name was that in today's database world, disclosing one's name is the key to unlocking the digital dossier and may lead to "an extensive fishing trip across government databases."[310] One of the arguments in favor of disclosure is the legitimate interest to determine whether the individual is wanted or dangerous. "Obtaining a suspect's name in the course of a *Terry* stop serves important government interests. Knowledge of identity may inform an officer that a

306 Evidence is already being collected from E-Zpass, Metrocard, and cell tower records, see e.g., Tresa Baldas, *High Tech Evidence*, THE NAT'L L.J. (Aug. 16, 2004) *available at* http://www.law.com/jsp/article.jsp?id=1090180376956; Adam L. Penenberg, *The Surveillance Society*, WIRED MAG. Dec. 2001, *available at* http://www.wired.com/wired/archive/9.12/surveillance_pr.html.

307 *See generally* Liberty Alliance Web Site *at* http://www.projectliberty.org/. Identity management is generally concerned with authentication of identity, and authorization and accountability for (or non-repudiation of) behavior within or across systems.

308 Hiibel v. Nevada, 124 S.Ct. 2451 (2004).

309 *See* Terry v. Ohio, 392 U.S. 1 (1968) (holding that police can detain a suspect for a reasonable period without reasonable cause to suspect a crime).

310 *See* Brief of Amice Curiae of Electronic Privacy Information Center (EPIC) and Legal Scholars and Technical Experts at 6, Hiibel v. Nevada, 124 S.Ct. 2451, *available at* http://www.epic.org/privacy/hiibel/epic_amicus.pdf.

suspect is wanted for another offense, or has a record of violence or mental disorder."[311]

Information systems based on pseudonymity, including the use of smart ID cards, can provide another alternative to meet these same needs. As noted in the example above, there are technical methods for an individual to be matched against a watch list (or any other list) without revealing explicit identifying data. Thus, development of a national ID card based on segmented data and pseudonymous identities could improve privacy over existing methods and still meet security needs.[312]

The Transportation Safety Administration has recently begun testing of a program in which *registered travelers* (sometimes also referred to as *trusted travelers*) are not subject to additional random checks based on having previously submitted to a background check.[313] The privacy lobby has opposed this program on a variety of grounds.[314] Nevertheless, the same system could be employed while still providing additional privacy protections by employing pseudonymous strategies, where individually identifying data would not need to be revealed at the point of verification but watch list matching could still occur.[315]

2. DEVELOPMENT IMPERATIVE

Pseudonymous technologies also have a development imperative separate from privacy and related to security generally. For example, developing pseudonymous transaction technologies and implementation architectures will be needed to enable secure online voting. Secure online voting will require maintaining ballot (user identity) secrecy but authenticating the right to vote (and preventing multiple votes). Additionally, various online payment systems[316] and *federated identity*

311 *Terry,* 392 U.S. at 7.
312 An example of such a card is described in Farmer, *supra* note 183. *But see* Froomkin, *supra* note 146; Sobel, *supra* note 146 (discussing issues relating to a national ID card).
313 *See* note 71*supra.*
314 *Id.*
315 Only hashed identifiers are matched, and authorization for travel authenticated.
316 For example, American Express Blue, Amazon Marketplace, and PayPal each use forms of third party authorization that could be adapted to enable

management systems[317] architectures could be adapted to enable pseudonymous transactions and access to data.

IX. TOWARDS A CALCULUS OF REASONABLENESS.

Assuming that anonymization and pseudonymization strategies are employed to separate identity from behavior (or data), and control over data attribution is enforced through privacy appliances, the policy issue still remains when and under what circumstances particular methods of inquiry might be used on specific data sets, and when and under what circumstances data attribution may occur. This section examines some of the variables that need to be considered.

It is not my intent in this section to recommend specific confidence intervals, predicates or oversight regimes for use of identification, data aggregation and automated analysis, or collection technologies in any particular application or by any particular agency but only to illustrate the interaction among certain relationships and issues that may be relevant in devising procedural guidelines, technical standards, or oversight structures in any particular context.

It is the general thesis of this section that procedural mechanisms relate to the concept of *reasonableness* (both in Fourth Amendment terms and as that term is more generally understood) through a complex policy calculus involving multiple independent and dependent variables that must be understood individually but considered together and in context.[318] Thus, guiding principles, not rigid standards to be determined *a priori* for every conceivable use, condition or context, must be derived within which specific administrative procedures, legislative oversight, and judicial intervention and review can be fashioned.

---

pseudonymous transactions. Additionally, research on so-called digital cash is relevant. *See, e.g.*, Chaum, *supra* note 297; Tatsuo Tanaka, *Possible Economic Consequences of Digital Cash*, 1 FIRST MONDAY 2 (Aug 5, 1996) *at* http://www.firstmonday.dk/ issues/issue2/digital_cash/

317 Federated identity refers to the use of a single authentication of identity to suffice for access across multiple trusted systems.

318 *Cf.* United States v. Cortez, 449 U.S. 411, 417 (1981) ("the whole picture -- must be taken into account").

There is some policy function, *f*, where reasonableness is a derivative of confidence interval, predicate, consequence, and procedure for error correction, (together, "due process") and scope of access, sensitivity of data, and method of query ("privacy/security" trade-off) as they relate to threat ("threat"). Policy guidelines are required to define the limits of and express the relationship among the due process, privacy and security, and threat variables.

f reasonableness = due process ~ privacy|security ~ threat

This is not intended to imply that there exists a formulaic policy outcome that can be pre-determined and simplistically applied in any circumstance. Rather, the construct of such a theoretical equation is used to illustrate the relationship among the issues to be discussed in this section.

A. DUE PROCESS

Due process is the means for ensuring fairness in a system.[319] Due process is essentially a function of four factors: the reasonableness of the *predicate* for action, the *practicality* of alternatives, the *severity and consequences* of the intrusion, and the procedures for *error control*.[320]

1. PREDICATE

Determining the appropriateness of predicate requires understanding error rates and assessing related confidence intervals – that is, it requires determining the probative weight of the indices of suspicion.[321] Confidence interval for policy purposes is simply the acceptable error rate for a given application.[322] As discussed above, for example, the confidence interval for a screening application can be viewed as a function of two competing relationships – the number of false positives (innocents identified) adjusted by the severity of the

319 *See generally* RONALD DWORKIN, LAW'S EMPIRE (1986).
320 *Cf. id.*
321 *See supra* notes 214-223 and accompanying text.
322 *See* Taipale, *supra* note 13, at n.104 discussing technical aspects of confidence intervals for decision making in knowledge discovery systems.

consequences to the individual and the number of the false negatives (terrorists not identified) adjusted by the consequences to security (and by the potential misallocation of resources from false positives).[323] Determining acceptable confidence intervals for any particular application requires assessing the probative value of the predicate procedures – for example, determining whether the observed behavior (or data analysis) reasonably justifies the consequences of the action or not.[324]

2. PRACTICAL ALTERNATIVES

The Supreme Court has also explicitly recognized that the requirement for individualized suspicion is not an "irreducible requirement"[325] and the *practicality* of requiring either a warrant or individualized suspicion needs to be considered. In *Treasury Employees v. Von Raab*,[326] the Court noted that where the government interest serves a need beyond routine law enforcement, the practicality of requiring individualized suspicion is also a relevant factor:

> [O]ur cases establish that where a Fourth Amendment intrusion serves special governmental needs, beyond the normal need for law enforcement, it is necessary to balance the individual's privacy expectations against the Government's interests to determine whether it is *impractical* to require a warrant or some level of individualized suspicion in the particular context.[327] (emphasis added)

The Court has used the same special-needs reasoning in upholding the use of sobriety checkpoints,[328] roving border checkpoints,[329] and random drug testing of student athletes.[330] Likewise, policy makers should consider the practicality (or impracticality) of requiring specific procedures or individualized

323 *See supra* Part VI.A.
324 *See supra* notes 225-230 and accompanying text.
325 *See* United States v. Martinez-Fuerte, 428 U.S. 543, 561 (1976).
326 489 U.S. 656, 666 (1989).
327 *Id.* At 665-66
328 Mich. Dep't of State Police v. Sitz, 496 U.S. 444 (1990).
329 United States v. Brignoni-Ponce, 422 U.S. 873 (1975).
330 Vernonia v. Acton, 515 U.S. 646 (1995).

predicate in cases of information processing systems for use in counter-terrorism.[331]

3. SEVERITY AND CONSEQUENCES OF INTRUSION

Another important factor to be considered is the reasonableness of the intrusion. "The reasonableness of a seizure under the Fourth Amendment is determined by balancing its intrusion on the individual's Fourth Amendment interests against its promotion of legitimate government interests."[332] Thus, where there is an important state interest, and the intrusion minimal and the consequences slight – for example, a brief stop and referral to a secondary inspection or minimal questioning – the courts are likely to find no Fourth Amendment violation.[333] In upholding roving traffic checkpoints in *Brignoni-Ponce*, the Court stated:

> Against this valid public interest we must weigh the interference with individual liberty that results when an officer stops an automobile and questions its occupants. The intrusion is modest. The Government tells us that a stop by a roving patrol "usually consumes no more than a minute." There is no search of the vehicle or its occupants, and the visual inspection is limited to those parts of the vehicle that can be seen by anyone standing alongside. According to the Government, "[a]ll that is required of the vehicle's occupants is a response to a brief question or two and possibly the production of a document evidencing a right to be in the United States."[334]

Thus, a legitimate inquiry for policy makers is to determine the severity and consequence of a particular intrusion

331 Compare Separate Statement of William T. Coleman in TAPAC Report, *supra* note 90, at 67, *with* Separate Statement of Floyd Abrams, *id.* at 63. Coleman argues that requiring certain predicate authorizations and procedures is impractical for use in automated information systems in the war on terror.
332 *Hiibel*, 124 S.Ct. 298, citing Delaware v. Prouse, 440 U. S. 648, 654 (1979).
333 *See* Martinez-Fuerte, 428 U.S. at 558.
334 Brignoni-Ponce, 422 U.S. at 879-880 (citations omitted).

in light of the state interest.[335] Where there is a significant state interest (for example, preempting terrorist attacks), minimal initial intrusion (for example, an automated review of data), and limited consequences (for example, a routine investigative follow-up that may only include cross-checking against additional data), the courts are likely to uphold the use of advanced information systems to screen data.[336]

4. ERROR CORRECTION

Reasonableness in this context also requires examining procedures for error detection and correction. Determining confidence levels means recognizing the potential for errors. "No system constructed by man is perfect. The only certainty is that there will be false positives — both in investigations and possibly (though less likely) in the mistaken imposition of collateral consequences on a misidentified subject."[337] Thus, an integral part of policy as well as system design is to recognize the potential for error and build in error detection and error correction procedures.[338]

Error, however, including falsely identifying suspects (false positives), is not unique to information systems. Thus, to the extent possible, error detection and error correction procedures for automated systems should embrace existing procedures.[339]

Additional complications arise, however, when one considers systems in which the subject may never become aware of the intrusion or the consequences of the query. In access control situations where permission for access is denied, the individual is on notice that their autonomy has been affected and corrective procedures can be triggered. More difficult is the

335 See Rosenzweig, *supra* note 71, at 677-85.
336 Note that in the context of an arrest, the Supreme Court has repeatedly held that even the drawing of blood constitutes only a minimally intrusive search, see Skinner v. Railway Labor Executives Ass'n, 489 U.S.602, 625 (1989) (blood tests do not "infringe significant privacy interests"); Winston v. Lee, 470 U.S. 753, 62 (1985) (not "an unduly extensive imposition"); Schmerber v. California, 384 U.S. 757, 771 (1966) ("commonplace"); Breithaupt v. Abram, 352 U.S. 432, 36 (1957) ("routine" and "would not be considered offensive by even the most delicate").
337 Rosenzweig, *supra* note 13, at 191-95.
338 *See id. Cf. supra* note 166 (discussing problems with watch lists).
339 Rosenzweig, *supra* note 13, at 191-195.

situation where the data subject never becomes aware of the query or its consequences.[340]

Also unresolved is whether *derived data* (that is, data generated from the query or analysis process) or other *meta-data* (data about the data that is attached to the data, for example, data labels) becomes part of the record and whether it too becomes subject to applicable laws that the underlying data may be subject to.[341]

B. PRIVACY AND SECURITY INFORMATION NEEDS

This section briefly examines the relationship between privacy and security information needs. In particular, it describes how *scope*, *sensitivity* and *method of query* implicate certain privacy and security considerations.

1. SCOPE OF ACCESS

Obviously, foreign intelligence, counter intelligence and law enforcement information is and should be available for appropriate domestic security purposes. The policy challenge lies in determining when and under what circumstance domestic intelligence access should be allowed to routinely-collected government data or to commercially available (or other third party) data.[342]

340 For example, a job applicant may never know they were denied a job because they were on a watch list.

341 Compare the treatment of private credit reports, where inquiries themselves become part of the report and subject to the Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681 et seq. Even national security investigations are subject to disclosure under the Act once the investigation is concluded. 15 U.S.C. § 1681b FCRA §604(b)(4)(B).

342 *See generally* Solove, *supra* note 63; Second Markle Report, *supra* note 23, at 30-37.

Note that privately held data (i.e., data not generally available to third parties or where access is restricted by statute), which is the most sensitive and requires the greatest protection under existing doctrine, is already protected to some extent under the procedural due process regimes applicable to methods of its collection, for example, judicial and statutory rules and procedures required for use of wire taps (Title III (governing electronic surveillance), 18 U.S.C. §§ 2510-2522 et seq. (2003), as amended by the USA PATRIOT Act, Pub. L. No. 107-56 (2001)), for accessing stored electronic data (Electronic Communications Privacy Act (governing access to stored

Routinely collected data – that is, government held data collected in the normal course of providing government services – is generally subject to restriction for other uses or data matching by the Privacy Act of 1974.[343] However, the Privacy Act has broad exceptions for data matching and inter-agency sharing for national security and law enforcement purposes,[344] thus, for practical purposes there are no restrictions on use for domestic security applications.[345] One policy question is whether there should be any additional procedural protections or guidelines for use of routinely-collected government data that subsequently come within the national security and law enforcement exceptions.

The more difficult question, however, involves deciding whether government should have access to, and use of, privately held third party data, particularly data from commercial data sources, and, if so, under what circumstances and what constraints.[346]

That the government should, and will ultimately, have access to this data seems foregone.[347] As already noted, it would

communications), 18 U.S.C. §§ 2701-2712 (2003), as amended by the USA PATRIOT Act), or for searching personal computer drives, storage media, or other physical assets. Once lawfully acquired under the appropriate procedures, this previously privately held data is then either foreign or counter-intelligence data (for example, if it is collected under Foreign Intelligence Surveillance Act of 1978, Pub. L. No. 95-511, 92 Stat. 1783 (providing a separate regime for "foreign intelligence"), 50 U.S.C. §§ 1801-1811, as amended by the USA PATRIOT Act) or law enforcement data (for example, if it is collected pursuant to a lawful warrant).

343 Privacy Act of 1974, 5 U.S.C. § 552a, as amended. The Privacy Act also contains certain procedural restriction on "matching" information from several government databases and for sharing data among agencies, requiring certain inter-agency agreements. *See* The Computer Matching and Privacy Protection Act of 1988, Pub L. No. 100-503, § 1, 102 Stat. 2507 (1988) (appears as a note amending the Privacy Act in 5 U.S.C. § 552a (2003).

344 The Privacy Act contains exemptions for both computer matching and for inter-agency data sharing for national security and law enforcement purposes. *See* 5 U.S.C. §§ 552a(b)(7), 552a(a)(8)(B)(vi), 552a(j) (2003).

345 *But see* Sean Fogarty & Daniel R. Ortiz, Limitations Upon Interagency Information Sharing: The Privacy Act of 1974, *in First Markle Report*, *supra* note 21, at 127-132.

346 *See* Second Markle Report, *supra* note 23, at 30-37; Solove, *supra* note 63.

347 *See generally* Federal Bureau of Investigation, Guidance Regarding the Use of ChoicePoint for Foreign Intelligence Collection or Foreign Counterintelligence Investigations (Sep. 17, 2001) *available at* http://www.epic.org/privacy/publicrecords/cpfcimemo.pdf; *see also* Solove,

be an unusual polity that demanded accountability from its representatives to prevent terrorist acts yet denied them access to available tools or information. Thus, it is the procedures under which such access should be allowed that need to be defined.[348] Here, developing clear goals and concomitant policy guidelines, requiring that the nexus between particular types of information and its use and value for counter-terrorism be clearly articulated,[349] and mandating strict oversight and review procedures, are needed to ensure that appropriate government access to potentially useful information is possible in a way that protects civil liberties.

Among the policy tools for dealing with access questions is the use of *categorization* to designate certain information sources or types subject to (or exempt from) particular procedures. For example, under Homeland Security Presidential Directive (HSPD-6) certain information is to be classified as "Terrorist Information" and provided to TTIC.[350]

---

supra note 63, at 1089 ("criminal investigations often require the gathering of data from third parties"); Second Markle Report, *supra* note 23, at 31-2 ("We … start from the premise that government must have access to [private sector data]").

348 *See* Solove, *supra* note 63, at 1151-67 (outlining an architecture of power based on a constitutional and statutory framework).

349 *See* Second Markle Report, *supra* note 23, at 31-32.

350 Homeland Security Presidential Directive/HSPD-6, 39 Weekly Comp. Pres. Doc. 1234-1235 (Sept. 16, 2003) (outlining procedures for integrating information about individuals who are known or suspected terrorists within the Terrorist Threat Integration Center ("TTIC"), the all-source intelligence fusion and analysis center announced by the President in January 2003. *See* White House Fact Sheet: Strengthening Intelligence to Protect America, *available at* http://www.whitehouse.gov/news/releases/2003/01/20030128-12.html. TTIC's role and responsibilities are set out in the classified Director of Central Intelligence Directive (DCID) 2/4 (effective May 1, 2003)). *See also* CONGRESSIONAL RESEARCH SERVICE, TERRORIST IDENTIFICATION, SCREENING AND TRACKING UNDER HOMELAND SECURITY PRESIDENTIAL DIRECTIVE 6 (Congressional Research Service 2004); Daniel Gallington, *The New Presidential Directive on "Screening" Terrorist Information,* Potomac Institute for Policy Studies, Waypoint Issue Paper (Oct. 6, 2003); Second Markle Report, *supra* note 23, at 19 (suggesting that HSPD-6 and TTIC may have "radically changed the balance of liberties" without "significant public debate on this fundamental question [*i.e.,* maintaining the U.S. person distinction]");
But cf. Daniel Gallington, *Better Information Sharing and More Privacy in the War on Terrorism – A New Category of Information is Needed*, Potomac Institute for Policy Studies, Waypoint Issue Paper (Jul. 29, 2003), *available at* http://www.potomacinstitute.org/research/072903_project_guardian.cfm;

Under the Executive Order for Sharing Terrorism Information certain requirements and procedures are to be applied to "terrorism information."[351] The Second Markle Report suggests that government should take steps to "concretely identify its true information needs" by identifying what private sector information is needed for "the government to carry out its homeland security responsibilities."[352] Additionally, it has been suggested that existing categorization procedures, for example, the classified procedures used in managing SIGINT, could be adopted for commercial data space.[353] However, all of these procedures require pre-determining what information is relevant – something that may not always be possible in counter-terrorism analysis.

Further, certain existing categorizations may not be appropriate for these new circumstances. For example, the procedures used under USSID-18 to manage SIGINT require additional exceptional procedures for handling information relating to U.S. persons (i.e., minimization).[354] This same arbitrary designation (arbitrary in the sense that it is a legal

---

Potomac Institute for Policy Studies, Oversight of Terrorist Threat Information: A Proposal (June 25, 2003), *available at* http://www.sainc.com/tapac/ library/sept29/Guardian_Proposal_0703.pdf

351 Exec. Order No. 13,356 (Aug. 27, 2004) 69 Fed Reg. 53599 (Sep. 1, 2004), at§ 6(d).

> The term "terrorism information" means all information, whether collected, produced, or distributed by intelligence, law enforcement, military, homeland security, or other United States Government activities, relating to (i) the existence, organization, capabilities, plans, intentions, vulnerabilities, means of finance or material support, or activities of foreign or international terrorist groups or individuals, or of domestic groups or individuals involved in transnational terrorism; (ii) threats posed by such groups or individuals to the United States, United States persons, or United States interests, or to those of other nations; (iii) communications of or by such groups or individuals; or (iv) information relating to groups or individuals reasonably believed to be assisting or associated with such groups or individuals.

352 Second Markle Report, *supra* note 23, at 31-32.

353 *Securing Freedom and the Nation: Collecting Intelligence Under the Law Before the House Permanent Select Committee on Intelligence*, 108th Cong. 41 (2003) (statement of Daniel Gallington, Senior Fellow, Potomac Institute for Policy Studies).

354 NSA/CSS United States Signal Intelligence Directive 18 ("USSID 18") (July 27, 1993), *available at* http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB23/07-01.htm.

categorization, not an attribute of the data) relating to nationality has also been used in various Congressional acts purporting to limit the use of certain technologies or techniques to non-U.S. citizens only.[355] The problem with these approaches arises when one considers the information space and data that is now available for analysis.

Historically, information in data sets collected in foreign intelligence operations related primarily to foreigners and U.S. data could be handled by exception. The information data sets that are being contemplated here – civilian transaction space – are essentially U.S. person-centric (or co-mingle data in such a way that makes it technically difficult, if not impossible, to separate out U.S. person data for handling as an exception) and therefore require the development of procedures based on a general rule, not by reference to the exception.

Note that the Second Markle Report also concludes that the "distinguishing line between domestic and foreign threats is increasingly difficult to sustain ... [requiring] new rules – rules to replace the old 'line at the border' . . . for information collection and use."[356]

Thus, any policy based on procedures for pre-designating information as relevant for counter-terrorism needs to address these issues: first, recognizing that some information may only be identifiable as relevant in the context of an ongoing investigation or in response to a particular threat not a priori;[357] second, that designations based on place, method of collection or nationality of subject may be outmoded in the context of a worldwide, distributed, networked database environment, thus, requiring more flexible standards based on or anticipating changing data mixes and different circumstances of initial collection; and, third, that classifying data into categories that do not relate to the purpose of the original data collection may

355 *See* Taipale, *supra* note 13, at n.28.
356 Second Markle Report, *supra* note 23, at 18.
357 Any rules limiting analysis or access to particular information should recognize that information may only become relevant during the process of analysis or investigation and should therefore contain "hot pursuit" exceptions or procedures.

not be possible post collection.[358] In any case, use of categorization or other designation requires that some technical means for data-labeling be built into systems.[359]

With respect to the U.S. person problem specifically, to the extent that technical means can be used to protect privacy they may need to be applied to all data – thus affording the highest protection to everyone – with subsequent identification of foreign or terrorist related information being treated as the exception, subject to procedures that selectively reveal additional information, including identity, based on an iterative analysis that increases the particular suspicion, thus predicate, with each pass. Privacy then becomes the norm (protected for everyone through data anonymization) rather than being something exceptional granted to particular data categories (for example, minimization of U.S. persons post collection or processing); and disclosure (determined by policy and controlled through selective revelation or selective attribution by privacy appliances) becomes the exception subject to appropriate authorization or due process procedures.[360]

2. SENSITIVITY OF DATA

Specific statutes already exist that protect particular classes of information deemed sensitive. These statutes generally require that use of these types of information conform to particular procedures. For example, census data, medical records, educational records, tax returns, cable television records, video rental, etc. are all subject to their own statutory protection, usually requiring an elevated level of predicate, for example, a warrant or court order based on probable cause instead of a subpoena based on mere suspicion, to gain access.[361]

358 For example, if nationality is not required for the transaction that generates the data, it may not be possible after the fact to determine if it is U.S. person data or not.

359 *See* Taipale, *supra* note 13, at n.76-77 and accompanying text.

360 Making privacy the default state might also help eliminate problems in sharing data with other jurisdictions. *Cf. e.g.,* Ryan Singel, *EU Travel Privacy Battle Heats Up*, WIRED NEWS (Dec. 22, 2003).

361 For example, U.S. Census data is protected under 13 U.S.C. § 9 (2003); certain medical records collected for research purposes under 42 U.S.C. § 242(m) (2003); educational records under 20 U.S.C. § 1232(g) (2003) (but see USA PATRIOT Act amendments, 20 U.S.C. 1232(g)-(j)); and tax records under The Tax Reform Act of 1976, Pub. L. No. 94-455, 90 Stat. 1590 (1976). With respect to state governments, the Driver's Privacy Protection Act of

Although some of these designations may need review in the context of domestic security, the general approach of dealing with particularly sensitive personal data by providing additional procedural protections is sound and can be applied to identification, data aggregation and analysis, and collection systems. Technical features, for example, data labeling discussed below, need to be developed to enable data categorization to be maintained when data is shared or exchanged.[362]

Policy decisions to designate higher standards for sensitive information or declaring certain information "off-limits" involve the same considerations that are necessary to pre-designate certain information or sources as relevant for counter-terrorism analysis. Additionally, there may be a trade-off between the sensitivity of information and its relevance for counter-terrorism that must be taken into account when such designations are made. If it turns out that certain information deemed sensitive by its nature, for example, financial records, is also quite specifically useful for counter-terrorism (for example, following the money trail), choice of policies (and technical features to support such choice) need to be developed taking both needs into account.[363]

1994, 18 U.S.C. § 2721 (2003), regulates the use and disclosure of personal information from state motor vehicle records. There is a broad exemption for use by any government agency, including law enforcement, for use in carrying out its functions. There are also a number of sector specific laws restricting the collection, use or disclosure of personal information by private sources. Among these, the Fair Credit Reporting Act, 15 U.S.C. §1681-1681v (2003), regulates the use of information by credit reporting agencies, the Video Privacy Protection Act of 1988, 18 U.S.C. § 2710 (2003), prohibits the disclosure of video rental records, the Cable Communications Policy Act of 1984, 47 U.S.C. § 551 (2003), limits disclosure of cable television subscriber data, and the Telecommunications Act of 1996, 47 U.S.C. § 222 (2003), limits the use and disclosure of customer proprietary network information. Additionally, individually identifiable health information is protected by the Standards for the Privacy of Individually Identifiable Health Information, 45 C.F.R. §§ 160, 164 (2003) pursuant to the Health Insurance Portability and Accountability Act of 1996, 42 U.S.C. § 1320(d) (Dept. of Health and Human Services 2003).

362 *See* Taipale, *supra* note 13, at n.76-77 and accompanying text (describing data labeling technologies).

363 Reconciling competing business process needs is fundamental to designing appropriate systems architectures.

3. METHOD OF QUERY

As previously noted, there is no absolute constitutional requirement for individualized suspicion,[364] and no inherent or presumptive constitutional problem with pattern-matching.[365] Nevertheless, for purposes of determining policy, it may be appropriate to recognize that different procedures may be appropriate for different query methods depending, for example, on whether they are subject-, link- or pattern-based.

Subject- and link-based queries generally raise the same issues as outlined above in the general discussion of scope – that is, what data can be accessed and under what circumstance. For some, pattern-matching may or may not also raise additional issues relating to its particularization depending on its efficacy in any particular application.[366]

Thus, there may be a legitimate policy question as to whether there should be additional technical or procedural protections applied for pattern-based queries based on the *perception* that these methods are more intrusive. Some have recommended that such additional procedures be used for processes using pattern-analysis derived from "data-mining."[367] If such policies are desired, technical features to support implementation, such as methods for selective revelation, anonymizing data, and pseudonymizing identity to ameliorate concerns related to non-particularized scrutiny of personal data, will be required.[368] These technical features would allow for judicial, administrative, or other procedural intervention before disclosure (or attribution) of identity or other personal information and would thus help protect individual autonomy through legal due process.

4. SUMMARY: SCOPE, METHOD AND SENSITIVITY

There is no magic policy formulation that perfectly balances the variables discussed in this section. What is

364 *United States v. Martinez-Fuerte*, 428 U.S. 543, 561 (1976); *see also* Part VII.B.2.
365 *See* Part VII.B.2 *supra.*
366 *Id.*
367 TAPAC Report, *supra* note 90; Rosenzweig, *supra* note 13.
368 These methods are discussed in this Article and in Taipale, *supra* note 13, at n.74-81 and accompanying text.

required then is an analytic framework, together with guiding principles, that can inform the public debate as these issues come up in varying contexts as new technologies develop and challenge existing doctrine or precepts. Judicious distinction between when rules (*what* you can do), procedures (*how* you can do something) and guidelines (constraints or *limits* within which you act to accomplish some goal) are appropriate requires understanding the dynamic character of the problem and the complex nature of the variables, as well as recognizing the inchoate nature of any solutions given the rapid pace of technological development and the evolving nature of the threat.

Further, these complexities and the wide variety of divergent interests involved and views expressed highlights the need for policy makers and technology systems developers to develop a shared common language for policy needs and technical requirements.

C. THREAT ENVIRONMENT AND REASONABLENESS

Reasonableness (including acceptable error rates) may also vary depending on the threat level and the particular security need. System bias towards more false positives and less false negatives may be appropriate (and reasonable) under certain high threat conditions or in applications requiring high security.[369] In other circumstances, system bias towards fewer false positives and more false negatives may be appropriate.

Policy considerations are also domain dependent. For example, decision heuristics used for the development of traditional defense systems are generally inappropriate for domestic security applications. In designing military defenses, the bias is to eliminate any false negatives by accepting additional false positives. On the battlefield, it is better to have a low threshold for triggering a response, say donning a gas mask, then to risk not being prepared. However, in the context of a civilian population, false positives may be as destructive of certain values (including security) as are false negatives by undermining trust in the system or creating intolerable burdens. Too many false positives and the resulting misallocation of

369 *Cf.* Rosenzweig, *supra* note 71, at 677-85.

resources will undermine both popular and political support for security measures as well as impact security itself.[370]

Thus, because of the dynamic nature of the threat and the changing security requirements, no system (technical or procedural) should be contemplated that is either constantly at ease or constantly at general quarters. Flexible systems and policy guidelines that can adapt proportionally to perceived threats faster and more efficiently are required.

At the same time, it seems premature to burden either policy development or technical research and development with a requirement to determine *a priori* what policy rules will apply in every conceivable case.[371] Technical development processes are not generally amenable to predictable development paths where ongoing research is in its early stages. An iterative process using value sensitive design procedures can help guide technical and policy development to achieve both required outcomes – security and privacy. However, achieving this outcome requires joint participation, not knee-jerk opposition.

Nevertheless, guiding policy principles can be developed even without knowing all the potential technologically enabled opportunities or constraints based on a deeper understanding of business process needs and how policy and technical features interact. Policy then develops rules of general applicability that are supported by technical architecture, while judicial review examines cases of specific application according to traditional notions of due process.

X. CONCLUSION

New technologies do not determine human fates; rather, they alter the spectrum of potentialities within which people

370 *See generally* Taipale, *Losing the War*, *supra* note 25; and *supra* notes 25 and 29.

371 *Cf.* ACLU, TOTAL INFORMATION COMPLIANCE: THE TIA'S BURDEN UNDER THE WYDEN AMENDMENT, 6 (2003), *available at* http://www.aclu.org/SafeandFree/SafeandFree.cfm?ID=12650&c=206 (suggesting that rules and technical capabilities need to be determined prior to research and development efforts that are intended specifically to determine feasibility and required rules).

act.[372] Advanced information technologies have the potential to help allocate domestic security and law enforcement resources more effectively. In particular, developing certain technical architectures may preclude opportunity for certain crimes to take place in the first place[373] and other technologies may enable preemptive allocation of resources to prevent future occurrences of crime.[374]

A. BUILDING IN TECHNICAL CONSTRAINTS

It is the premise of this article that disaggregating privacy into identity and behavior for analytic purposes, and designing technical systems to help manage the circumstances of attribution, can help achieve a practical resolution to the apparent conflict between privacy-security interests.

This Article has argued that *anonymization* and *pseudonymization* strategies designed to control data attribution at the *privacy divide* can significantly mitigate privacy concerns in the context of certain domestic security and law enforcement applications by enabling existing legal doctrines and related procedures to function within technical systems.

Systems requirements (and related technologies) to support these strategies include:[375]

- rule-based processing and a distributed database architecture, which can limit the scope of inquiry and the subsequent use of data within policy guidelines;

- selective revelation, which can be used to control the attribution of observed behavior and identity (or identity with behavior or data); and

372 ROBERT MCCLINTOCK & K. A. TAIPALE, EDUCATING AMERICA FOR THE 21ST CENTURY, INSTITUTE FOR LEARNING TECHNOLOGIES 2 (1994), *available at* http://www.taipale.org/ilt/ILTplan.html.
373 *See also* Katyal, *Digital Architecture, supra* note 137; Taipale, *supra* note 137.
374 *See generally* Taipale, *supra* note 13; see *also* Robert Popp, *et al., Countering Terrorism Through Information Technology*, 47 COMM. OF THE ACM 36 (Mar. 2004).
375 Taipale, *supra* note 13, at 74.

- strong credential and audit features, and diversified authorization and oversight, which can make misuse and abuse "difficult to achieve and easy to uncover."[376]

Rule-based processing technologies include the use of intelligent agents to query distributed databases according to pre-determined rules, and data labeling to ensure appropriate processing when data is exchanged and include research in proof carrying code, data labeling (DRM), and analytic filtering tools.[377]

Selective revelation technologies include research in entity resolution, searching on encrypted data, and one-way hashing technologies.[378] Strong credential and auditing requires developing immutable logging and self-reporting data.[379] Additional development requirements include a common language for expressing privacy and other policy rules across systems, general computer and network security, user authentication, encryption and compliance checking and reporting technologies.[380]

Further, this article contends that developing these features for use in domestic security applications will lead to significant opportunities to enhance overall privacy protection more broadly in the U.S. (and elsewhere) by making these technical procedures and supporting features available for voluntary or legislated adoption in the private sector. In addition, the development of these technologies will have significant beneficial "spill-over" uses for commercial and scientific applications, including improved information infrastructure security (better user authentication, encryption, and network security), protection of intellectual property (through rule-based processing), and the reduction or elimination of spam (through improved analytic filtering). Other economic sectors that stand to benefit from developments in these technologies include bioinformatics and pharmaceutical design, medical research, corporate knowledge management, environmental resource management, basic science, and others

376 Rosenzweig, *supra* note 13, at 196-97.
377 Taipale, *supra* note 13, at 75-78.
378 Taipale, *supra* note 13, at 79-80.
379 Taipale, *supra* note 13, at 80.
380 Taipale, *supra* note 13, at 81.

requiring the management of vast data volumes that may or may not include sensitive data.

B. OVERRIDING PRINCIPLES

Development and use by government of advanced identification, aggregation and analysis, and collection technologies in domestic security applications raise legitimate privacy and related civil liberties concerns. Nevertheless, such development and use is inevitable and strategies premised on opposition to research or banning certain uses or deployments through law alone are destined to fail, and, in any case, provide little security and brittle privacy protection. Protecting civil liberties in an information-based society requires that value sensitive development strategies be used to design technical systems that include features that enable familiar due process mechanisms and related procedures to function, in particular by providing intervention before attribution of identity with data (or data with identity) occurs.

This article proffers certain guiding principles for the development and use of these technologies, particular in the context of their use in preemptive counterterrorism applications:

First, automated predictive or screening technologies should be used only as investigative, and not evidentiary, tools. That is, they should be used as predicates for further investigation and not to provide proof of guilt. Moreover, their use should be restricted to investigations of activities about which there is a political consensus that aggressive preventative strategies are appropriate (for example, in the realms of counter-terrorism and national security).

Second, specific implementations should be subject to congressional authorization, oversight and review. Executive agencies that employ these technologies should adopt appropriate administrative procedures to control their use. And, judicial review should be available in accordance with existing due process doctrines.

Third, specific technical features should be developed and built into the technologies to protect privacy by providing opportunities for existing doctrines of due process and related

procedures to function effectively. These features include rule-based processing, selective revelation, and secure credentialing and tamper-proof audit functions.

C. IN SUM

Reconciling competing requirements for security and privacy requires an informed debate in which the nature of the problem is better understood in the context of the interests at stake, the technologies at hand for resolution, and the existing resource constraints. Key to resolving these issues is designing a policy and information architecture that can function together to achieve both outcomes, and is flexible and resilient enough to adapt to the rapid pace of technological development and the evolving nature of the threat.

XI. FINALE

With familiarity, Frankenstein's monster is no longer as frightening as when he first appears.[381] So too, technology's potential to protect civil liberties and security should be considered equally with its potential for harm. The mythology of privacy built on absolute secrecy should not keep us from considering opportunities to improve both security and privacy in a world of changing base conditions. The early Luddites were killed or shipped to Australia to little effect, while later movements in their name built collaborative institutions to control new technologies. To ensure security with liberty we should learn from their example.

381 *Cf.* Jay Stanley & Barry Steinhardt, ACLU, *Bigger Monster, Weaker Chains: The Growth of an American Surveillance Society* (2003).