SENSITIVE INFORMATION

PAUL OHM*

ABSTRACT

Almost every information privacy law provides special protection for certain categories of "sensitive information," such as health, sex, or financial information. Even though this approach is widespread, the concept of sensitive information is woefully undertheorized. What is it about these categories that deserves special protection? This Article offers an extended examination of this question. It surveys dozens of laws and regulations to develop a multi-factor test for sensitivity.

From this survey, the Article concludes that sensitive information is connected to privacy harms affecting individuals. Consistent with this, at least for the case of privacy in large databases, it recommends a new "threat modeling" approach to assessing the risk of harm in privacy law, borrowing from the computer security literature. Applying this approach, it concludes that we should create new laws recognizing the sensitivity of currently unprotected forms of information—most importantly, geolocation and some forms of metadata—because they present significant risk of privacy harm.

TABLE OF CONTENTS

* Professor, Georgetown University Law Center, and Faculty Director, Center on Privacy and Technology. I thank workshop participants at the Privacy Law Scholars Conference and the Georgetown University Law Center. Thanks in particular to Danielle Citron, Julie Cohen, Allan Friedman, Dorothy Glancy, James Grimmelmann, Deven McGraw, Helen Nissenbaum, Joel Reidenberg, Stuart Shapiro, Peter Swire, and Dan Solove. Thanks also to Chelsea Brooks, Arielle Brown, and Kurtis Zinger for their fine research assistance.

INTRODUCTION

Privacy law seems stuck. Congress has not enacted any meaningful new privacy laws or overhauled any old laws in at least a decade, with one important but narrow exception relating to genetic information.[1] Courts evolve the common law of privacy very slowly, for example, interpreting the now venerable privacy torts essentially as William Prosser first articulated them over a half century ago.[2] Although there has been some activity in the state legislatures, the most significant advances there involve security not privacy.[3]

The quiescence in privacy law seems puzzling given the seismic shifts that have occurred in the collection, use, and distribution of information about individuals.[4] Companies increasingly monitor the once private

1. The exception is the Genetic Information Nondisclosure Act ("GINA"), enacted in 2008. Genetic Information Nondiscrimination Act of 2008, H.R. 493, 110th Cong. § 2 (2008).

2. *See* RESTATEMENT (SECOND) OF TORTS (1965) (reflecting Prosser's views on privacy torts as articulated in William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383 (1960)).

3. *See, e.g.*, CAL. CIV. CODE §§ 1798.29(a), 1798.82(a) (West 2014) (adding security breach provisions to the California Civil Code); MASS. GEN. LAWS. ANN. ch. 93H, §§ 1–6 (West 2014) (establishing minimum standards for protecting personal information). For a discussion of the differences between security and privacy, see Derek E. Bambauer, *Privacy Versus Security*, 103 J. CRIM. L. & CRIMINOLOGY 667 (2013).

4. JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 107–55 (2012); HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY,

behavior and even thoughts of their customers.[5] Chasing profits, they hoard this data for future, undefined uses; redistribute it to countless third parties; and repurpose it in ways their customers never imagined.[6] And they do all of this in ways that engender confusion and anxiety, and sometimes lead to great risk of harm.[7]

All of this activity redounds to the benefit of governments too, which have relatively unobstructed access to this data and use it not only to build sophisticated prediction machines to combat terrorism, but also to watch political dissidents and spy on friends and rivals alike.[8] Even before Edward Snowden's revelations about unprecedented spying by the NSA,[9] many had called for the sensible, meaningful expansion of laws protecting privacy from the government.[10] Policymakers, legal scholars, and advocates have proposed sweeping privacy reform, including the creation of new laws, the revamp of old ones, and no less than a reconceptualization of how we protect privacy with law.[11] The drumbeat for reform has only quickened and grown louder since the Snowden leaks. Yet nothing ever changes.

This Article seeks to spur a wave of new privacy law by focusing on one critically important but undertheorized concept: sensitive information. The great variety of regulations, laws, technical standards, and corporate

POLICY, AND THE INTEGRITY OF SOCIAL LIFE 21–67 (2010); DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 22–26 (Ex Machina ed., 2004).

5. JULIA ANGWIN, DRAGNET NATION: A QUEST FOR PRIVACY, SECURITY, AND FREEDOM IN A WORLD OF RELENTLESS SURVEILLANCE 3–6 (1st ed. 2014); SOLOVE, *supra* note 4, at 22–27; Danielle Keats Citron, *Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age*, 80 S. CAL. L. REV. 241, 243 (2007); Paul Ohm, *The Fourth Amendment in A World Without Privacy*, 81 MISS. L.J. 1309, 1310 (2012) [hereinafter Ohm, *World Without Privacy*]; Paul Ohm, *The Rise and Fall of Invasive ISP Surveillance*, 2009 U. ILL. L. REV. 1417, 1417 [hereinafter Ohm, *ISP Surveillance*]; Julia Angwin, *The Web's New Gold Mine: Your Secrets,* WALL ST. J.: WHAT THEY KNOW SERIES (July 30, 2010), http://online.wsj.com/news/articles/SB10001424052748703940904575395073512989404 [hereinafter Angwin, *The Web's New Gold Mine*].

6. Angwin, *The Web's New Gold Mine*, *supra* note 5.

7. DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 1–12 (2008). See *infra* Part II (discussing the harm that can come from the misuse of sensitive information).

8. Ohm, *World Without Privacy*, *supra* note 5, at 1325–28.

9. Glenn Greenwald, *NSA Collecting Phone Records of Millions of Verizon Customers Daily*, THE GUARDIAN (June 5, 2013), http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order.

10. DANIEL J. SOLOVE, NOTHING TO HIDE 55–146 (2011); Patricia Bellia, *Surveillance Law Through Cyberlaw's Lens*, 72 GEO. WASH. L. REV. 1375, 1434–38 (2004). *See also* Digital Due Process, http://digitaldueprocess.org (last visited Aug. 24, 2015) (calling for ECPA reform in this website for coalition group).

11. Angwin, *The Web's New Gold Mine*, *supra* note 5.

practices that have been implemented to protect the privacy of information stored in databases share at their core this unifying construct.[12] Some categories of information—information about health, sex, financial activity, and education, to name only a few—are singled out, and database owners owe special duties and face burdensome constraints regarding sensitive information.[13]

Sensitive information is a showstopper. Otherwise lax regulations become stringent when applied to it.[14] Permissive laws set stricter rules for the sensitive.[15] The label plays a central role in rhetoric and debate, as even the most fervent advocate for free markets and unfettered trade in information will concede that companies should refrain from selling certain forms of sensitive information—especially health information—regardless of the cost.[16]

Despite the importance of sensitive information, very little scholarship has systematically studied this important but woefully undertheorized category.[17] What makes a type of information sensitive?[18] Are the sensitive categories set in stone, or do they vary with time and technological advance? What are the political mechanisms that lead categories of information into or out of the designation? Why does the designation serve as such a powerful rhetorical trump card? We rarely ask, much less try, to answer questions like these.

We should try to answer these questions because if we are going to expand privacy law, sensitive information is probably the only game in town. In the United States, we are unlikely to enact a new, comprehensive

12. ELOISE GRATTON, UNDERSTANDING PERSONAL INFORMATION: MANAGING PRIVACY RISKS (Lexis Nexis Canada, 2013); Andrew B. Serwin, *Privacy 3.0-the Principle of Proportionality*, 42 U. MICH. J.L. REFORM 869, 900 (2009).

13. *Infra* Part I.B.

14. *Infra* Part I.B.

15. *Infra* Part I.B.

16. J. Howard Beales, *Modification and Consumer Information: Modern Biotechnology and the Regulation of Information*, 55 FOOD & DRUG L.J. 105, 117 (2000); Adam Thierer, *Advertising, Commercial Speech, and First Amendment Parity*, 5 CHARLESTON L. REV. 503, 517 (2011).

17. It appears that only two scholarly works, both by practitioners, have tried to catalog the many definitions of sensitive information. GRATTON, *supra* note 12, at 91–145; Serwin, *supra* note 12, at 900–06.

18. Although very few scholars have attempted to define sensitive information with rigor, *see* GRATTON, *supra* note 12, at 147, most privacy scholars have used the phrase as a central, if vaguely defined, concept. *E.g.*, Woodrow Hartzog, *Reviving Implied Confidentiality*, 89 IND. L.J. 763 (2014); Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119 (2004); Joel R. Reidenberg, *Setting Standards for Fair Information Practices in the U.S. Private Sector*, 80 IOWA L. REV. 497 (1995); Lior Jacob Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHI. L. REV. 919 (2005).

privacy law along the lines of the European Data Protection Directive[19] because we lack widespread agreement that the general problem of privacy invasion is significant enough to justify such a sweeping approach.

This leaves us to regulate privacy piecemeal, by recognizing new forms of sensitive information. Just as we once enacted the Health Insurance Portability and Accountability Act ("HIPAA") to protect sensitive health information[20] and the Family Educational Rights and Privacy Act ("FERPA") to protect sensitive education records,[21] it is time to enact new protections for categories of sensitive information, old and new. But we cannot do so until we make the concept of sensitive information coherent.

At present, we define single categories of sensitive information in multiple, different, inconsistent ways, and this variation cannot be explained merely by differences in context or cosmetic variations in language. For example, similarly situated online advertising platforms prohibit advertising related to the sensitive category of health in dramatically different ways; some extend the prohibition to genomic information while others do not; some cover disabilities but others do not; some cover symptoms while others do not.[22]

Because of this confusion, new categories of sensitive information are rarely added to the positive law of privacy, and categories already enshrined in law are never removed. In the United States, a sectoral and primarily legislative approach to privacy law means that new categories are recognized as sensitive only when the slow—and at present dysfunctional—wheels of the legislative process turn to recognize them, often as a reaction to mere anecdote and folklore.[23] Even European Union lawmakers drafted a single list of sensitive information into the data protection directive more than fifteen years ago and have never seen fit to modify the initial list.[24]

To better understand the nature of sensitive information, the Article surveys the landscape of privacy law, identifying dozens of examples of special treatment for sensitive information in rules, laws, policy statements,

19. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC).

20. Heath Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996).

21. Family Educational Rights and Privacy Act, Pub. L. No. 93-380, 88 Stat. 57 (1974) (codified at 20 U.S.C. § 1232g (2012)).

22. *Infra* Part I.B.

23. Paul M. Schwartz, *Preemption and Privacy*, 118 YALE L.J. 902, 908–14 (2009).

24. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC).

academic writing, and corporate practices from a wide number of jurisdictions, in the United States and beyond.[25] From this survey, the Article reverse engineers the meaning of sensitive information. It then develops a multi-factor test that may be applied to explain, ex post, the types of information that have been deemed sensitive in the past and also to predict, ex ante, types of information that may be identified as sensitive going forward.

The factors number four. First, sensitive information can lead to significant forms of harm. Second, sensitive information is the kind that exposes the data subject to a high probability of such harm. Third, sensitive information often is information transmitted in a confidential setting. Fourth, sensitive information tends to involve harms that apply to the majority of data subjects.

The factors suggest we need better, more formal, and more rigorous frameworks for assessing the risk of privacy harm. Rather than focusing on anecdotes and intuition, we should design "privacy threat models," learning from computer scientists who study security and risk.[26]

With properly constructed threat models, we can rigorously assess the sensitivity of three forms of data that have been proposed as candidates for legal protection for many years. Threat models suggest that precise geolocation should be considered sensitive, especially given the underappreciated threats from company insiders; some forms of communications metadata should be considered sensitive, such as the list of URLs visited by users of the web; and remote biometric information, such as facial recognition data, may not yet qualify as sensitive, but the more this information is used to track location or identity, the more likely it will cross into the sensitive category.

Privacy threat models suggest two other expansions of privacy law. First, laws should more often protect sensitive information, such as health information, regardless of the identity of the person handling the data. HIPAA applies only to doctors, hospitals, insurers, and their "business associates," but it should be expanded to include any company possessing sensitive health information. Second, privacy laws should cover companies that create pools of sensitive information out of seemingly benign sets of

25. *Infra* Part II.B.

26. ADAM SHOSTACK, THREAT MODELING: DESIGNING FOR SECURITY 111–23 (2014); FRANK SWIDERSKI & WINDOW SNYDER, THREAT MODELING 25–33 (2004); Vineet Saini et al., *Threat Modeling Using Attack Trees*, 23 J. COMPUTING SCI. IN COLLEGES 124, 125 (2008). *See* Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. COLO. L. REV. 1117, 1149–51 (2013) (suggesting the need for privacy threat models).

unstructured data. Google has created a massive database of health symptoms from its users' search queries for its Flu Trends project, and privacy law should regulate the handling of this massive, sensitive new database.

By refocusing privacy law on harm and by giving new rigor to the way we assess harm, this Article paves the way to meaningful and necessary legal reform.

The Article proceeds in three parts. Part I defines sensitive information, contrasts it to related concepts, and outlines the problems that arise from our lack of understanding of what makes information sensitive. Part II surveys the law of sensitive information and from this survey, develops a model, a multi-factor test, for deciding whether a form of information is sensitive. Finally, Part III applies this model to address the problems presented earlier and to propose a framework for conducting privacy law reform.

I. THE MEANING AND IMPORT OF SENSITIVE INFORMATION

Sensitive information categories relate to types of privacy harm. One can usefully distinguish categories of sensitive information from what I am calling "protected channel laws."

Sensitive information is an undertheorized concept, one that tends to be inconsistently defined. This has led to significant confusion about privacy harm, and it has contributed to stasis in the development of new privacy law. Scholars have been reluctant to develop theories of sensitive information because they tend to downplay or even criticize the importance of privacy harm.

A. SENSITIVE INFORMATION

Although everybody in information privacy law and policy uses the term "sensitive information," few have tried to define the term, and even fewer have done so with rigor.[27] This subpart proposes a tentative definition, one developed extensively in Part II; gives a few examples of the kinds of rules that govern the handling of information deemed sensitive; and develops a new taxonomy for categorizing types of privacy law, among which sensitive information law is but one.

27. *See e.g.*, GRATTON, *supra* note 12 (discussing a more comprehensive definition of sensitive information).

1. What is Sensitive Information?

Sensitive information describes information that can be used to enable privacy or security harm when placed in the wrong hands. According to one definition, sensitive information is "any information that, when lost, can lead to significant contractual or legal liabilities; serious damage to your organization's image and reputation; or legal, financial, or business losses."[28] Other definitions vary slightly, but all focus on a risk of harm resulting from a loss of control over information.[29]

This Article focuses on information sensitive to people, putting aside definitions that focus in part or in whole on information, such as national security secrets or trade secrets, that could harm governments or other organizations.[30] These are important topics, but they sit outside this Article's focus on individual privacy.

Although sensitive information is used informally and loosely in many contexts, this Article focuses most of its attention on formal definitions for categories of sensitive information, as found in statutes, rules, and contracts. In all of these settings, the conclusion that a particular type of information should be treated as sensitive gives rise to special rules of collection, use, and disclosure as a means to prevent security or privacy harm.[31]

It is rarely the case that a law uses the term "sensitive information," or some variant, directly. Instead, laws single out specific, enumerated categories of information, such as health or financial, for special treatment. This Article pays attention to the way categories like these have been defined in statutes, regulations, and the common law, which I will refer to collectively as, "sensitive information laws."

28. *Protect Your Organization's Sensitive Information and Reputation with High-risk Data Discovery*, PRICEWATERHOUSECOOPERS (2010), http://www.pwc.com/us/en/it-risk-security/assets/high-risk-data-discovery.pdf (defining "high-risk data," which is used synonymously with "sensitive information").

29. TRANSP. SEC. ADMIN., OFFICE OF THE SPECIAL COUNSELOR, TSA MGMT. DIRECTIVE NO. 3700.4, HANDLING SENSITIVE PERSONALLY IDENTIFIABLE INFORMATION (Dec. 9, 2008), http://www.tsa.gov/video/pdfs/mds/TSA_MD_3700_4_FINALv3_081209.pdf; NAT'L INSTS. OF HEALTH, DEP'T OF HEALTH & HUMAN SERVS., GUIDE FOR IDENTIFYING AND HANDLING SENSITIVE INFORMATION AT THE NIH (Nov. 8, 2010), http://oma.od.nih.gov/public/MS/privacy/Documents/Guide%20for%20Handling%20Sensitive%20Information%20at%20NIH.pdf.

30. *See e.g.*, 47 U.S.C. § 929 (2012) (entitled "National security and other sensitive information"); FBI, DEP'T OF JUSTICE, WHITE PAPER: HIGHER EDUCATION AND NATIONAL SECURITY: THE TARGETING OF SENSITIVE, PROPRIETARY AND CLASSIFIED INFORMATION ON CAMPUSES OF HIGHER EDUCATION (Apr. 2011).

31. *Infra* Part I.A.2.

A paradigmatic example of a sensitive information statute is the European Union's data protection directive.[32] The directive, which has been transposed into national law by each European Union member state, prescribes rules for the processing of data across industries in the European Union.[33] While the law obligates "data processors" to obey a long list of rules regarding any personal information, for certain kinds of information, it requires even more diligence, broadly proscribing the processing of a list of defined sensitive categories of data: "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life."[34]

A regulatory example is the Federal Trade Commission's definition of "personal information" in the Children's Online Privacy Protection Act ("COPPA") rule.[35] The rule defines personal information to include, inter alia, "[a] first and last name; [a] home or other physical address including street name and name of a city or town; . . . [a] telephone number; [and] [a] Social Security Number."[36]

An example from the common law includes a requirement that the privacy tort of disclosure must involve the disclosure of facts that would be "highly offensive to a reasonable person."[37] Courts have found that many different kinds of information satisfy this definition, including Social Security Numbers, nude photographs, and information about sexual activity.[38]

2. What Rules Govern the Handling of Sensitive Information?

Sensitive information laws command the custodian of sensitive information to comply with special handling rules. For the most part, this Article focuses on the threshold definition of sensitive information and will not focus on the rules that apply to sensitive information, setting these aside for later works. But to provide a complete picture of the topic, consider a few examples.

32. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC).
33. *Id.*
34. *Id.* ¶ 1.
35. 16 C.F.R. § 312 (2014).
36. *Id.* § 312.2.
37. RESTATEMENT (SECOND) OF TORTS § 652D (1965).
38. Bodah v. Lakeville Motor Express, Inc., 649 N.W.2d 859, 862–63 (Minn. Ct. App. 2002) (Social Security Number); Michaels v. Internet Entm't Grp., 5 F. Supp. 2d 823, 842 (C.D. Cal. 1998) (sex); McCabe v. Village Voice, Inc., 550 F. Supp. 525, 529 (E.D. Pa. 1982) (nude photos).

Often, possessing sensitive information is a threshold requirement of a law or regulation. The rules simply do not apply to information deemed not sensitive. COPPA extends obligations only to those who collect or disclose "personal information."[39]

In addition to threshold requirements, many sensitive information laws restrict the handling of sensitive information. These special handling rules derive mostly from the Fair Information Practice Principles ("FIPPs"), a set of widely discussed background principles for data collection, use, and disclosure that form the heart of modern privacy law.[40] For example, sensitive information can sometimes be maintained only if the data holder obeys the FIPPs of "purpose specification" and "use limitation," which together require the data handler to state a purpose for their collection or use of the data and to restrict their use to that purpose alone.[41] Quite often, holders of sensitive information must protect the information with adequate data security, another FIPP.[42]

Consider a few concrete examples of the way various sensitive information laws have implemented the FIPPs. HIPAA restricts the way hospitals and other covered entities may use protected information to a narrow set of permitted uses, unless patients have consented to additional uses, which is the reason nearly every American adult has signed a privacy waiver at the doctor's office in recent years.[43] COPPA requires, among other things, that websites directed to children obtain verifiable parental consent before collecting personal information from a child under thirteen.[44] The Video Privacy Protection Act ("VPPA") limits the disclosure of some forms of video viewing habits.[45] The European Union sweeps much more broadly, proscribing entirely the processing of information deemed sensitive absent explicit consent or unless the processing falls within a narrow list of enumerated purposes.[46]

39. Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506 (2012).

40. Fred Cate, *The Failure of Fair Information Practice Principles*, *in* CONSUMER PROTECTION IN THE AGE OF THE INFORMATION ECONOMY 342, 342–45 (2006), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1156972; Robert Gellman, *Fair Information Practices: A Basic History* (Feb. 11, 2015), http://bobgellman.com/rg-docs/rg-FIPShistory.pdf.

41. *Id.*

42. *Id.*

43. 45 C.F.R. § 164.502(a) (2014).

44. Children's Online Privacy Protection Act of 1998 (COPPA), 15 U.S.C. § 6502(b)(1)(A)(ii).

45. Video Privacy Protection Act of 1988, 18 U.S.C. § 2710 (2012).

46. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC).

3. What Is Not Sensitive Information Law: Protected Channels

Another way to define sensitive information is by considering what falls outside the definition. Consider a previously unrecognized distinction: there are two, and only two, distinct types of privacy law, which I label "sensitive information law" and "protected channel law."

Protected channel laws, as the name suggests, protect the privacy of particular channels of communication. As a key example, the U.S. Wiretap Act proscribes the collection, use, or disclosure of information obtained from particular providers of real-time communication services, such as telephone companies and broadband Internet providers.[47] Similarly, the Stored Communications Act protects communications stored with some types of online intermediaries.[48] Recently, states have been enacting laws prohibiting employers from looking at the private social networking profiles of employees and prospective employees.[49] These laws treat Facebook and other social network sites as the newest protected channels. These three categories of laws, and others like them, protect the channel of communication, irrespective of the content of the communication transmitted or stored.

The distinction between sensitive information and protected channel approaches to privacy law tends to blur in the middle. Health information transiting a hospital network is protected by both HIPAA and the Wiretap Act.[50] But for the most part, legislatures tend to choose one or the other of these two approaches.

4. The Only Game in Town

For those desiring new privacy rules to reduce the amount of information collected about individuals, there is a pragmatic reason to pay attention to sensitive information: it may be the only game in town. Aside from the social networking password laws[51]—laws that limit employer access to the social networking profiles of employees and job applicants—we have not seen a significant new protected channel law in decades. The modest privacy law reforms we have seen during that time have come only when Congress, state legislatures, and administrative agencies have

47. 18 U.S.C., ch. 119, §§ 2510–2522.

48. 18 U.S.C., ch. 121, §§ 2701–2712.

49. California Social Media Privacy Act of 2012, CAL. LAB. CODE § 980 (West 2012).

50. Heath Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996); 18 U.S.C.A. § 2511 (2008).

51. LAB. § 980.

tightened restrictions in laws based on sensitive information.[52]

Most importantly, Congress enacted the Genetic Information Nondiscrimination Act ("GINA") in 2008.[53] This marked the first new substantial category of sensitive information to gain Congressional recognition in over a decade.[54] Under GINA, health insurers may not base coverage decisions or set premiums based solely on genetic predisposition to disease, and employers may not consider genetic information in making personnel decisions about employees.[55]

If we broaden our time horizon, we see that this is not a recent trend. Throughout the evolution of privacy law in this country, the vast majority of expansion has come from sensitive information approaches.[56]

Why has the sensitive approach succeeded where others have failed? This Article does not offer a comprehensive theory behind the political economy of sensitive information, but considers a few arguments in brief. Calls to protect sensitive information tend to be salient, even sensationalistic.[57] They tend to provide advocates with "poster children"—victims of privacy harm with a human face. This has led to an overreliance by policymakers on anecdote to support calls for enshrining new categories of sensitive information into law, a problem discussed later.[58]

Sensitive information approaches tend also to be amenable to narrow contexts, such as the health industry or financial industry.[59] This gives advocates a narrower set of adversaries and a more tailored set of arguments. It also provides those who might be regulated by a new law an opportunity to tailor the law's protections to lessen the regulatory impact. Unlike protected channel laws, which tend to sweep broadly and overprotect intentionally, sensitive information laws fit well with the "cost/benefit" regulatory approach that dominates the modern regulatory

52. *See, e.g.*, HITECH Act, Pub. L. No. 111-5, 123 Stat. 226 (2009) (codified as amended in scattered sections of 42 U.S.C.); 16 C.F.R. § 312 (2014).

53. Genetic Information Nondiscrimination Act of 2008, H.R. 493, 110th Cong. § 2 (2008).

54. Prior to GINA, the next most recent federal sensitive information law was Gramm-Leach-Bliley, enacted in 1999. Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999).

55. *Id.* §§ 202–205.

56. A leading privacy law casebook lists twenty privacy statutes enacted by Congress, perhaps only one of which, the Electronic Communications Privacy Act, does not qualify as a sensitive information law. DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 36–37 (3d ed. 2009).

57. Neil M. Richards, *The Perils of Social Reading*, 101 GEO. L.J. 689, 694–95 (2013).

58. *Infra* Part I.C.1.

59. Heath Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996) (codified as amended in scattered sections of 42 U.S.C.); Family Educational Rights and Privacy Act, 20 U.S.C. § 1232g (2012).

state.[60]

This Article thus seeks to refocus attention on sensitive information but in a reinvigorated and expanded fashion. Sensitive information laws can and should do more work, they can and should apply more often, and their remedies should be easier to obtain. Legislatures should enact new laws recognizing new forms of sensitive information; law enforcement and administrative agencies should recognize the way changing circumstances have expanded the meaning of what is sensitive in the law; and judges should reverse the way they have begun to clamp down on discussions of sensitivity.

B. INCONSISTENCIES IN DEFINING SENSITIVE INFORMATION

We lack a coherent theory of sensitive information. For proof, consider an example from the private sector. The online behavioral advertising industry targets online advertising to consumers based on evidence of individual preferences or behavior tracked at some point in the past.[61] Private actors in this industry always draw self-imposed lines between sensitive and nonsensitive information, separating the facts from a person's observable history that legitimately can form the basis for a targeted ad from the facts that should never play this role. They are probably motivated to draw these lines by a combination of moral compunction, ethical norms, market demand, and fear of consumer backlash or government regulation.

The entities that have defined categories of sensitive information for online behavioral advertising ("OBA") have tackled the special category of sensitive health information in oddly inconsistent ways. The Network Advertising Initiative ("NAI"), a trade association established to set principles of responsible OBA for its members,[62] promulgated principles in 2008 that defined sensitive information to include "[p]recise information about past, present, or potential future health or medical conditions or treatments, including genetic, genomic, and family medical history."[63]

60. Exec. Order No. 12866, 58 Fed. Reg. 51735 (Oct. 4, 1993); CASS R. SUNSTEIN, THE COST-BENEFIT STATE: THE FUTURE OF REGULATORY PROTECTION 14 (2002).

61. JOSEPH TUROW, THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH 65–88 (2013).

62. National Advertising Initiative, *About the NAI*, (2012), http://www.networkadvertising.org/about-nai/about-nai ("The NAI's mission is to create, draft, and oversee the self-regulation of the third-party online advertising industry through enforceable standards and ongoing compliance efforts."). National Advertising Initiative, History, http://www.networkadvertising.org/about-nai/history (last visited February 8, 2013).

63. Network Advertising Initiative, 2013 *NAI Code of Conduct*, 4 (2013),

Meanwhile another trade group, the Digital Advertising Alliance ("DAA"), released its own voluntary principles in July 2009, one year after the NAI's latest version, and defines sensitive information to include "pharmaceutical prescriptions, or medical records about a specific individual."[64] Facebook, a member of neither association, does not use the term "sensitive information" in this context but requires advertisers to agree not to target ads to "disability or medical condition (including physical or mental health)."[65] Google's corresponding list prohibits ads based on the very broad category of "health or medical information."[66]

We must remember that privacy is deeply contextual, and we should not expect consistency across contextual boundaries.[67] Norms shift and differ. Nevertheless, the inconsistency of definitions of sensitive information seems hard to explain away as merely contextual variation. Even when we hold most things equal, lists of sensitive information tend to differ.

Consider the plight of an online advertiser hoping to advertise on Facebook, Google, an NAI platform, and a DAA platform based on their health information. Advertisers can definitely target ads to people suffering from a particular disability on DAA platforms, definitely not on Facebook, and probably not on Google or NAI. Genomic information is only expressly prohibited within the NAI definition, arguably within Google's, and likely not Facebook's or DAA's. Ads targeted to symptoms might be barred by Google and maybe NAI, but probably not by Facebook or DAA.[68] There is no rhyme or reason here,[69] and it suggests that categories of sensitive information are not being thoughtfully or rigorously generated. It suggests a need for a theory of sensitive information and perhaps better

http://www.networkadvertising.org/sites/default/files/imce/principles.pdf.

64. American Association of Advertising Agencies, *Self Regulatory Principles for Online Behavioral Advertising*, 4 (July 2009), http://www.aboutads.info/resource/download/seven-principles-07-01-09.pdf.

65. *Facebook Advertising Guidelines*, (June 4, 2014), https://www.facebook.com/ad_guidelines.php.

66. Google, *Policy for Advertising Based on Interests and Location*, https://support.google.com/adwordspolicy/answer/143465?hl=en.

67. NISSENBAUM, *supra* note 4, at 186–231.

68. *Cf.* Michael Barbaro & Tom Zeller Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES (Aug. 9, 2006), http://www.nytimes.com/2006/08/09/technology/09aol.html (reporting that users will often enter their own symptoms into search engines).

69. Jim Brock, *Yet Another (Better) Definition of Sensitive Boundaries for Ad Targeting*, PRIVACYCHOICE (Dec. 14, 2011), http://blog.privacychoice.org/2011/12/14/yet-another-better-definition-of-sensitive-boundaries-for-ad-targeting/. ("The self-regulatory effort would be more credible if the various standards for sensitive boundaries were unified and strengthened along the lines of Google's definition.").

coordination among those who generate lists of sensitive information.

C. HOW DO WE DECIDE SOMETHING IS SENSITIVE?

To date, policymakers have identified categories of sensitive information in an informal, nonscientific, and anecdotal manner. Some categories of information, such as health, financial, and education, have seemingly always been considered sensitive, so further analysis seems unnecessary today, or so the attitude seems to be. New categories of information such as records held by the department of motor vehicles or video stores were created in response to a single, salient story. Meanwhile, as technology creates new categories of information that seem potentially sensitive, such as precise geolocation, biometric, and metadata, we are left without a rigorous method of assessment, biding our time until the next data disaster.

1. Ad Hoc, Anecdotal Development

Congress has enacted numerous privacy statutes that categorize particular types of information as sensitive.[70] These statutes, commonly referred to as "sectoral," delineate the borders of privacy protection according to types of information: health, financial, and education information are protected, while web browsing history information (for the most part) is not.[71]

Congress's approach to this endeavor has been, to put it mildly, haphazard. Usually, it has taken an anecdotal, nonscientific approach in deciding whether something is sensitive. Most famously, Congress enacted the Video Privacy Protection Act ("VPPA") in 1998,[72] which treats as sensitive (without using the label, "sensitive"), "title, description, or subject matter of any video tapes or other audio visual material."[73] According to the generally accepted story, Congress created this Act almost entirely because a reporter obtained the video rental records for Judge Robert Bork during his confirmation hearings regarding his doomed nomination to the Supreme Court.[74]

Another famous example is the Driver's Privacy Protection Act

70. *Infra* Part II.B.

71. *Id.*

72. Video Privacy Protection Act of 1988, Pub. L. No. 100-618, 102 Stat. 3195 (codified as amended at 18 U.S.C. §§ 2710 (2012)).

73. 18 U.S.C. § 2710(b)(2)(D)(ii).

74. Michael Dolan, *The Bork Tapes*, THE AM. PORCH (Oct. 1, 1987), http://www.theamericanporch.com/bork5.htm.

("DPPA"), enacted in 1994.[75] The DPPA, more than the VPPA, is a classic "sensitive information" statute, defining two separate categories of information deserving of privacy protection. The sensitive category, called "highly restricted personal information," includes "an individual's photograph or image, social security number, medical or disability information."[76] This law was modeled after several similar state laws, some of which were inspired directly by the murder of actress Rebecca Schaeffer, killed by a deranged fan who located her using records he purchased from the California DMV.[77]

There are many problems with this ad hoc and anecdotal approach. The process arguably protects information it should not. Worse, it fails to respond quickly, meaning categories of information that can lead to harm and thus probably should be protected are not, while Congress awaits the next horrifying and salient anecdote.

Because Congress creates sensitive information laws often in response to anecdote, these laws tend to seem unprincipled, making them easy for critics to criticize as protecting unimportant values. Netflix waged a multi-year campaign against the VPPA, saying the Act restricted innovative social sharing features it wanted to build into its service.[78] A key argument Netflix advanced was that video records are not terribly sensitive.[79] In the waning days of the last Congress, a time when almost nothing moved on Capitol Hill, Netflix finally prevailed, convincing both houses of Congress and the President to amend the VPPA to reduce its privacy protections.[80]

Although Congress was wrong to capitulate to Netflix, the company may on one level have had a point. If the question is whether we should protect video watching habits, the answer is probably yes.[81] But if the question is whether it makes sense to protect video watching habits so severely when other, arguably more sensitive types of information receive little or no protection, then the answer is less obviously yes. The same could probably be said for other privacy statutes. The DPPA singles out

75. 18 U.S.C. § 2721–2725 (2012).

76. 18 U.S.C. § 2725(4).

77. Darrell Dawsey et al., *Actress Rebecca Schaeffer Fatally Shot at Apartment*, L.A. TIMES, July 19, 1989.

78. Julianne Pepitone, *Why Netflix's Facebook App Would Be Illegal*, CNN MONEY (Mar. 27, 2012), http://money.cnn.com/2012/03/27/technology/netflix-facebook/.

79. In re Netflix Privacy Litig., 5:11-CV-00379(EJD), 2012 WL 2598819 (N.D. Cal. July 5, 2012).

80. Video Privacy Protection Act Amendments Act of 2012, Pub. L. No. 112-258, 126 Stat. 2414 (2013).

81. Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 418 (2008).

driver's license records[82] and the Cable Privacy Protection Act singles out cable subscription information in ways that seem overprotective when compared to the fact that search queries and web history tend not to be protected.[83] Although the rest of this Article will argue for the expansion of privacy law, it is important to acknowledge that a less-anecdotal, more rigorous approach to defining categories of sensitive information may lead to rolling back some privacy laws, too.

Because legislative bodies take an ad hoc, anecdotal approach to defining sensitive information, the categories they define as sensitive change very slowly and infrequently. Consider, for example, the European Union's Data Protection Directive.[84] First enacted in 1996, it applies special rules to a narrow class of sensitive information.[85] Not once in nearly two decades has the EU seen fit to expand or contract this list.

Or consider once again GINA, Congress's newest creation of a category of sensitive information and arguably the only meaningful expansion of privacy law Congress has enacted in the last decade.[86] GINA took a long time to enact. It was first introduced in 1995, and then reintroduced in every subsequent Congress until it was finally enacted in 2008.[87] The legislative effort took this long despite the work of many advocacy groups.[88] These groups understood well the importance of anecdote, highlighting throughout those thirteen years the stories of people suffering under the type of discrimination GINA is intended to prevent.[89] And although GINA's supporters perhaps lacked a single unifying anecdote with the power of Robert Bork or Rebecca Schaeffer, many focused on Aldous Huxley's *Brave New World*[90] and the movie *Gattaca*[91] as two salient fictional portraits of a future without GINA.

82. Driver's Privacy Protection Act of 1994, 18 U.S.C. §§ 2721–2725 (2012)).

83. 47 U.S.C. § 551.

84. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1.

85. *Id.*

86. Genetic Information Nondiscrimination Act of 2008, H.R. 493, 110th Cong. § 2 (2008).

87. *E.g.*, The Genetic Privacy and Nondiscrimination Act of 1996, S. 1416; Genetic Information Nondiscrimination Act, H.R. 1910 (2003); Genetic Information Nondiscrimination Act, H.R. 1227.

88. Coalition for Genetic Fairness Resource, in partnership with the National Partnership for Women & Families, *Faces of Genetic Discrimination: How Genetic Discrimination Affects Real People* (July 2004), http://go.nationalpartnership.org/site/DocServer/FacesofGeneticDiscrimination.pdf.

89. *Id.*

90. ALDOUS HUXLEY, BRAVE NEW WORLD (Perennial Classics 1998) (1932).

91. GATTACA (Columbia Pictures & Jersey Films 1997).

2. Three Candidates for Sensitive Information

This slow development of sensitive information law has had trouble keeping up with advances in technology. Today, candidates for new sensitive information categories seem to arise with alarming frequency. Consider three: precise geolocation, biometric information, and communications metadata.

a. Precise Geolocation

The last few years have seen an explosion in the technology of precise geolocation tracking and with it, mounting concerns about privacy harms that might result.[92] As I and others have documented, two technologies have contributed most to this change: cheap GPS sensors embedded in small mobile devices and smartphone operating systems and apps on those devices that disclose location information to many people.[93]

Geolocation has already been recognized as sensitive by some government bodies, but this conclusion has not yet been integrated broadly into law. The FTC included "precise geolocation information" in its recent best practices, agenda-setting Privacy Report in a list of sensitive information categories that should not be collected absent meaningful and affirmative consent.[94] At least six bills in Congress have been introduced that would constrain either government or corporate uses of geolocation data, although none has yet been enacted.[95]

b. Remote Biometric

Commentators have also focused on the potential privacy problems of the spread of biometric information, such as iris scan, fingerprint, and facial recognition data.[96] To date, legislatures have been very slow to adopt laws focused on this information.[97]

Biometric information raises at least two sets of privacy harms. First, biometric techniques that work from a distance, such as facial recognition of images captured through closed-circuit television, can record the precise location of individuals. This privacy concern raises most of the same

92. Ohm, *World Without Privacy*, *supra* note 5, at 1353–55.

93. *Id.*

94. Fed. Trade Comm'n, *2012 Privacy Report* (Mar. 26, 2012), http://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/ftc-privacy-report.

95. *Geolocation Privacy Legislation*, GPS.GOV (June 18, 2014), http://www.gps.gov/policy/legislation/gps-act/.

96. Laura K. Donohue, *Technological Leap, Statutory Gap, and Constitutional Abyss: Remote Biometric Identification Comes of Age*, 97 MINN. L. REV. 407, 409 (2012).

97. Margaret Hu, *Biometric ID Cybersurveillance*, 88 IND. L.J. 1475 (2013).

concerns as precise geolocation.

Biometric information will also give rise to potential harm if system designers begin to use it as a password for access to personal account information—the way social security numbers have been used for many years. For example, if hospital websites or apartment doors or bank vaults began to use iris or fingerprint scan data to regulate authorized access, then the risk of mischief and harm that might result if that data fell into the wrong hands would increase.

c. Metadata

Finally, for at least a decade, scholars have argued for more privacy protection for the metadata associated with communications technologies, such as telephone, email, and web browsing.[98] These calls have spilled into the public consciousness as a result of Edward Snowden's revelations about the NSA,[99] and in large part, President Obama, who, in the earliest days after the first revelations, tried to assure the public that the information gathered under the so-called Section 215 program was "only" metadata and thus not sensitive.[100] Many have contested this claim, and some have called for amendments to privacy law to clarify that metadata deserves full protection.[101] The White House has more recently conceded the need for some reforms, which are still being hashed out.[102]

D. ARGUMENTS AGAINST FOCUSING ON SENSITIVE INFORMATION

Although it may seem intuitive to focus on sensitive information, other scholars have suggested that this focus misses what is important in privacy. Helen Nissenbaum has argued specifically against theories of privacy that focus on sensitive information. Many other scholars have argued against theories that focus on privacy harm, which sits at the heart of the definition of sensitive information. I respectfully disagree.

98. *See, e.g.*, Patricia L. Bellia & Susan Freiwald, *Fourth Amendment Protection for Stored E-mail*, 2008 U. CHI. LEGAL F. 121, 125; SOLOVE, *supra* note 10.

99. Greenwald, *supra* note 9.

100. Jonathan Weisman, *Obama Pledges Quick Action on Economic Stimulus*, WALL ST. J., Nov. 8, 2008.

101. Jennifer Stisa Granick & Christopher Jon Sprigman, *The Criminal N.S.A.*, N.Y. TIMES, June 27, 2013; PRIVACY & CIV. LIBERTIES OVERSIGHT B., REPORT ON THE TELEPHONE RECORDS PROGRAM CONDUCTED UNDER SECTION 215 OF THE USA PATRIOT ACT AND ON THE OPERATIONS OF THE FOREIGN INTELLIGENCE SURVEILLANCE COURT (2014) [hereinafter PCLOB Report], http://www.pclob.gov/SiteAssets/Pages/default/PCLOB-Report-on-the-Telephone-Records-Program.pdf.

102. Charlie Savage, *Obama to Call for End to N.S.A.'s Bulk Data Collection*, N.Y. TIMES, Mar. 24, 2014, at A1.

1. The Argument Against from *Privacy in Context*

Helen Nissenbaum, in her influential book, *Privacy in Context*,[103] offers an approach to privacy called "contextual integrity" as the superior alternative to other privacy theories, including theories premised on sensitive information.[104] Contextual integrity defines privacy as a "right to appropriate flow of personal information."[105] "[I]t is a right to live in a world in which our expectations about the flow of personal information are, for the most part, met."[106] Translated into practice, contextual integrity requires a rigorous accounting of information flow in narrow contexts, an attempt to divine preexisting norms by studying four categories: contexts, actors, information types (attributes), and transmission principles.[107] Nissenbaum would probably see proposals to single out particular types of information as sensitive as sweeping far too broadly and completely. Health information, such as from a hospital database, is sensitive, except when it is not, such as public testimonials by people battling a disease.

It may be that there is no conflict at all. This Article may simply be an implementation of privacy in context. The primary subject—the creation and use of large databases of personal information collected by commercial entities—investigated in this Article may itself qualify as a proper "context" in the Nissenbaumian sense, and if so, this Article may simply be an investigation around contextual integrity with respect to this context and harmful flows of information.

On the other hand, it may be that this context sweeps too broadly for Nissenbaum's model. Although Nissenbaum does not define the proper size or scope of a context with precision, her examples of contexts tend to be narrower than the one I have just proposed.[108] But even if "all commercial databases" is too broad a context, this Article's prescriptions could provide a framework for contextual integrity in narrower contexts, such as "health care databases" or "search engine databases."

103. NISSENBAUM, *supra* note 4.
104. *Id.* at 3.
105. *Id.* at 127.
106. *Id.* at 231.
107. *Id.* at 141–47.
108. *Id.* at 132 ("Contexts are structured social settings characterized by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)."). Nissenbaum offers contexts as broad as "education or health care" and as narrow as "social networking site" or "voting stations, courtrooms, and . . . church services." *Id.* at 135. In at least two instances, she highlights potential "contexts" that sweep too broadly to be useful. First, we cannot divide the world into two contexts, "public" and "private." *Id.* at 89–102. Second, we cannot hope to enact an "omnibus" privacy law that is supple enough for contextual integrity. *Id.* at 237–38.

Even if there is general alignment between the approaches taken by Nissenbaum and myself, I still need to address another critique Nissenbaum lodges specifically against approaches focusing on "public" versus "private" information, a distinction that overlaps somewhat with sensitive information. She cites philosophers like William Parent, Raymond Wacks, and Tom Gerety, as each suggesting that "information can be divided into two categories, public and private, and that we need only worry about imposing constraints on the flow of private information."[109] I do not embrace the argument she is attacking. Sensitive information law deserves to be only one tool in a broader toolkit of privacy protection. Information that is not sensitive might still lead to very important privacy problems and might deserve very different regulatory responses. I have written about some of these in my other work.[110]

Still, I take the point that regardless of my intent, some may take from my call to focus on sensitive information the argument that sensitive information deserves not just some focus but all of the focus of privacy protection. Legislatures that "fix" the problems of sensitive information might feel their motivation to protect privacy in other ways sapped and diminished. This is certainly not my intent.

But although I think my approach is consistent with Nissenbaum's, I must at least acknowledge the possibility that Nissenbaum and I might part company in some ways. I am concerned that Nissenbaum's approach to defining privacy in context requires a debate of precision and nuance that our political systems can simply not achieve. In parts of the book, Nissenbaum worries about the tendency for approaches other than hers to overprotect or underprotect privacy.[111] My approach to sensitive information is likely to overprotect or underprotect in some cases, but I think this is an unavoidable by-product of the messiness of the way we make and apply rules. Privacy in context is an important approach to privacy, but it should not supplant efforts to craft rules that rely more on heuristics or second-best results, including rules that identify sensitive information, when these are all the political processes can produce.

2. The New Privacy Scholars De-Emphasis of Harm

Many other privacy law scholars have resisted theories of privacy that turn on questions of individual harm, which might give them reason to

109. *Id.* at 120.

110. Paul Ohm, *Branding Privacy*, 97 MINN. L. REV. 907 (2013); Ohm, *ISP Surveillance*, *supra* note 5.

111. NISSENBAUM, *supra* note 4, at 89–102.

criticize my attempt to renew attention on sensitive information.[112] I refer to a group of scholars referred to by some as the New Privacy Scholars, including Dan Solove, Paul Schwartz, Julie Cohen, Priscilla Regan, Anita Allen, and others.[113]

It is likely that some of these scholars will resist a call to focus time and energy on sensitive information and harm, time and energy better directed elsewhere. These scholars focus on what I call "modern privacy harms" in Part II. These are harms that are broad and not very concretely defined, affecting individuals, groups, and society and rooted in theories of liberal philosophy, such as dignity, autonomy, and deliberative democracy, or theories from postmodern thought.[114]

My response is much more pragmatic than principled. On principle, I share the concerns and have in the past embraced the reconceptualizations these fine scholars have advanced.[115] But I find it difficult to continue to follow their path to the exclusion of seeking other solutions because they seem not to be gaining enough traction for my impatient tastes. Especially in the United States, the small victories for privacy we have achieved have come only when lawmakers have focused on concrete harm and specifically on sensitive information.

I believe that at this political moment, in this country with its idiosyncratic history of information privacy protection, policymakers are not ready to embrace the way New Privacy has framed the issues. Unfortunately, policymakers too often shrug their shoulders at the New Privacy harms because they lack the salience of traditional harms and are thus easy to ignore or outweigh; are stated so abstractly as to be incommensurable to other interests like security or economic efficiency; and do not lend themselves to testing or falsifiability.[116]

I think these scholars gave up on traditional harm back when harm

112. To be clear, many privacy scholars, including some of the New Privacy Scholars such as Dan Solove and Paul Schwartz, write regularly about traditional harm. This seems particularly true of a new generation of privacy scholars—people like Danielle Citron, Citron, *supra* note 5, Neil Richards, Richards, *Intellectual Privacy*, *supra* note 81, Bill McGeveran, William McGeveran, *The Law of Friction*, 2013 U. CHI. LEGAL F. 15 (2013), and Ryan Calo, M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1131 (2011)—who are developing new theories of harm and new prescriptions to combat harm.

113. *See e.g.*, Paul M. Schwartz & William M. Treanor, *The New Privacy*, 101 MICH. L. REV. 2163 (2003).

114. *Infra* Part II.C.1.c.

115. *See e.g.*, Ohm, *ISP Surveillance*, *supra* note 5.

116. Fed. Trade Comm'n, *supra* note 94, at C-3–C-8 (dissenting statement of Comm. J. Thomas Rosch).

stories were harder to identify than they are today. Most of these scholars began writing about privacy more than a decade ago and were motivated by alarming changes they saw coming in the earliest days of the Internet.[117] Back in 1994 through 2000, these prescient observers understood how changes both technological—such as cookies, DRM, and packet sniffers—and institutional—such as the online advertising business model—were significantly impacting our historical notions of personal privacy, even if they were not simultaneously leading to cognizable privacy harms.[118]

These scholars gave up on harm a few years too early. The subsequent decade has seen the rise of conditions that now seem ripe (to me, at least) for an explosion of more traditionally defined harms tied to these massive databases. Far too many people continue to suffer harm from data in databases without protection or possibility for redress. Attackers use databases to stalk victims, particularly women, with the worst cases ending in physical abuse and sometimes murder.[119] People are exposed to threats of blackmail, harassment or unlawful discrimination at the hands of companies and their employees who know an astonishing amount of information about each of us.[120] Governments leverage the power of private databases to watch people in their borders and stifle dissent.[121]

Worse, new technologies are amplifying the frequency with which harms like these can be committed and the magnitude of the harm. Pervasive monitoring of location raises significant concerns about physical security.[122] Other sensors now watch the way we drive, eat, exercise, and sleep,[123] giving our adversaries more information to use in causing us harm. Big data analytics help adversaries make more out of less information, acting like a force multiplier for harm.[124]

Given this recent history and poor track record, those who seek sweeping privacy reform need to acknowledge that they are not about to

117. Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373, 1373–1438 (2000); Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609 (1999); Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393 (2001).

118. Schwartz, *supra* note 117, at 1655–58.

119. *Infra* Part II.B.4.

120. *Infra* Part III.B.1.

121. Gary King et al., *How Censorship in China Allows Government Criticism but Silences Collective Expression*, 107 AM. POL. SCI. REV. 1 (2013).

122. *Infra* Part II.B.4.

123. Scott R. Peppet, *Unraveling Privacy: The Personal Prospectus and the Threat of a Full-Disclosure Future*, 105 NW. U. L. REV. 1153, 1154 (2011).

124. Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010) [hereinafter Ohm, *Broken Promises of Privacy*].

get it. Reinvigorating and expanding sensitive information law serves as a good second best alternative, indeed one with the potential to significantly expand and strengthen privacy law.

II. WHAT MAKES INFORMATION SENSITIVE?

This Part starts from the idea that underneath the surface variety of laws that define sensitive information lies a common approach. It derives the unstated factors that help us determine whether a particular type of information is sensitive and worthy of protection. This serves as a descriptive, not normative, restatement of the factors that add up to conclusions of sensitivity. Standing alone, it fills a gaping hole in our understanding, presenting for the first time a descriptive theory of sensitive information. Additionally, this analysis sets the groundwork for an argument for reinvigorating and expanding sensitive information laws, presented in Part III.

A. METHODOLOGY

This Part summarizes a thorough, but not comprehensive, survey of positive privacy law. In order to reverse engineer the factors legislatures, regulators, judges, and companies have used to develop lists of sensitive information, my research assistants and I surveyed dozens of laws, primarily from the United States along with a few others from around the world.

This Part surveys dozens of different laws, regulations, and self-regulatory principles that name particular types of information for special handling. For each item in the list, it tries to divine from the construction of the law or rule the principle factors that get at the nature of sensitive information. For example, how are sensitive categories specified? Are they listed in general or specific terms?

For many of the categories, the survey considers the history behind the enactment of each item. Do contemporaneous statements provide insights into the justifications of the rule? Finally, it looks to the way some of the laws, rules, or principles have been enforced since enactment to develop further insights into the meaning of sensitive information. How do adjudicants press claims of right under the rule? How do adjudicators decide the merits of those claims?

The goal is a list of factors that together explain why certain types of information have been deemed sensitive. The goal is descriptive accuracy, and the factors are meant to reflect the text and legislative history behind

pronouncements as best as we can.[125]

B. LIST OF SENSITIVE INFORMATION

Before reverse engineering the "why" of sensitive information, I present a list of the "what." This is an attempt to summarize a fairly complete list of categories of information that have been treated as sensitive in laws primarily from the United States but also in other laws from around the world.[126]

1. Health

The sensitivity of health information has been recognized for millennia. The Hippocratic Oath, developed over 2,000 years ago, acknowledged the fundamental role privacy plays in the provision of healthcare: "What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself holding such things shameful to be spoken about."[127] It is no surprise, then, that many centuries later there are modern protections that attempt to ensure individuals' sensitive health information is kept between the individual and the physician. As discussed in Part I, there is a broad range of information protected as private health information. But what is the definition of "health information" under the law? What information is actually protected?

At the federal level, HIPAA provides the basic framework for extensive privacy protections with respect to health information and records. But, as it turns out, Congress was barely concerned with issues of privacy when it passed HIPAA. Rather, Congress instituted new standards for the electronic transmission of health information and records "with the goal of reducing administrative costs."[128] Lawmakers eventually recognized the omission of substantive privacy protections within the original iteration of HIPAA, however, and addressed privacy by delegating the duty to promulgate protective privacy measures to the Secretary of the

125. This is similar to taxonomies and surveys that others have undertaken. Practitioners like Andrew Serwin have published similar surveys. Serwin, *supra* note 12. Scholars like Dan Solove and Chris Hoofnagle have tried to capture not only the positive law in their taxonomies, but also the privacy problems not well addressed by privacy law. Daniel J. Solove & Chris Jay Hoofnagle, *A Model Regime of Privacy Protection (Version 3.0)*, 2006 U. ILL. L. REV. 357.

126. For privacy law experts, this subpart will be review. You are encouraged to skip to Part II.C.

127. Ben A. Rich, *Postmodern Medicine: Deconstructing the Hippocratic Oath*, 65 U. COLO. L. REV. 77, 87 (1993) (quoting LUDWIG EDELSTEIN, THE HIPPOCRATIC OATH: TEXT, TRANSLATION, AND INTERPRETATION 3 (Henry Sigerist ed., 1943)).

128. H.R. Rep. No. 104-497(I), at 61 (1996).

Department of Health and Human Services ("HHS").[129]

Today, the regulations eventually promulgated by the HHS, collectively referred to as the "Privacy Rule," protect a large swath of health information, including genetic information, that relate "to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual."[130] In enacting the final Privacy Rule, which sets the federal "floor" for privacy protections related to health information, the HHS explained the harm that can ensue from a breach of health privacy: "[a breach] can have significant implications well beyond the physical health of [a] person, including the loss of a job, alienation of family and friends, the loss of health insurance, and public humiliation."[131] President Bill Clinton recognized the particular importance and sensitivity of health information when introducing the Privacy Rule, noting that "[n]othing is more private than someone's medical or psychiatric records."[132] But the Privacy Rule, at the time of its enactment, was not necessarily the most stringent health privacy law in the nation. In the absence of federal standards, many states had begun to take measures to protect the privacy of health information.[133]

State privacy laws often protect specific types of health information in addition to protecting health information in more generalized terms. Various state statutes protect evidence about specific types of medical conditions, like information about HIV/AIDS[134] or sexually transmitted diseases.[135] Other state laws protect information about certain types of medical treatment, such as for cancer.[136] In modern society, perhaps the single most prominent example of protected evidence of treatment is evidence of abortion, both because of the stigma attached as well as the threat of harm to the patient. Yet no state or federal law explicitly protects

129. Standards for Privacy of Individually Identifiable Health Information, 67 Fed. Reg. 53,182, 53,182 (Aug. 14, 2002) (codified at 45 C.F.R. pts. 160 & 164).

130. 45 C.F.R. § 160.103 (2014).

131. Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82,462, 82, 468 (Dec. 28, 2000) (codified at 45 C.F.R. pts. 160 & 164).

132. Press Release, The White House, Remarks by the President on Medical Privacy (Dec. 20, 2000), http://www.hhs.gov/ocr/privacy/hipaa/news/whpress.html.

133. For an exhaustive list of state laws prior to the enactment of HIPAA's Privacy Rule, see JOY PRITTS ET AL., THE STATE OF HEALTH PRIVACY: A SURVEY OF STATE HEALTH PRIVACY STATUTES (2d ed. 1999), http://ihcrp.georgetown.edu/privacy/pdfs/statereport1.pdf.

134. *See, e.g.*, ALA. CODE § 22-11A-14, 22, 54 (1975).

135. *See, e.g.*, 410 ILL. COMP. STAT. 325/8 (2013).

136. *See, e.g.*, CAL. HEALTH & SAFETY CODE § 103875 (West 2013).

information pertaining to evidence of abortion.[137] Moving down the scale of frequency, some laws protect evidence of hospital visits or treatments that do not include diagnoses or prescriptions. Mere records of admittance or appointments are sometimes protected. Given the relative lack of guidance states had prior to the issuance of HHS' Privacy Rule, it is fairly remarkable how uniform state protections for health information privacy have become. Nearly every state has protections in place for health information regarding sexually transmitted diseases, cancer, HIV/AIDS, and genetics.[138]

Perhaps the fastest growing category of protected health information includes evidence of propensity to disease or disability, especially through genetic information. In 2008, Congress passed GINA, which disallowed discrimination on the basis of genetic information with respect to employment and health insurance.[139] In passing GINA, lawmakers recognized that "discrimination based on a person's genetic identity is just as unacceptable as discrimination on the basis of race or religion."[140] The law was based on the notion that "[a] person's unique genetic code contains the most personal aspects of their identity," and the law was a response to Americans' "legitimate fears about how this deeply private information will be used."[141] The discriminatory potential derived from genetic information had already been seen in cases like *Norman-Bloodsaw v. Lawrence Berkeley Laboratory*,[142] which involved an employer that included genetic tests within its pre-employment medical screening of job applicants.[143] States have also taken steps to protect the genetic privacy of their citizens.[144] In the form of the human genome, the definition of private health information has added yet another sizeable category.

Oftentimes, health privacy laws must walk the line between absolute individual privacy and public health concerns. In this way, health information is somewhat unique in the realm of information privacy. The

137. *See* Alice Clapman, Note, *Privacy Rights and Abortion Outing: A Proposal for Using Common-Law Torts to Protect Abortion Patients and Staff*, 112 YALE L.J. 1545, 1547 (2003) (discussing state and federal statutory laws that have helped protect the privacy of abortion patients even though most laws were not passed for that purpose).

138. *See* JOY PRITTS ET AL., *supra* note 133 (outlining each states' privacy laws on health information regarding sexually transmitted diseases, cancer, HIV/AIDS, and genetics).

139. Genetic Information Nondiscrimination Act of 2008, H.R. 493, 110th Cong. § 2 (2008).

140. 154 Cong. Rec. S3363-01 (Apr. 24, 2008) (statement of Rep. Kennedy).

141. *Id.* at S3364.

142. Norman-Bloodsaw v. Lawrence Berkeley Lab., 135 F.3d 1260, 1269 (9th Cir. 1998).

143. *Id.* at 1265.

144. Serwin, *supra* note 12, at 912–13 (citing ALASKA STAT. §§ 18.13.010–18.13.100 (2006); IDAHO CODE § 39-8303(1)(a)-(d) (2008)).

original Privacy Rule acknowledged that overly strong health privacy laws have the potential to harm public health benefits.[145] An example of the costs and benefits of health privacy can be seen with many states' cancer registry systems. In California, for example, the state maintains a system for collecting information for the purpose of identifying cancer hazards to the public health and potential remedies.[146] Repositories of medical information like cancer registries necessarily contain a wealth of identifiable health information, and the risk of harm in inadvertently releasing identifiable medical information is high. But cancer registries also provide a public health benefit in that they give medical professionals unprecedented access to high volumes of valuable medical data. Cancer registries are just one example of the health privacy benefits and costs. Some state laws go even further and protect mere symptoms, but many do not.

Thus, personal health information is just as sensitive and important to individuals today as it was to individuals in the days of Hippocrates. Though health information has become more portable and granular because of advanced technology and medicine, the sensitivity of and value in protecting health information has been left relatively unquestioned.

2. Sex

Information about sex or a person's sexual behavior is often considered sensitive, private, and personal, but in reality, relatively few laws in the United States explicitly protect information about sex. From a global perspective, the European Union's Data Protection Directive provides more explicit protections for information regarding sexual behavior than does any analogous American law.[147] The European Union Data Protection Directive protects as sensitive information any data pertaining to an individual's "sex life."[148] No similar American law exists, although some American scholars have advocated an approach resembling

145. Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82,462, 82, 468 (Dec. 28, 2000) (codified at 45 C.F.R. pts. 160, 164) ("National standards for medical privacy must recognize the sometimes competing goals of improving individual and public health, advancing scientific knowledge, enforcing the laws of the land, and processing and paying claims for health care services. . . . We do not suggest that privacy is an absolute right that reigns supreme over all other rights. It does not. However, the case for privacy will depend on a number of factors that can influence the balance—the level of harm to the individual involved versus the needs of the public." (internal citations and quotations omitted)).

146. CAL. HEALTH & SAFETY CODE § 103875 (West 2013).

147. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1.

148. *Id.*

that of the European Union's.[149]

Sensitive information regarding an individual's sex includes not only information about a person's sex life, but also an individual's sexual orientation.[150] Although scholarship around sexual orientation has primarily focused on constitutional privacy protections, tort or constitutional remedies may also be available to those who are "outed," indicating that information regarding one's sexual orientation is regarded as extremely sensitive.[151] But as societies around the globe adapt to treat sexual orientation as both less stigmatized and less important for treatment under law, one wonders whether this may wane as a form of sensitive information.

While no American law explicitly protects imagery of a person's nakedness, there has been some scholarship that argues nakedness with respect to airport security scanners is sensitive information, exposure of which could cause serious harm.[152]

This category of sensitive information has received attention of late through debates about proposals to criminalize revenge porn, the term used to describe videos posted on the Internet of people engaging in sex without the consent of at least one of the people depicted.[153] Proponents of these state laws point to the harm people feel from revenge porn, even if they consented originally to the creation of the videos.[154] To date, eleven states have enacted such laws.[155]

149. *See* Tanith L. Balaban, *Comprehensive Data Privacy Legislation: Why Now is the Time?*, 1 CASE W. RES. J.L. TECH. & INTERNET 1, 32 (2009) (arguing that United States should follow the European Union's lead with respect to European Union's definition of sensitive information under the Privacy Directive).

150. For a survey of the sexual orientation privacy legal landscape, see Adam J. Kretz, Note, *The Right to Sexual Orientation Privacy: Strengthening Protections for Minors Who are "Outed" in Schools*, 42 J.L. & EDUC. 381 (2013).

151. *Id.* at 384 (discussing tort and constitutional claims for violations of privacy).

152. Yofi Tirosh & Michael Birnhack, *Naked in Front of the Machine: Does Airport Scanning Violate Privacy?*, 74 OHIO ST. L.J. 1263, 1286, 1306 (2013) ("It is not only the fact that scanning produces an image of passengers' unclothed bodies that is problematic: passengers' lack of control of their own bodies and of the information gathered about them is what defies their reasonable expectations and harms privacy.").

153. Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345 (2014).

154. Erica Goode, *Victims Push Laws to End Online Revenge Posts*, N.Y. TIMES, Sept. 23, 2013.

155. National Conf. of State Legislatures, State 'Revenge Porn' Legislation, http://www.ncsl.org/research/telecommunications-and-information-technology/state-revenge-porn-legislation.aspx, (Aug. 15, 2014).

3. Financial

Another widely recognized category of sensitive information is financial information. Financial information tends to be easily amenable to measureable economic harm, which makes it politically desirable to try to protect. Even some types of nonfinancial information are protected because of the potential for financial harm, such as through identity theft. For example, unique government identifiers, such as social security numbers, are often protected because they once could be easily used to obtain new identification cards and open lines of credit.

The Gramm-Leach-Bliley Act ("GLB") was enacted in part to acknowledge the privacy concerns surrounding consumer financial information.[156] If an individual's financial information is placed in the wrong hands, it can have significant consequences. Title V of the GLB, for example, was developed after a public outcry surrounding a high-profile case where a bank was discovered to have sold costumer financial information to telemarketers, leading to numerous customers being charged for credit cards without their permission.[157] This instance is not unique; there have been countless stories of how financial information can and has been used to harm individuals.[158]

Another reason financial information appears to be classified as sensitive is because of the expectation of confidentiality individuals have surrounding this type of information.[159] As one Senator noted during the floor debates of the GLB, "I think most of us have this vague concept that when we are dealing with our bank . . . that stuff is confidential . . . . You certainly do not have the expectation that that information is going to be shared."[160] The Senator compared this expectation you have about interacting with your bank to the relationships between individuals and their doctors, lawyers, and even religious advisors.[161]

156. Jolina C. Cuaresma, *The Gramm-Leach-Bliley Act*, 17 BERKLEY TECH. L.J. 497, 497 (2002).

157. Eric Poggemiller, *The Consumer Response to Privacy Provisions in Gramm-Leach-Bliley: Much Ado About Nothing?*, 6 N.C. BANKING INST. 617, 618 (2002).

158. *See* 145 CONG. REC. 13,891 (1999). The Record provides several anecdotes depicting the harm that can be caused by the abuse of financial information. For example, it was stated that a convicted felon ran up $45.7 million in charges on customer credit cards when a San Fernando Valley Bank sold their customer information to a telemarketer. *See id.*

159. *See id*.

160. *Id.*

161. *Id.*

4. Personal Safety

Some laws protect information in order to safeguard personal safety. Often these laws are spurred by concerns about past patterns of violence against women or children.[162] Sensitive information for these purposes includes home and work addresses,[163] and even online contact information, like telephone numbers and email addresses. In two notable tragic cases from the 1990s, stalkers used information stored in databases to find and kill the women they were stalking. In 1999, Liam Youens used an Internet-based service called Docusearch.com to learn the work address of a woman named Amy Lynn Boyer.[164] Youens used the information to find and fatally shoot Boyer as she left work before killing himself.[165] A few years earlier, a deranged fan murdered actress Rebecca Schaeffer after obtaining her home address from the California Department of Motor Vehicles. Schaffer's murder figured prominently in the push for state laws limiting access to DMV records, which culminated in the passage of the DPPA.[166]

Consider also the Nuremberg Files website.[167] An anti-abortion group hosted a website called Nuremberg Files that contained detailed profiles of abortion providers, including names, home addresses, photographs, driver's license numbers, social security numbers, and information about family members—including the names of schools that abortion providers' children attended.[168] "After the website's creation, two abortion doctors were shot at their homes."[169] Today, it has become so commonplace to publish private information as a means to criticize or intimidate a speaker online, the act has its own label: doxing.[170]

Despite all of these well-documented cases, in his survey of sensitive

162. *See* 16 C.F.R. § 312.2 (2014); FTC, PRIVACY ONLINE: A REPORT TO CONGRESS 4–5 (1998), http://www.ftc.gov/sites/default/files/documents/public_events/exploring-privacy-roundtable-series/priv-23a.pdf (detailing concerns over the availability of children's personal information in chat rooms where predators may have easy access to information).

163. Drivers' Privacy Protection Act, Pub. L. No. 103-322, 108 Stat. 2099-2102 (1994) (codified at 18 U.S.C. §§ 2721–2725 (1994 & Supp. 1998)); Danielle Keats Citron, *Mainstreaming Privacy Torts*, 98 CALIF. L. REV. 1805, 1817 (2010). *See* Remsburg v. Docusearch, Inc., 816 A.2d 1001, 1009 (N.H. 2003) (A stalker received a woman's date of birth, Social Security Number, home address, and work address from an information broker and later murdered the woman.).

164. Remsburg v. Docusearch, Inc., 816 A.2d 1001, 1005–06 (N.H. 2003).

165. *Id.* at 1006.

166. *See* 140 Cong. Rec. 7924-25 (1994) (statement of Rep. Moran).

167. Citron, *supra* note 163, at 1817.

168. *Id.*

169. *Id.*

170. Bruce Schneier, *Doxing as an Attack*, SCHNEIER ON SECURITY (Jan. 2, 2015), https://www.schneier.com/blog/archives/2015/01/doxing_as_an_at.html.

information, practitioner Andrew Serwin oddly concludes that email addresses, telephone numbers, and addresses are "non-sensitive information" that should be allowed to be collected without consent.[171]

5. Criminal Records

Many jurisdictions treat evidence of past crimes, such as booking records or records of incarceration, as protected private information. The Supreme Court has endorsed the notion that a person's privacy interest in "avoiding disclosure of personal matters" extends to his criminal record.[172] The Court has also long recognized a person's interest in preventing disclosure of the fact of mere criminal suspicion of that person.[173]

One of the four distinct common law privacy rights identified by William Prosser in 1960 is the tort of "public disclosure of embarrassing private facts."[174] Knowledge about past crimes and criminal accusations stigmatize the individual.[175] This stigmatization can lead to reputational harm when it comes to social standing, future employment, and relationships.[176]

6. Education

Another area of information that has traditionally been protected as sensitive is educational information. Several laws aim to protect student records, particularly records of student achievement—individual grades and grade point averages—and discipline.

Congress enacted the Federal Educational Rights and Privacy Act ("FERPA") in an attempt to give parents and students access to education records, as well as to protect individual privacy.[177] Prior to 1974 there was an increasing number of reported abuses of student education records in the United States.[178] Thousands of upset parents approached public policy groups with stories about how schools had abused their children's

171. Serwin, *supra* note 12, at 905–06.

172. U.S. Dep't of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 762 (1989) (quoting Whalen v. Roe, 429 U.S. 589, 599 (1977)).

173. United States v. Procter & Gamble Co., 356 U.S. 677, 682 (1958) (quoting United States v. Rose, 215 F.2d 617, 628–29 (3d Cir. 1954)).

174. William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383, 389 (1960).

175. *Id.*

176. U.S. v. Marion, 404 U.S. 307, 320 (1971) ("Arrest is a public act that may seriously interfere with the defendant's liberty, whether he is free on bail or not, and that may disrupt his employment, drain his financial resources, curtail his associations, subject him to public obloquy, and create anxiety in him, his family and his friends.").

177. 120 CONG. REC. 39,862 (1974).

178. 121 CONG. REC. 13,990 (1975).

educational records.[179] Many of the concerns brought forth by parents to these public policy groups dealt with the secrecy in which educational records were maintained.[180] Because parents and students had a hard time accessing the records, they could not challenge potentially harmful (and in many cases false) information contained within them.[181] Factual inaccuracies in educational records could significantly harm students in their future educational and professional endeavors.[182]

7. Information about Children

Many jurisdictions around the world treat the privacy of information about children as a special category.[183] The justifications for treating children's information as more sensitive than information on adults are numerous but include the idea that children are vulnerable and give information freely (especially online), which interferes with the gatekeeping role of the parent.[184] The most obvious potential harm resulting from the availability of children's information online is the harm associated with sexual predators.[185]

Protections for information about children began to take root in response to online privacy concerns. Congress enacted the Children's Online Privacy Protection Act ("COPPA"), which assigned rulemaking authority to the Federal Trade Commission ("FTC"), which implemented the Children's Online Privacy Protect Rule ("Children's Privacy Rule").[186] The Children's Privacy Rule sought to safeguard children's personal data from the harms that concerned legislators, including potential abuse by online marketers, deceptive trade practices, and general safety.[187] The Children's Privacy Rule defines the "personal data" it seeks to protect to include a child's first and last name, home address, e-mail or any other

179. *Id.*

180. *Id.*

181. *Id.*

182. *Id.*

183. This should be distinguished from attempts to keep particular forms of information—notably pornography—away from children.

184. Fed. Trade Comm'n, PRIVACY ONLINE: A REPORT TO CONGRESS 4–5 (1998), http://www.ftc.gov/sites/default/files/documents/public_events/exploring-privacy-roundtable-series/priv-23a.pdf.

185. *Id.*

186. Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6506 (2012); Laurel Jamtgaard, *Big Bird Meets Big Brother: A Look at the Children's Online Privacy Protection Act*, 16 SANTA CLARA COMPUTER & HIGH TECH. L.J. 385, 387 (2000).

187. Fed. Trade Comm'n, PRIVACY ONLINE: A REPORT TO CONGRESS 4–5 (1998), *available at* http://www.ftc.gov/sites/default/files/documents/public_events/exploring-privacy-roundtable-series/priv-23a.pdf.

online contact information, phone number, social security number, or the combination of a photograph of an individual coupled with the person's last name.[188]

8. Political Opinion

Political opinion, while not considered sensitive information by statute in the United States, is considered sensitive information in England specifically[189] and in the European Union at large.[190] Once again, the EU Personal Data Directive more explicitly protects certain types of information than does U.S. law, as it did with personal information regarding sex life. The United States' regime for protecting political opinion, however, is relatively nonexistent.[191]

9. Public Records

Many laws protect the privacy of information submitted to public institutions. States keep records memorializing almost every contact they have with their residents.[192] For example, when a person obtains a license to drive, receives a traffic citation, registers to vote, or gets married or divorced, a record is made.[193]

Protection of public records is governed by both state and federal regulation. Regulations surrounding access to public records attempt to balance two important, and often conflicting, interests: privacy and transparency.[194] While privacy aims to protect an individual's interest in access to records that contain personal information, transparency involves

188. 16 C.F.R. § 312.2 (2014).

189. Data Protection Act, 1998, c. 29, pt. I, § 2(b) (Eng.), http://www.legislation.gov.uk/ukpga/1998/29/section/2.

190. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC).

191. *See* Tanith L. Balaban, *Comprehensive Data Privacy Legislation: Why Now Is the Time?*, 1 CASE W. RES. J.L. TECH. & INTERNET 1, 32 (2009) ("One of the things a comprehensive United States data privacy scheme should take from the EU Directive is the bar on exploiting sensitive data including 'racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union memberships, and the processing of data concerning health or sex life.'") (quoting Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC)).

192. It is worth noting that the right of access to court records often differs from the right to access other public records. Nixon v. Warner Commc'ns, Inc., 435 U.S. 589 (1978). *See also* United States v. McVeigh, 119 F.3d 806, 811 (10th Cir. 1997) (citing Nixon for the proposition that right of access is not absolute); United States v. Amodeo, 71 F.3d. 1044, 1047–50 (2d Cir. 1995) (applying a balancing test to determine if public access is proper).

193. Daniel Solove, *Access and Aggregation: Public Records, Privacy, and the* Constitution, 86 MINN. L. REV. 1137, 1143–44 (2002).

194. *Id.* at 1140.

the desire to expose government bureaucracy to public scrutiny.[195]

Public records have often been considered sensitive, not necessarily because the information could cause reputational harm or embarrassment, but as a reflection of the sensitive relationship citizens have with their government. One idea behind laws like the Privacy Act is that if we must require our citizens and residents to divulge information to the government to participate in our society, we should protect that information from being used for purposes other than the purpose for which data was first required.[196]

As information becomes easier to access by the public, however, the need for increased privacy protections has become even more apparent. Until recently, public records were difficult to access. In order to obtain personal information about someone, a person would have to physically go to the place where the record was stored and sift through information.[197] The difficulty in accessing this information offered an inherent level of protection; however, as we enter the information age, this built-in protection is waning and new forms of privacy protections may need to be considered.

10. Historically Protected but Waning?

Two protected categories seem to be waning in prominence somewhat; perhaps a reflection of shifting historical conditions. In Europe in particular, membership in a trade union is still protected as private information.[198] In the United States, membership in the Communist Party was once considered protected information.[199]

11. Miscellaneous

Many other categories of information are protected as sensitive but only in a few, narrowly defined laws. These laws often reflect a political response to a specific controversy. American law protects the privacy of video watching habits, both at video rental stores and on cable television.[200]

195. *See id.* (discussing privacy and transparency interests).

196. 5 U.S.C. § 552a (2012). For a detailed look at the creation of the Privacy Act, see PRISCILLA A. REGAN, LEGISLATING PRIVACY 71–83 (1995).

197. Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CALIF. L. REV. 1, 21 (2013).

198. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC).

199. *See* ALAN F. WESTIN, COMPUTERS, PERSONNEL ADMINISTRATION, AND CITIZEN RIGHTS 143 (1979) (discussing the U.S. Government's discussion to cease asking about affiliation with the Communist Party in security clearance documents).

200. Video Privacy Protection Act of 1988, 18 U.S.C. § 2710 (2012).

Both COPPA and the Driver's Privacy Protection Act protect photographs as sensitive, presumably because stalkers may use photographs to track their targets.[201]

C. THE FACTORS

By synthesizing the items in this list, we can begin to develop the unstated rules for defining sensitive information. Those who promulgate privacy laws and rules—whether for the government or to order private behavior—tend to focus on four factors in assessing whether a given piece of information seems sensitive: the possibility of harm; probability of harm; presence of a confidential relationship; and whether the risk reflects majoritarian concerns.

1. Can Be Used to Cause Harm

The first, and perhaps only necessary, factor is a connection between the category of information and harm. Information is deemed sensitive if adversaries (to use the computer scientific term) can use it to cause harm to data subjects or related people. This formulation raises more questions than it answers, because harm tends to be a deeply contested topic in information privacy debates. In other work, I am attempting to create a new and better taxonomy for privacy harm, which I will only summarize here.[202] Privacy harm may usefully be divided into three groups, roughly tracing the evolution of information tort law: ancient, traditional, and modern.

The point of this summary is to provide a menu of privacy harms not to argue that all of these harms deserve a legal remedy. My analysis isolates privacy harm as the core concern of sensitive information, but it leaves for further research and debate which of the following privacy harms the law should recognize and seek to remedy. The method Part III proposes for classifying sensitive information categories—privacy threat modeling—is agnostic about which theory of harm we should recognize; it treats that decision as an input, something decided prior to the threat modeling step, which builds upon different conceptions of privacy harm, from narrow to expansive and everything in between.

201. 16 C.F.R. § 312.2 (2014) (COPPA rule); 18 U.S.C. § 2725 (2012) (DPPA Definitions).

202. Paul Ohm, Ancient, Traditional, and Modern Privacy Harms (unpublished manuscript) (on file with author).

a. Ancient Harms

Ancient information harms focus on an old, short, and slowly evolving list of information wrongs. These harms have been recognized for decades (in some cases centuries) and find expression in descriptive accounts of positive common law and statutes. The hallmark of these harms is the sense that they lend themselves to an easy-to-measure, if not strictly monetary, harm.[203] Examples include breach of confidentiality,[204] defamation,[205] infliction of emotional distress,[206] and blackmail.[207] More modern examples include identity theft[208] and stalking.[209]

Non-information-based harms extend beyond merely financial harm, and include even physical harm. Many categories of information are considered sensitive because they can be used to locate or contact an individual. The crimes of harassment[210] and stalking,[211] often directed at women, are abetted by rich databases that help offenders locate their targets. COPPA labels personal information collected from a child as sensitive, because it can be used "to contact" a child.[212]

b. Traditional Harms

At the end of the nineteenth and throughout the twentieth centuries, legal scholars and judges slowly expanded the type of information harms recognized by tort law. Warren and Brandeis are credited with precipitating this shift practically out of whole cloth with their foundational article, *The Right to Privacy*.[213] As soon as a decade after Samuel D. Warren and Louis D. Brandeis's articles, judges had already began recognizing causes of action based on their theories.[214]

The evolution continued with the work of William Prosser, the influential tort scholar, who synthesized the earliest cases and shaped their eventual expression in both a famous law review article[215] and the

203. Daniel Solove, *Privacy and Data Security Harms*, CONCURRING OPINIONS BLOG (Aug. 4, 2014), http://www.concurringopinions.com/archives/2014/08/privacy-and-data-security-harms.html.
204. RESTATEMENT (SECOND) OF TORTS § 652A (1965).
205. *Id.* § 652E.
206. *Id.* § 46
207. *Id.* § 623A.
208. *Id.* § 217.
209. *Id.* § 652B.
210. *E.g.* N.Y. PEN. L. § 240.25 (McKinney 2014) ("Harassment in the first degree").
211. *E.g.* CAL. PEN. CODE § 646.9 (West 2014) ("Stalking").
212. Children's Online Privacy Protection Act of 1998, 15 U.S.C. § 6501 (2012).
213. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890).
214. Pavesich v. New England Life Ins. Co., 50 S.E. 68 (Ga. 1905).
215. William Prosser, *Privacy*, 48 CALIF. L. REV. 383 (1960).

Restatement.[216] His four privacy torts were public disclosure of private facts, intrusion upon seclusion, false light, and appropriation.[217] For example, public disclosure of private facts arises when one discloses a private matter that is "highly offensive to a reasonable person" and "is not of legitimate concern to the public."[218] Although these torts have traditionally been applied outside the database context, plaintiffs have alleged every one of these four torts based on activity associated with a database,[219] and judges have, on rare occasion, accepted these theories.[220]

Traditional privacy harms differ from ancient harms because the former are harder to measure and often involve injuries to dignity or emotion. Warren and Brandeis criticized the state of tort law as they found it for failing to remedy "mere injury to the feelings" and called for the recognition of new tort principles in order to safeguard "an inviolate personality."[221]

These principles have impacted how we regard sensitive information, not merely in tort law but also in statute. Information tends not to be deemed sensitive if its release might do little more than embarrass a data subject.[222] Instead, the improper disclosure of sensitive information usually leads to stronger feelings, of humiliation, abasement, or ostracism.[223] Dignitary harms reflect the privacy tort requirements that information be "highly offensive to a reasonable person" and "not of legitimate concern to the public."[224]

Consider some examples from the survey in Part II of categories protected by law because of concerns about dignitary harm. Health information is protected because people tend to feel more than mere embarrassment when details about symptoms, diagnoses, or treatments become known to others.[225] These feelings reflect a complicated set of societal reactions. Some diagnoses are connected generally to practices considered shameful, such as certain features of one's sex life or drug

216. RESTATEMENT (SECOND) OF TORTS §§ 652B, 652C, 652D, 652E (1965).

217. *Id.*

218. *Id.* § 652D.

219. *See* Potocnik v. Carlson, 9 F. Supp. 981 (D. Minn. 2014).

220. Dayton Newspapers, Inc. v. Dep't of Air Force, 107 F. Supp. 2d 912 (S.D. Ohio 2009).

221. Warren & Brandeis, *supra* note 213, at 205.

222. Perere v. Louisiana Television Broad. Corp., 812 So. 2d 673 (La. Ct. App. 2001).

223. Patricia Sanchez Abril, *Recasting Privacy Torts in a Spaceless World*, 21 HARV. J.L. & TECH. 1, 8 (2007).

224. Cox Broad. Corp. v. Cohn, 420 U.S. 469 (1975).

225. Lawrence O. Gostin, *Health Information Privacy*, 80 CORNELL L. REV. 451, 490 (1995) ("Disclosure of some conditions can be stigmatizing, and can cause embarrassment, social isolation, and a loss of self-esteem.").

use.[226] Other diagnoses might tend to inspire feelings of horror, extreme pity, or disgust because they are related to death or contagion.[227]

Similarly, education information is protected for dignitary reasons, because education records can lead to inferences (even if false) about intelligence and diligence.[228] Such information can suggest prospects for advancement and success.[229] Such feelings are probably not as deeply felt for educational records as they tend to be for health information, however, demonstrating that the sensitivity of information tends to sit along a continuum.

c. Modern Harms

Over the past few decades, many scholars have recommended expanding or reconceptualizing information privacy law to account for harms beyond the concerns about dignity and emotion that underlie the privacy torts.[230] These scholars represent a large, growing, and diverse set of individuals, whose theories continue to evolve and spread, so any brief discussion necessarily commits the twin sins of reductionism and incompleteness. What tends to bind these scholars together, however, is concern about the rise of computerization, and the large databases, persistent monitoring, and new techniques of analysis it enables.

Like Warren and Brandeis, many of these scholars focus on effects on individuals. These scholars talk about the harm of losing control over one's information[231] or define privacy as limited access to self[232] or for the protection of intimacy.[233]

Others measure the effects not only on individuals but also on larger groups, including on society at large. These scholars have been dubbed the "New Privacy" school[234] or, alternatively, the "Information Privacy Law Project."[235] Instead of focusing on harm, these scholars have identified

226. *Id.*

227. *Id.*

228. Joint Statement in Explanation of Buckley/Pell Amendment, 120 Cong. Rec. 39862 (Dec. 31, 1974).

229. *Id.*

230. Neil Richards, *The Information Privacy Law Project*, 94 GEO. L.J. 1087, 1088–94 (2006) (summarizing scholarly trend); Schwartz & Treanor, *supra* note 113, at 2163–64 (same).

231. ALAN F. WESTIN, PRIVACY AND FREEDOM 169 (1967).

232. Ruth Gavison, *Privacy and the Limits of Law*, 89 YALE L.J. 421 (1980).

233. JULIE INNESS, PRIVACY, INTIMACY, AND ISOLATION 124 (1992).

234. For more information on the "New Privacy" school's perspective, see, for example, Schwartz & Treanor, *supra* note 113.

235. Gostin, *supra* note 225, at 490.

other "problems" (some are even reluctant to use the word "harms")[236] that arise when governments and companies amass large databases full of information about people.[237] Compared to the harms of the first hundred years of privacy law, these problems tend to be nontraditional, divorced from tort law, focused on societal concerns over individuals ones, and abstract.

Julie Cohen, a leading privacy scholar, rarely focuses on individualized, traditional harms. Cohen's theories of privacy have evolved, and today her work strongly supports two theoretical strands, both of which encourage others to shift their focus away from these kinds of harms. More than a decade ago, she forcefully advanced information privacy law as a way to provide "the conditions for meaningful autonomy."[238] In perhaps the most cited passage in all of information privacy law, "[p]ervasive monitoring of every first move or false start will, at the margin, incline choices toward the bland and the mainstream."[239] The concern here is not the kind of blackmail and harassment and severe embarrassment I have discussed above, but rather "a blunting and blurring of rough edges and sharp lines."[240]

In her more recent work, Cohen has distanced herself from the liberal theory framework that once supported her work, in particular what she now sees as unrealistic views of the way individuals develop autonomously.[241] Instead she argues for a more postmodern conception of privacy providing the conditions necessary for engaging in the "play of everyday practice," thus enabling "evolving subjectivity."[242]

Paul Schwartz has argued that new privacy invasions blunt the capacity for the Internet to become a powerful new force for deliberative democracy if people are chilled from speaking and listening.[243] Still others have talked about privacy's crucial role in protecting diversity of thought and equality.[244]

236. SOLOVE, *supra* note 7, at 74–77 (explaining use of "problem" not "harm" as proper label for privacy law taxonomy).

237. *Id.*

238. Cohen, *supra* note 117, at 1423.

239. *Id.* at 1426.

240. *Id.*

241. COHEN, *supra* note 4, at 110–15.

242. *Id.* at 131.

243. Schwartz, *supra* note 117, at 1675.

244. *E.g.*, Richards, *Intellectual Privacy*, *supra* note 81, at 443–45.

d. The Role of Shifting Social Norms

As we connect privacy and harm, we should be mindful that social norms might shift to diminish or extinguish harms relating to certain types of facts. Some of the harms described above, and particularly harms to reputation, may change as society decides to care about different things.

Some critics of information privacy law advance this concern forcefully.[245] These critics argue that information privacy is a dying notion, akin to Victorian era concepts of propriety that we now find hard to imagine people once holding.[246] According to these people, if everybody has a skeleton in his or her closet, then as technology continues to evolve to force or entice people to reveal those secrets, soon we will stop caring.[247] They point as proof to evolving moral standards of sex, drug use, profanity, and entertainment as examples of this trend.[248]

There are many responses. Most of the ancient harms listed in the prior discussion have nothing to do with shame and reputation. Thus, the harms of stalking, harassment, and invidious discrimination will persist even in a hypothetical future world of no shame.

Second, even for harms to reputation, just because we stop caring about some kinds of secrets does not mean we will have no secrets at all. One can imagine a future world where sexual mores are loosened beyond where they sit today, yet it still is considered improper to discuss another person's specific sexual habits in most contexts. This is merely a prediction, not an attempt to try to set a normative moral floor on evolving norms. And sex is not the only example. Other sources of shame—family history, drug use, addiction, past criminality, and financial trouble—are still unlikely to lose their stigma entirely, at least for the foreseeable future. So long as there are some secrets that people in a society find shameful, this will allow for harms like blackmail and the infliction of emotional distress.[249]

245. Mark Zuckerburg has forcefully advanced the claim that social media has changed social norms around privacy. Bobbie Johnson, *Privacy No Longer a Social Norm, Says Facebook Founder*, THE GUARDIAN (Jan. 10, 2010), http://www.theguardian.com/technology/2010/jan/11/facebook-privacy.

246. Felix Salmon, *How Technology Redefines Norms*, COLUMBIA JOURNALISM REV. AUDIT BLOG (May 20, 2013), http://www.cjr.org/the_audit/how_technology_redefines_norms.php.

247. JEFF JARVIS, PUBLIC PARTS: HOW SHARING IN THE DIGITAL AGE IMPROVES THE WAY WE WORK AND LIVE 56 (2011).

248. *Id.*

249. *See generally* Richard H. McAdams, *Group Norms, Gossip, and Blackmail*, 144 U. PA. L. REV. 2237 (1996) (discussing the connection between law and norms with a focus on the intersection of group norms and informational blackmail law).

2. Sufficiently High Probability of Harm

Information is not classified as sensitive information simply because it might conceivably lead to one of the harms listed above. Instead, information earns the label only if a legislature (or other rulemaking institution) decides that the probability of harm is sufficiently high. Thus, the legislature engages in a risk assessment, taking into consideration the likelihood that a particular type of information—say, a health diagnosis—will lead to a particular type of harm—say, unlawful discrimination or severe embarrassment.

It is in the assessment of this factor that anecdote and intuition loom large. In this step, salient stories like the tales of Rebecca Schaeffer or Robert Bork described above bear great weight.[250] The salience of these stories might trigger in individual lawmakers cognitive biases like the availability heuristic, which tends to elevate disproportionately the importance of events that are easy to conjure up in one's mind.[251]

Judges make these risk calculations too. In data security cases, judges often weigh the risk of harm in deciding standing or remedies.[252] Just because a massive database has been breached does not necessarily mean that the breacher will use the information to target the data subjects. Judges seem to feel most comfortable with completed, measureable harm, such as completed identity theft or anxiety so severe that it leads to a visit to the doctor.[253] Judges seem less willing to credit uncompleted but likely harm or completed but difficult-to-measure harm.[254]

Finally, legislatures and judges alike tend to incorporate several sliding scales in deciding whether certain risks of harm are acceptable or not. First, the higher the magnitude of the harm, the more attenuated the risk of harm may be. In COPPA, because the harm sought to be avoided is every parent's worst nightmare—a stranger stalking his or her child—the likelihood of risk required is relatively slight.[255] COPPA covers the collection of a mere street name plus name of city, even when a child lives

250. *See supra* notes 74–77 and accompanying text.

251. Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics And Biases*, *in* JUDGMENT AND DECISION MAKING: AN INTERDISCIPLINARY READER 38, 42–46 (Hal R. Arkes et al. eds., 2d ed. 2000).

252. Petriello v. Kalman, 576 A.2d 474 (Conn. 1990).

253. Clapper v. Amnesty Int'l USA, 133 S. Ct. 1138 (2013).

254. In re Google, Inc. Privacy Policy Litig., No. C 12-01382 PSG, 2012 WL 6738343 (N.D. Cal. Dec. 28, 2012).

255. S. Rep. 107-240, at 50 (2002).

on a major boulevard in a city of millions.[256]

Second, and related, when the magnitude of harm is high, we permit overprotection. HIPAA protects all information about patients, and FERPA protects all information about students, even though some of that information may be quite disconnected from harms like blackmail, discrimination, and even embarrassment. FERPA, for example, protects even the public release of "honor roll" information, lists that celebrate and honor rather than expose individuals to embarrassment or shame.[257]

3. Shared Confidentially

Many sensitive information laws seem to protect information that might cause harm only when held by particular parties, those who owe a duty of confidentiality to data subjects due to a special relationship. Some of these reflect venerable and formally defined promises of confidentiality, such as the doctor-patient relationship protected in HIPAA.[258] Others may protect relationships less widely protected in other areas of law, yet still widely accepted in society, such as the teacher-student relationship at the heart of FERPA[259] or the bank-account holder relationship protected by GLB.[260] Still others represent very new relationships, some barely recognized outside the privacy law in question, such as the relationship between video store owner and renter protected in the VPPA.[261]

These statutes thus seem to borrow as much from the law of confidentiality as the law of privacy. In tort law, as in many of these statutes, a breach of confidentiality action requires a particular kind of relationship.[262] The tort law seems to protect against harmful revelations of information, but only in the context of protected relationships.[263] At least in the United States, this has led to a crabbed, narrow confidentiality tort, one that rarely applies.[264]

256. Children's Online Privacy Protection Act, 15 U.S.C. § 6501(8)(b) (2014).

257. Lynn M. Daggett, *Bucking Up Buckley I: Making the Federal Student Records Statute Work*, 46 CATH. U.L. REV. 617, 664–65 (1997).

258. Heath Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936. (codified as amended in scattered section of 18, 26, 29, and 42 U.S.C.).

259. Family Educational Rights and Privacy Act, Pub. L. No. 93-380, 88 Stat. 57 (codified at 20 U.S.C. § 1232g (2012)).

260. Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (codified as amended in scattered section of 12 and 15 U.S.C.

261. Video Privacy Protection Act of 1988, 18 U.S.C. § 2710 (2012).

262. Neil M. Richards & Daniel J. Solove, *Privacy's Other Path: Recovering the Law of Confidentiality*, 96 Geo. L.J. 123, 148–58 (2007).

263. *Id.*

264. *Id.* at 156–57 (describing the "stunted development" of confidentiality law in the United

4. Reflects Majoritarian Concerns

Because lists of sensitive information tend to be defined by majoritarian institutions, most importantly legislatures and administrative agencies, they tend to reflect majoritarian interests. This is an underappreciated bias that takes three distinct forms. First, categories of information are likelier to be deemed sensitive when a large segment of the population can imagine being harmed by the uncontrolled revelation of the information.

Health information is a classic example. Every member of society can imagine having oneself or a close relative afflicted by a secret malady, the revelation of which would cause harm. Financial information and educational records share this characteristic.

Second, categories that do not lead to harm to a large segment of the population are nevertheless protected as sensitive if the ruling majority can relate to the affected minority. Consider the absence of legal rules forcing foreign banks to reveal the identity of Americans holding offshore bank accounts.[265] The vast majority of the American public have no offshore holdings, and probably care very little about protecting the privacy of those who do. Yet we might imagine that the politically powerful and connected have a vested interest in keeping this information protected.[266] This should cause us concern about the noncategorization of classes of information that affect maligned minority groups and groups without significant political support.

Third, the mechanisms that define sensitive information do not account for idiosyncratically sensitive information, categories of information that trigger harm, but only for a very small number of people. I speculate that almost every person living in technologically sophisticated societies has some idiosyncratically harmful information stored about them in some database somewhere, forming a core part of what I have called a "database of ruin,"[267] yet idiosyncratic forms of this are not covered by definitions of sensitive information.

Because information harms experienced by a very small number of relatively powerless individuals will not be recognized by majoritarian

States).

265. Carolyn Michelle Najera, Note, *Combating Offshore Tax Evasion*, 17 SW. J. INT'L LAW 205 (2011).

266. *See* Laura Saunders & Anita Greil, *Swiss Will Give Up Names of U.S. Taxpayers*, WALL ST. J., Nov. 18, 2009 (discussing U.S. investigation into identities behind offshore Swiss accounts).

267. Ohm, *World Without Privacy*, *supra* note 5.

legal institutions, many privacy harms are never addressed with categories of sensitive information.

D. THE THREE KINDS OF SENSITIVE INFORMATION

Focusing on the three categories of privacy harm reveals another unappreciated form of variety among at least three types of sensitive information, which I am calling inherently sensitive, inferentially sensitive, and instrumentally sensitive information.

Inherently sensitive information is information that causes concrete harm merely by being known to another. These categories of sensitive information are usually connected to the category of ancient harm, such as blackmail and harassment. Often, inherently sensitive information is the kind of information we tend to regard as embarrassing (or worse, such as humiliating or abasing). As such, inherently sensitive information tends to be held closely and kept secret, passed along only to trusted confidences. Information in this category tends also to ebb and flow over time, as social norms shift to treat some things as humiliating or leading to ostracism than others. Many forms of health information are inherently sensitive.

Inferentially sensitive information is connected to harm through at least one inferential or predictive step. Education records such as grades might be considered sensitive because they suggest—sometimes quite imperfectly—a trait of the student, such as inability or laziness. Past criminal history is sensitive not only for the inherent shame sometimes associated with criminal activity but also for the possibility of future danger or recidivism. Some information is sensitive not because people make accurate inferences from it, but instead from unfounded stereotype. Thus, we might consider a diagnosis of sexually transmitted disease sensitive because some regard it as a sign of promiscuity. In some places, religion or ethnic background might also be considered sensitive because of negative stereotypes people associate with the group.

Many big data techniques focus on drawing accurate inferences about people from data.[268] As these techniques increase, we might expand the use and breadth of categories of inferentially sensitive information. For example, banks are investigating methods for inferring credit worthiness from seemingly nonsensitive characteristics, and insurers are doing the same for health risks.[269] If inferences begin to be drawn in ways we feel

268. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 127 (2014).

269. *Id.*

unfair or unjust in some way, we might begin to classify information as sensitive if it can be used to enable such inferences.

Finally, some forms of information are only *instrumentally sensitive*. Information in this category can lead to harm, but not merely through the knowledge or inference of another person. The harm is not from being known; it is from being used. A classic example is home address. Stalkers or child predators use home addresses in order to find their victims. Another example is a social security number, which is useful by would-be identity thieves.[270]

A single piece of information may fall within more than one of the categories listed above. Health diagnoses may be inherently sensitive because of social stigma but also inferentially sensitive if it suggests another trait about a person. Home address information can be not only instrumentally sensitive to stalkers, but also potentially inferentially sensitive if the address suggests somebody who lives on the proverbial "wrong side of town."

III. THE FUTURE OF SENSITIVE INFORMATION

The analysis of Part II, and in particular the four-factor test for identifying sensitive information, can be used to address the problems of inconsistency and anecdotal development presented in Part I. Once we understand that the classification of sensitive information requires us to accurately calculate the probability of harm, we can develop new approaches to assessing these probabilities rigorously in order to help us decide whether to add or subtract new forms of information into the category.

To begin modeling privacy harm rigorously, we should look to the computer security threat modeling literature. This literature offers a rigorous framework for identifying and assessing threats, and others both within and outside the law have begun to apply its lessons to assessing alleged threats to privacy. I synthesize and extend much of this work in this Part.

Rigorous threat modeling will support calls to enact new laws to protect categories of sensitive information, including geolocation, biometric, and metadata information; limit the cases in which we conflate confidentiality and sensitivity (health information at risk should be

270. Bodah v. Lakeville Motor Express, Inc., 649 N.W.2d 859, 862–63 (Minn. Ct. App. 2002) (finding that SSNs "facilitate access by others to many of our most personal and private records and can enable someone to impersonate us to our embarrassment or financial loss.").

protected regardless who holds it); and extend privacy law to cover sensitive information stored in large, unstructured databases, at least when we have reason to believe the sensitive information may easily be extracted.

A. BUILDING THREAT MODELS TO IDENTIFY PRIVACY HARMS

Part I criticized the type of risk assessment we tend to use today to assess whether a particular form of information poses a sufficiently high risk of harm.[271] Legislatures and judges favor easy-to-measure harm over harms that might be hard to quantify but are nevertheless significant and concrete. Throughout the enterprise of risk assessment about information, too many people favor anecdote and intuition over rigorous analysis.

To remedy all of these problems, and to give lawmakers, judges, and policymakers a richer set of tools with which to consider privacy harm, we should embrace what computer security experts call threat modeling. We can build threat models to characterize the kind of harms that might lie latent in a database of personal information with precision, rigor, and concreteness.

1. Threat Modeling: Borrowing from Security

Computer security experts build threat models to enumerate and prioritize security risks to a computer system.[272] Although "threat modeling" can mean different things,[273] one good working definition is "the activity of systematically identifying who might try to attack the system, what they would seek to accomplish, and how they might carry out their attacks."[274] Threat modeling is brainstorming about a system trying to find ways to subvert the goals of the system designers.

Everybody builds threat models, even if only informally.[275] Adam Shostack gives as an example the highway driver, working out how fast and recklessly to drive, factoring in the "threats" of the police, deer, or rain.[276] But computer security experts have developed tools and formal models that make their threat modeling seem very different from everyday

271. *Infra* Part II.C.1.

272. SHOSTACK, *supra* note 26, at 111–23; SWIDERSKI & SNYDER, *supra* note 26, at 100; Wu, *supra* note 26, at 1149–51.

273. SHOSTACK, *supra* note 26, at xxii–xxiii (noting at least four definitions for "threat model"); Wu, *supra* note 26, at 1149 (noting two definitions for "threat modeling").

274. Wu, *supra* note 26, at 1149.

275. SHOSTACK, *supra* note 26, at xxii.

276. *Id.*

threat modeling, deploying tools and complex methodologies, building things called "attack trees" and murmuring acronyms like STRIDE and DREAD.[277] But the formal veneer should not obscure the fact that threat modeling is just brainstorming for pessimists; when done well, it is a formal and comprehensive game of "what-if" focused on worst-case scenarios.

Computer experts have used threat modeling techniques primarily to assess and improve security, not privacy. They build threat models to identify and prioritize the steps needed to secure systems against hackers, scammers, virus writers, spies, and thieves, for example.[278] Recently, scholars from opposite sides of the law-technology divide have begun to adapt threat modeling for privacy too. From the law side, scholars, most prominently Felix Wu, have talked about building threat models for privacy law. Wu has constructed the unstated, implied threat models of existing privacy law, trying to study statutory text or common law court opinions to reveal the implicit threat lawmakers and judges held in mind when they created the laws.[279] He also tries to find "mismatches," where the implicit threat model of a privacy law does not seem to address real world concerns.[280]

From the technology side, computer scientists such as Mina Deng[281] and Adam Shostack[282] have asked whether the rigorous and well-documented threat models for security might be extended to privacy. Just as we build attack trees to decide where to prioritize scarce computer programming resources to shore up security, so too can we build attack trees to decide how best to tackle possible privacy harms.

This Article continues this work, bringing these two still quite distant strands of scholarship closer to meeting in the middle. It extends the work of the legal scholars like Wu by bringing even more substance and rigor to the work of threat modeling. And it extends the work of the computer scientists, by bringing in more nuances to the question of adversaries and threat models and by focusing on threats for which technological fixes fail and legal fixes are needed. All of this marks a vast improvement from the

277. SHOSTACK, *supra* note 26, ch. 1.

278. Threat modeling is also used to improve security in noncomputer contexts. Much of what the Transportation Safety Administration does around airport security, for example, aids the security of a physical system.

279. Wu, *supra* note 26, at 1149.

280. *Id.*

281. Mina Deng et al., *A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements*, 16 REQUIREMENTS ENG'R 3 (Mar. 2011).

282. SHOSTACK, *supra* note 26.

informality with which the identification of sensitive information has been approached to date.

2. Threat Models for Privacy Harm

Although the threat model framework developed can be used to analyze any privacy situation, it is important to once again state the context under immediate focus. Given a particular category or class of data that tends to be stored in databases and might lead to harm, should we enact or amend a privacy law to classify that class of data as "sensitive," subjecting it to restrictions on how that kind of data may be collected, distributed, or used? I will focus on proposals to create new laws or amend old laws, which is different from Wu's primary focus on revealing the implicit threat models of extant laws. My threat modeling process comprises four phases: threat enumeration, risk assessment, nonlegal remediation, and finally legal remediation. Many of these steps bleed into one another, and the model is not meant to be applied rigidly.

a. Step One: Enumerating Adversaries and Harms

The first phase is threat enumeration, which requires us to enumerate two things: adversaries and harms. "Adversary" is the term used to describe the human actor who might use the information being studied to cause another person harm. Wu describes three characteristics of the adversary: goals, tools, and capabilities (e.g. sophistication and computational power).[283]

We turn next to privacy harms, which computer scientists might refer to as the "impact" of the adversary's misuse of the information.[284] This is where the analysis of the history of privacy law and the taxonomy enter the discussion. Recall that Part II talked about three kinds of harms: ancient, traditional, and modern. We have tended to focus most of our attention to date on ancient harms.

Privacy threat modeling incorporates contemporary debates about privacy harm. It does not try to short-circuit them. One who believes that privacy law should protect us against all ancient, traditional, and modern privacy harms can build an accordingly broad threat model. One who

283. Wu, *supra* note 26, at 1148. Wu also discusses the adversary's "background information," which is especially relevant to the problem of reidentification but may not be generally applicable to other privacy harms. *Id.*

284. SHOSTACK, *supra* note 26, at 112 (describing Dan Solove's privacy problems taxonomy and concluding that it describes "harms [that] are analogous to threats in many ways, but also include impact").

would restrict privacy law to ancient harms (or even only a subset of ancient harms) can build a correspondingly narrow model. Which choice is "correct" is external and prior to the process of threat modeling itself and gives those with differing viewpoints a coherent framework for having their debates.

For as much as this Article has criticized the casual use of anecdote thus far, it anticipates an important role for anecdote in step one. As "brainstorming for pessimists," threat modeling cannot escape turning to anecdote in the search for adversaries and harms. Knowing that a specific adversary has committed a specific harm, even if only once, gives useful context to this step.[285] And it is far better to brainstorm based on a documented case instead of a hypothetical case or pure folklore.

As we enumerate both adversaries and harms, we should resist two, polar-opposite, competing temptations: focusing on impossible adversaries and harms or dismissing adversaries and harms too quickly as unlikely. In other words, we will waste time and come to bad conclusions if we imagine a world populated by all-powerful "superusers" who can bend software to their wills.[286] But we will under-protect if we too often dismiss threats as unrealistic. To square these concerns in tension, our discussion of the adversary should be grounded and realistic but also creative and broad. Adversaries think outside the box, and so should we. At this step, we should resist the urge to focus too much about incentives and motives, which we give full consideration to in step two.[287]

But if in doubt, we should err on the side of potential overprotection and consider a potential adversary and harm, even if it is a little implausible. For one thing, we will have another opportunity to weigh likelihood in step two, at which time we can cull possibilities that are too unlikely to merit response. For another, Shostack highlights the importance of thoroughness and creativity at this phase.[288] Given the complexity and dynamism of modern information systems, it helps to enumerate a comprehensive list. If we instead forego thinking about a particular type of harm as "unlikely" or claim "nobody would ever do this," we will miss

285. If the anecdote is a true aberration, it will be discounted as such in step two.

286. Paul Ohm, *The Myth of the Superuser: Fear, Risk, and Harm Online*, 41 U.C. DAVIS L. REV. 1327 (2008); SHOSTACK, *supra* note 26, at 12 ("Focus on feasible threats").

287. A. Michael Froomkin, *The Metaphor Is the Key: Cryptography, the Clipper Chip, and the Constitution*, 143 U. PA. L. REV.709 (1995) (describing "threat assessment" as "long-established intelligence approach in which one assumes the worst about everyone and attempts to measure their capabilities for harm without regard to their likely or actual motives").

288. SHOSTACK, *supra* note 26, at 150.

likely harms.

At the completion of the first phase, we will have a list of adversaries and the harms they might cause. Computer scientists use the evocative phrase, "misuse cases," analogs to the more widely known "use cases" for a system.[289] Going forward, policymakers should develop as a first step a long list of misuse cases for any data nominated for protection as sensitive information.

b. Step Two: Risk of Harm

In step two, we focus on prioritizing the threats we have enumerated in step one. Our goal should be to try to quantify risk. The question to ask is, how likely is this misuse case to occur? It is alluring to imagine we can do this precisely, calculating risk with the precision of a weatherperson or actuary. "This data will result in privacy harm X 0.25% of the time," we would like to say. But we cannot, for a host of reasons.

First, it may be impossible to be precise. A clear trend in the computer security threat modeling literature is to eschew attempts to be unduly quantitative about security risk.[290] Computer security threat modelers have recognized that calculating the risk of something that depends so much on human behavior is a tricky proposition. Ideally, computer security threat models will point to documented cases of completed harm or studies that measure a rate of harm. But anecdotes can be helpful, assuming we keep the limitations of anecdote in mind.

The fact that computer security experts have given up on precision should give us pause because in many ways their task is more straightforward than what faces the privacy threat modeler. Security experts tend to model threats to a specific piece of software or network architecture: what is the threat model for SSL or Windows 10 or Twitter? In contrast, with privacy, we are focused instead on categories or types of information—what is the threat model for precise geolocation information or remote biometric information? If security threat models lack precision about risk calculations, privacy threat models must be even less precise.

Second, we should take into consideration not only the risk of a particular harm occurring but also the magnitude of that harm. When the

289. These are also sometimes referred to as "anti-goals."

290. SHOSTACK, *supra* note 26, at 181 ("Probability assessments in information security are notoriously hard to get right."); GAO, INFORMATION SECURITY RISK ASSESSMENT: PRACTICES OF LEADING ORGANIZATIONS at 8 (Nov. 1999) ("[L]ack of reliable and current data often precludes precise determinations of which information security risks are the most significant and comparisons of which controls are the most cost-effective.").

harm identified is slight, only a large risk will suggest the need to intervene. On the other hand, when the harm is high, we should respond even to avoid slight risks.

Third, we need to build into our model of risk the way evolving technology can alter the probabilities. If the misuse case is not a very realistic risk today but seems to be much more likely in the future, we might act proactively to regulate now before waiting until it is too late.

The path to GINA provides a good case study of reasoning about the changing nature of risk. At the time it was first proposed, whole genome sequencing was a futuristic possibility.[291] Even at the time it was enacted and still today, scientists have not yet connected particular genetic sequences to conclusions we might consider harmful, such as a propensity to develop a maligned disease, so the harm is still best characterized as futuristic.[292] But Congress recognized the inevitability that the "structural constraints" protecting our privacy today were likely to fall in the near future and worked proactively.[293]

To summarize the second step: we should undertake careful but necessarily limited risk assessment of each misuse case. If the harm is extremely severe, it probably makes sense to classify the category of information as sensitive and thus subject to some regulatory control. Even when the harm is less severe but nevertheless significant, if a technical analysis suggests that the misuse case is very likely to occur, we should classify the information as sensitive. If the harm is both not severe and unlikely to occur, then a sensitive information law is probably not warranted.

c. Step Three: Nonlegal Responses and Remediation

The third step enshrines an important value: a bias against regulation. Sometimes, technology, markets, and social institutions can and will successfully prevent (or significantly limit) privacy harms, and in those cases, we should resist the temptation to intervene legally. This choice reflects not an unthinking libertarianism but instead represents a more modest recognition of the "second best" nature of legal solutions.

The question for step three is, short of law, how might we respond? It is crucial that this step involve not only technological but also economic and social considerations. Just because a privacy enhancing technology

291. S. Rep. No. 110-48, at 7 (2007).
292. H.R. Rep. No. 110-28, at 25 (2007).
293. *Id.* at 30; Harry Surden, *Structural Rights in Privacy*, 60 SMU L. REV. 1605 (2007).

would work in theory is not enough, if market forces or social acceptance might prevent that technological fix from being implemented.

The privacy literature is replete with proposals for fixes that fit into step three. From the engineering side, this includes the wide variety of privacy-enhancing technologies that have been proposed and built. Some of these require the user to exercise self-help, such as TOR[294] or disk encryption. Others can be built and implemented by service providers, such as SSL.[295]

From the organizational side, we should consider the many principles of accountability that have been proposed over the years.[296] Some of these are hopelessly vague and amount mostly to empty slogans, such as "accountability." Others teeter on the brink of emptiness such as "Privacy by Design," yet might amount to something with more teeth given the many who are focusing on the concept.[297]

Into this category we should also place self-regulatory efforts. This includes attempts to set technical standards to improve user control, from P3P[298] to Do Not Track[299] to the Department of Commerce's efforts to develop codes of conduct around privacy.[300] It also includes third party certification programs, such as TrustE.[301]

Any one of these solutions should be weighed in the narrow context of a particular form of information. How likely is a given proposal to successfully limit a form of likely harm, taking into account not only technical possibility but also the psychology and economics around adoption of the solution?

294. Tor Project, https://www.torproject.org (last visited Aug. 23, 2014).

295. Christopher Soghoian, *Caught in the Cloud: Privacy, Encryption, and Government Back Doors in the Web 2.0 Era*, 8 J. ON TELECOMM. & HIGH TECH. L. 359, 378 (2010).

296. CENTER FOR INFORMATION POLICY LEADERSHIP, DATA PROTECTION ACCOUNTABILITY: THE ESSENTIAL ELEMENTS (Oct. 2009) http://www.huntonfiles.com/files/webupload/CIPL_Galway_Accountability_Paper.pdf.

297. Ira S. Rubinstein & Nathaniel Good, *Privacy by Design: A Counterfactual Analysis of Google and Facebook Privacy Incidents*, 28 BERKELEY TECH. L.J. 1333 (2013).

298. William McGeveran, Note, *Programmed Privacy Promises: P3P and Web Privacy Law*, 76 N.Y.U. L. REV. 1812 (2001).

299. W3C, Tracking Preference Expression (DNT) Working Draft, http://www.w3.org/TR/tracking-dnt/ (Apr. 24, 2014).

300. NTIA, Privacy Multistakeholder Process: Mobile Application Transparency (Nov. 12, 2013), http://www.ntia.doc.gov/other-publication/2013/privacy-multistakeholder-process-mobile-application-transparency.

301. TRUSTe, http://www.truste.com (last visited Aug. 24, 2014).

d. Step Four: Crafting the Regulatory Response

Finally, we come to the fourth and final step, when we have concluded that a significant risk of harm cannot be mitigated without a legal solution: crafting the law to mitigate the risk of harm. Once again, this step is outside the scope of this Article, which focuses solely on the decision to label a type of data as sensitive or not, but I will say a few words about mitigation through law.

It is important to try to tailor the regulatory obligation to the identified harm. Thus, the many data breach notification laws enacted by states over the past decade tailored a potential privacy harm—risk of identity theft—into an obligation to notify. Although these laws are not without critics, many think the added publicity has encouraged norms and practices that have protected user privacy in many ways.[302]

HIPAA, by contrast, is premised on a much more significant set of harms and higher risk, and it accordingly creates much more onerous obligations on regulated entities. But this Article stops short of laying out all of the necessary pathways from identification of sensitive information to the design of privacy regulation. This deserves much more attention.

B. WHAT NEW CATEGORIES OF INFORMATION SHOULD WE CONSIDER SENSITIVE?

We are now ready to put our threat models into action. Armed with a better understanding of the meaning of sensitive information and a rigorous threat modeling methodology, we can revisit the problems arising in the context presented at the beginning—the privacy of information stored in large electronic databases.

Consider again the three candidates for entry into the ranks of the sensitive: geolocation, metadata, and remote biometric information. For each, rather than merely turn to anecdote and await a sensationalistic story of misuse, consider what a rigorous threat modeling approach will reveal. This analysis will reveal different arguments for treating these categories as sensitive, some that will spur us to act today, others that suggest we wait and see.

302. Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247 (2011).

1. Geolocation Information: The Insider Threat

Geolocation seems poised to be classified as sensitive. Like a home or work address, geolocation information can be used by stalkers to prey on victims.[303] Anecdote suggests that women and children are most often the victim of this kind of threat.[304] Addresses have been protected as sensitive in the DPPA and COPPA for this reason.[305]

Step one of the threat modeling process instructs us to enumerate threats tied to theories of harm in order to develop misuse cases. Wireless phone providers have long maintained a historical record of each customer's cell tower registrations, which an adversary can use to reconstruct a fairly accurate historical record of a person's geolocation.[306] Three recent shifts have increased the quantity and quality of geolocation information stored.

First, all new cell phones come with a GPS chip that can calculate location even more accurately than cell tower trilateration.[307] Second, smartphone operating systems make location available to app developers through an API.[308] Sometimes apps coax users to share their location voluntarily,[309] and sometimes apps take location much to the surprise of their users.[310]

Third, the wireless providers have started for the first time to monetize user location information. Verizon changed its privacy policy to inform its

303. Kate Murphy, *Web Photos That Reveal Secrets, Like Where You Live*, N.Y. TIMES (Aug. 11, 2010), http://www.nytimes.com/2010/08/12/technology/personaltech/12basics.html.

304. Danielle Keats Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 MICH. L. REV. 373, 385 (2009).

305. Driver's Privacy Protection Act of 1994, 18 U.S.C. §§ 2721–2725 (2012)); Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6502 (2012).

306. ECPA Reform and the Revolution in Location Based Technologies and Services: Hearing Before the Subcomm. on the Constitution, Civil Rights, and Civil Liberties of the H. Comm. on the Judiciary, 111th Cong. 20, 21 (2010) (statement of Matt Blaze, Associate Professor, University of Pennsylvania).

307. Renee McDonald Hutchins, *Tied Up in Knotts? GPS Technology and the Fourth Amendment*, 55 UCLA L. REV. 409, 417–18 (2007).

308. Google, Android Developers: Location APIs, https://developer.android.com/google/play-services/location.html (last visited Aug. 27, 2015); Apple, iOS Developer Library: Core Location Framework Reference, https://developer.apple.com/library/ios/documentation/CoreLocation/Reference/CoreLocation_Framework/_index.html (last visited Aug. 27, 2015).

309. Peppet, *supra* note 123, at 1171.

310. Jason Hong, *Analysis of Most Unexpected Permissions for Android Apps*, JASON HONG'S CONFABULATIONS (Nov. 30, 2012), http://confabulator.blogspot.com/2012/11/analysis-of-top-10-most-unexpected.html.

users that location information would be used in new ways.[311] To allay concerns, the company and its corporate partners tout rather innocuous and useful examples such as distributed monitoring of highway traffic conditions.[312]

The sum of these three changes is that more precise geolocation information is collected about more people, stored for a longer amount of time, distributed to a broader group of people, and used for more purposes than ever before. Strong economic incentives are spurring each of these results. If we had reason to worry in the past about stalkers knowing a single, generalized piece of location information—say home or work address—about their prey, we should worry much more about them being able to access discrete and granular records of a person's comings and goings stretching back years.[313] And we should recognize misuse cases beyond stalkers, such as insurers and employers checking up on workers who claim to be recovering from on-the-job injuries, people tracking their spouses, and angry people spoiling for a fight.[314]

After enumerating these misuse cases, the threat model approach moves to step two: measuring the risk of harm. How often have stalkers taken advantage of location information in electronic databases, and how will this rate change as the availability of information increases?

Stalkers have used GPS and smartphone location technology to track the whereabouts of their victims.[315] Andre Leteve, convicted and sentenced to death for killing his two children, had tracked his wife's location using her phone's GPS.[316] James Harrison killed his five children and himself

311. Anton Troianovski, *Phone Firms Sell Data on Customers*, WALL ST. J., May 21, 2013; Kashmir Hill, *Verizon Very Excited That it Can Track Everything Phone Users Do and Sell that to Whoever Is Interested*, FORBES (Oct. 17, 2012), http://www.forbes.com/sites/kashmirhill/2012/10/17/verizon-very-excited-that-it-can-track-everything-phone-users-do-and-sell-that-to-whoever-is-interested/.

312. Jessica Leber, *How Wireless Carriers are Monetizing Your Movements*, MIT TECH. REV., Apr. 12, 2013, http://www.technologyreview.com/news/513016/how-wireless-carriers-are-monetizing-your-movements/.

313. *See Tell All Telephone*, ZEIT ONLINE (Aug. 31, 2009), http://www.zeit.de/datenschutz/malte-spitz-data-retention.

314. We are focused almost entirely on ancient harms that flow from pervasive monitoring of geolocation, such as stalking. Many have argued that location tracking implicates traditional and modern harms as well. *See, e.g.*, United States v. Jones, 132 S. Ct. 945, 955 (2012) (Sotomayor, J., concurring) ("GPS monitoring generates a precise, comprehensive record of a person's public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations.").

315. Justin Scheck, *Stalkers Exploit Cellphone GPS*, WALL ST. J., Aug. 3, 2010.

316. *Id.*

after using location tracking to determine his wife was having an affair.[317]

But, in this step, we should look to proof of systemic activity that goes beyond mere anecdote. Domestic violence shelters have documented many cases of estranged husbands using GPS to find their secret locations.[318] In one survey, 72 percent of such programs nationwide had helped victims who had been tracked by location technology.[319] A study by the U.S. Department of Justice of data collected in 2006, when far fewer people possessed smartphones, estimated that approximately 25,000 stalking victims reported being tracked by GPS technology.[320]

Almost all of these anecdotes and statistics describe tracking done by outsiders, meaning by people who are not employed by the wireless phone company collecting the information. The tendency in debates about information privacy is to focus almost all attention on the threat from outsiders, hackers, and spies who break into databases in order to steal valuable information.[321]

These discussions neglect an emerging security literature around internal threats.[322] From this, we know that the insider threat is vast. In many companies and other large organizations, dozens if not hundreds or thousands of employees, contractors, service providers, and others are given access to databases containing sensitive information. The rise of the cloud has increased and consolidated the power of the biggest online services, giving companies like Google, Facebook, and Amazon access to vast databases that require huge workforces to maintain. Every one of these insiders poses some risk of using the information to cause harm.[323]

The threat model for insiders differs from the outsider threat model in many important ways. Insiders tend to have specialized knowledge about the victims of their invasion, and specialized knowledge might significantly raise the risk of a successful attack, on anonymity for example.[324] They

317. *Id.*

318. *Id.*

319. Nat'l Network to End Domestic Violence, New Survey: Technology Abuse & Experiences of Survivors and Victim Service Agencies (Apr. 29, 2014), http://techsafety.org/blog/2014/4/29/new-survey-technology-abuse-experiences-of-survivors-and-victim-services.

320. KATRINA BAUM ET AL., BUREAU OF JUSTICE STATISTICS SPECIAL REPORT: STALKING VICTIMIZATION IN THE UNITED STATES 5 (Jan. 2009).

321. Citron, *supra* note 5, at 251.

322. SALVATORE STOLFO ET AL., INSIDER ATTACK AND CYBER SECURITY: BEYOND THE HACKER 69–73 (2008).

323. Peter P. Swire, *Peeping*, 24 BERKELEY TECH. L.J. 1167 (2009).

324. Wu, *supra* note 26, at 1154 ("[I]t can be . . . difficult to protect against disclosure to insiders, who can exploit special knowledge gained through their relationships with a target individual to deduce

also have motives to peek that outsiders interested only in identity theft might not. These people might peek at the information stored about former paramours or ex-spouses.[325] They may search through the data to learn things about celebrities or politicians or notorious figures. Considering factors like these, Felix Wu concludes that "[p]rotecting against privacy insiders may therefore require far greater restrictions on data release than protecting against outsiders."[326]

The evidence suggests further that the insider threat cannot be dismissed as an edge case, but instead happens at high enough levels to reflect significant concern.[327] For example, many have documented how employees of government agencies and hospitals are often caught engaging in peeping, particularly through celebrity records.[328] Violations like these happen inside Silicon Valley giants as well.[329]

Turning to step three, it is naïve to assume that even if insider curiosity is inevitable, the insider threat can be significantly mitigated through access controls and other forms of IT security. Computer security experts have long recognized the difficulty of designing a system that grants or denies access to data based on purpose or motive or intent.[330] As legal scholar Derek Bambauer puts it, "it is not possible to enforce selective access, so that your dermatologist can see information about your sunscreen use but not your antidepressant use."[331]

Finally, we can quickly dispense with the antiquated notion that data anonymization will let companies extract utility from this data while also preventing the kind of harm I am describing here. As I have said, "data can be useful or perfectly anonymous, but never both."[332] Even if anonymization holds a faint hope of promise in some contexts, recent

more about that individual from released data than the general public would.").

325. Edward Moyer, *NSA Offers Details on 'LOVEINT'*, CNET (Sept. 27, 2013), http://www.cnet.com/news/nsa-offers-details-on-loveint-thats-spying-on-lovers-exes/.

326. Wu, *supra* note 26, at 1154.

327. Shaun Spencer, *Security vs. Privacy*, 79 DENV. U. L. REV. 519 (2002) (citing several statistics about incidence of insider breach and reciting common rule of thumb that "one percent of an organization's staff will always 'be willing to sell or trade confidential information'").

328. Charles Ornstein, *Hospital to Punish Snooping on Spears*, L.A. TIMES, Mar. 15, 2008.

329. Adrian Chen, *GCreep: Google Engineer Stalked Teens, Spied on Chats*, GAWKER (Sept. 14, 2010), http://gawker.com/5637234/gcreep-google-engineer-stalked-teens-spied-on-chats (reporting Google fired a Gmail engineer for peeking into the email messages of underage teens.).

330. *See also* Wu, *supra* note 26, at 1154 ("In security threat modeling, analysts regard insider attacks as 'exceedingly difficult to counter,' in part because of the 'trust relationship . . . that genuine insiders have.'").

331. Bambauer, *supra*, note 3, at 669.

332. Ohm, *Broken Promises of Privacy*, *supra* note 124, at 1704.

research suggests that the technique is particularly fruitless at masking detailed tracks of location information.[333]

The sum result of this analysis is clear: we need new laws protecting precise geolocation information. The harm we are trying to avoid is plain: fear and injury from stalkers. The risk seems high and irreducible once we account for the precision of the data, the vast amounts of information that are stored, and the intractable problem of preventing insider breach.

2. Metadata: Restricting Government Access?

In 2013, Edward Snowden revealed that the NSA is conducting periodic bulk downloads of call records for millions of Americans, the vast majority of whom are not under suspicion of any terrorist activity.[334] This has sparked a vigorous debate over whether metadata should be considered sensitive, particularly because President Obama has argued repeatedly that metadata is far less subject to abuse and harm than other communications data.[335]

Rather than focus on the telephone number call records that have been the focus of attention, I want to focus on another form of metadata currently being tracked: the record of URLs visited by users as they read websites across the web. Begin with step one of the threat modeling framework, enumerating the threat of harm. Neil Richards has written extensively about the harms that befall individuals whose reading habits are placed under constant surveillance.[336] The government desperately wants access to the trove of URL information currently being tracked by websites, ISPs, social networking services, and advertisers. The harm of allowing the government to access this kind of information is suggested by past events.[337] History is littered with government attempts to monitor thoughts by looking at reading habits.[338] The FBI conducted a well-documented wiretapping campaign against Martin Luther King, Jr.[339] The contemporary equivalent may be the NSA's assertion that it is entitled to

333. Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, 3 NATURE SCI. REP. 1376 (2012) (concluding four location points sufficient to uniquely identify individuals in large database of cell phone tracks).

334. Greenwald, *supra* note 9.

335. William Saletan, *Meta Man*, http://www.slate.com/articles/news_and_politics/frame_game/2013/06/nsa_metadata_obama_s_non_answers_to_questions_about_government_surveillance.html (last visited Aug. 27, 2015).

336. *See, e.g.*, Richards, *Intellectual Privacy*, *supra* note 81; Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934 (2013) [hereinafter Richards, *Dangers*].

337. Richards, *Dangers*, *supra* note 336, at 1953–55.

338. *Id.*

339. SOLOVE, *supra* note 10, at 8.

pore through the records of communications of anybody within "three hops" of a suspected terrorist.[340] Monitoring the surfing habits of millions of people cannot be justified by the unquestionable need to detect terrorist plots before they occur.

We should thus strengthen protections limiting government access to databases that record what we read online. We should, for example, strengthen the Electronic Communications Privacy Act ("ECPA") to require more proof for access to this kind of information.[341]

But the much more difficult question is: should we limit private party access to what we read as an attempt to limit government access? In other words, in order to prevent the NSA from obtaining information, should we restrict what Google or Verizon can store? Once again, the threat modeling framework helps us answer that question rigorously.

Under step two, we need to decide whether this threat—government access to location trail data stored by a private company—is likely to occur. Unfortunately, too many advocates who debate the threat of government access to sensitive information stored by commercial third parties, from both sides, often possess very little reliable knowledge or even a decent understanding about the motivations of government agents. Federal law enforcement agents[342] are probably neither as cunning nor as innovative as privacy advocates imagine nor as plodding or hidebound as company advocates suggest. Instead, law enforcement agents practice a tradecraft, an evolutionary honing of skills and techniques.

On today's web, numerous third parties such as online advertising networks track users as they travel from website to website.[343] Over time, these third parties can amass impressive digital dossiers about the Internet habits of thousands or millions of individuals.[344] Some claim that these dossiers are appealing targets for government search warrants and subpoenas, one stop shopping for agents who want to retrace a person's reading and viewing habits and communications patterns.[345]

But how often do law enforcement agents subpoena this information

340. Spencer Ackerman, *NSA Warned to Rein in Surveillance as Agency Reveals Even Greater Scope*, THE GUARDIAN (July 17, 2013), http://www.theguardian.com/world/2013/jul/17/nsa-surveillance-house-hearing.

341. Richards, *Intellectual Privacy*, *supra* 81, at 423.

342. Because my direct professional experience is in law enforcement (both criminal and civil), I will focus on this type of investigation rather than on national security.

343. Angwin, *The Web's New Gold Mine*, *supra* note 5.

344. *Id.*

345. *E.g.*, Richards, *Intellectual Privacy*, *supra* note 81, at 437–41.

from these companies? It probably has never happened. Not even once. At least given today's conditions, FBI agents and other law enforcement officials probably do not think it would be worth it to pursue such avenues. Anybody claiming that the risk of law enforcement access to third party advertiser records is high, either today or in the near future, is probably mistaken.

But one would make a more serious mistake to conclude that such a risk could not increase dramatically, and virtually overnight, for one simple reason: third party advertiser digital dossiers contain information that is probably difficult to gather from any other single source.[346] Law enforcement agents may not today understand the investigatory goldmines these dossiers represent, but in time, they probably will.

All it would take is for one creative law enforcement agent to sit down with one talkative third party advertiser for there to spring forth a new tool in the agent's toolkit. The agent would come to understand that the biggest third party advertising networks watch each user or target across hundreds, if not thousands, of different websites.[347] This kind of cross-site trail is very difficult to obtain from any other single source on the Internet and very likely relevant to many different crimes, from the distribution of child pornography, to criminal copyright infringement, to theft of trade secrets, and even to terrorist activity.[348] A smart FBI agent would realize that with a single unique ID stored in a cookie from a seized laptop or smartphone, a subpoena or warrant to a third party advertiser would yield months if not years of past web surfing behavior.[349]

Something very similar has happened before. Not too long ago, the FBI rarely, if ever, solicited the vast stores of precise geolocation information held by cell phone providers, even though these have been available for many years.[350] Today, spurred by the value and great detail of the information in this kind of data, FBI tradecraft has shifted. Cell phone providers now routinely receive government subpoenas and warrants.[351]

346. Angwin, *The Web's New Gold Mine*, *supra* note 5.

347. TUROW, *supra* note 61, at 65–88.

348. One possibility, of course, is from a target's upstream ISP. Traditionally, ISPs have not kept pervasive log files of user activity across the web. But as I have written, business motivations are enticing them to change this historical reticence. Ohm, *ISP Surveillance*, *supra* note 5.

349. *Cf.* Ashkan Soltani et al., *NSA Uses Google Cookies to Pinpoint Targets for Hacking*, THE WASH. POST (Dec.10, 2013), http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/nsa-uses-google-cookies-to-pinpoint-targets-for-hacking/ (describing the way in which the NSA leverages Google cookies to aid target surveillance).

350. Hutchins, *supra* note 307.

351. *Verizon Transparency Report for the First Half of 2014*, VERIZON (2014),

And once a single FBI agent adds this tool to his toolkit, he will propagate it agent-to-agent across the national social network of FBI agents, with great speed. FBI agents and prosecutors constantly trade stories of tradecraft, around the office, at conferences, and through electronic communications.

Next, we can dispense with step three in short order. Companies can prevent government access technologically only if they encrypt all of their data and give the key only to their users.[352] But doing so deprives the companies themselves of access to the data, something they resist doing for both quality assurance and monetization reasons. Technical access controls are almost entirely irrelevant to this question.[353]

We are left to conclude that the massive databases of URLs collected by advertisers and ISPs will likely fall into government hands some day, and probably on a relatively large scale. But once again, whether this is a reason to try to limit URL collection by private parties in the first place is a much more debatable question.

I conclude that the risk is high enough that we should consider laws limiting URL collection. At the very least, it should justify unnecessary collection, as many companies probably hold onto more URL information than they need to justify their business model, clinging to the remote possibility that they find a new profitable use for it some day.

---

http://transparency.verizon.com/us-report (reporting more than 300,000 requests from U.S. law enforcement).

352. This is analogous to Apple's recent decision to encrypt the storage on iOS 8 devices (iPhones and iPads), which has been criticized by law enforcement officials. Sam Frizell, *The FBI and NSA Hate Apple's Plan to Keep Your iPhone Data Secret*, TIME (Sept. 27, 2014), http://time.com/3437222/iphone-data-encryption/.

353. Everything said in the prior section about the underappreciated insider threat essentially applies to government access to privately generated and held data. As noted computer security scholar Ed Felten puts it, "a court order is an insider attack." Ed Felten, *A Court Order Is an Insider Attack*, FREEDOM TO TINKER BLOG (Oct. 15, 2013), https://freedom-to-tinker.com/blog/felten/a-court-order-is-an-insider-attack/. He compares an employee responding to a government subpoena to a faithless employee stealing data to give to a drug cartel:

> From a purely technological standpoint, these two scenarios are exactly the same: an employee copies user data and gives it to an outside party. Only two things are different: the employee's motivation, and the destination of the data after it leaves the company. Neither of these differences is visible to the company's technology—it can't read the employee's mind to learn the motivation, and it can't tell where the data will go once it has been extracted from the company's system. Technical measures that prevent one access scenario will unavoidably prevent the other one.

Id.

3. Remote Biometric: The Lessons of the Social Security Number

Recall that remote biometric information raises two different possible privacy harms. Armed now with the language of threat modeling, we can speak of these as two separate misuse cases: using biometric data to track an individual's location and using biometric data to spoof identity. The first category—location tracking—raises most of the same issues posed in the discussion of precise geolocation information. The analysis below thus focuses on the threat to identity.

Turning to step one, remote biometric information is tied to ancient harms, namely the harm of identity theft.[354] Nobody suffers a traditional harm—say humiliation or abasement—when his facial print or gait print becomes known. Remote biometric information thus operates like social security numbers or signatures, frequent targets for today's identity thieves.[355]

This is a classic example of instrumentally sensitive information. As with so many other forms of information in this category, the sensitivity turns on how we choose to use the information. Ten years ago, almost nobody used faceprint authentication for securing access to spaces or information.[356] At that time, it would not have made sense to classify faceprints as sensitive. Now, as we grow increasingly frustrated with traditional passwords and seek simpler, more seamless access controls, fingerprints and faceprints are used more often, sometimes to secure very valuable spaces.[357] System architects are thus making decisions that produce the side effect of increasing the sensitivity of this information.

A few conclusions flow from this observation. First, as policymakers weigh the sensitivity of remote biometric information, they must chase a moving target. Biometric information is as sensitive as the information it protects, so as more entities begin to embrace biometric access as a means to protect valuable information, the more harm unauthorized access might bring and thus the more sensitive remote biometric information becomes. Policymakers should weigh not only current implementations but also plans for future implementations.

It is difficult to undertake a detailed step two analysis measuring the

354. Alessandro Acquisti, Facial Recognition Study — FAQ, http://www.heinz.cmu.edu/~acquisti/face-recognition-study-FAQ/ (last visited Aug. 27, 2015).

355. Daniel J. Solove, *Identity Theft, Privacy, and the Architecture of Vulnerability*, 54 HASTINGS L.J. 1227 (2003).

356. Donohue, *supra* note 96.

357. *Id.*

likelihood of the harms identified in step one because so much depends on the future evolution of the design of biometric systems. We should turn in step three, therefore, to the ways in which system designers can make step three choices that mitigate the risk of harm.

System designers can use biometric information in ways that limit its sensitivity. Doing so may not only put fewer individuals at risk of identity theft but also avoid exposing one's industry to privacy regulation. Even if it means a more cumbersome process for users and a less streamlined process for the company, these tradeoffs may be worth it to avoid creating new categories of sensitive information.

Would-be purveyors of remote biometrics might learn a lesson from the plight of the social security number.[358] It was not the creation of a unique identifier for almost every American that created a new sensitive category; it was the way government and nongovernment actors began using the identifier as a password for access to benefits, proof of identity, and more.[359] If these actors had treated the social security number as a unique number but not as a password, it would have been less useful for identity theft and thus less in need of protection as sensitive information.[360]

C NEW DIRECTIONS FOR SENSITIVE INFORMATION

In addition to helping us induct new categories of information into the ranks of the sensitive, threat modeling points to other ways we might broaden our traditional approaches to privacy law. First, we tend to conflate privacy and confidentiality, obligating the protection of some types of sensitive information only when held by certain types of entities. For example, under HIPAA, we protect the undeniably sensitive category of health information only when held by hospitals, doctors, and their service providers. Threat modeling suggests we should expand laws like HIPAA, to cover any entity that intentionally collects health information. Second, we tend not to cover within privacy law sensitive information "hidden" within larger pools of data, intermixed with nonsensitive information. For example, Google's massive database of search queries contains health information, yet we do not consider the database itself to be sensitive health information. Although privacy threat modeling has supported this conclusion to date, trends in both the power of data analysis and business

358. Adrianne Jeffries, *Identity Crisis: How Social Security Numbers Became Our Insecure National ID*, THE VERGE (Sept. 26, 2012), http://www.theverge.com/2012/9/26/3384416/social-security-numbers-national-ID-identity-theft-nstic.

359. *Id.*

360. *Id.*

incentives are pushing companies more often to isolate the sensitive from these unstructured databases. This might encourage us to expand privacy law to sometimes cover data such as search query data as sensitive.

1. Sensitive No Matter Who Holds It

No privacy law covers a category of sensitive information regardless of how it is collected or who possesses it. For example, no law regulates the handling of health diagnoses you tell your friend or the behavior of the stranger who finds your misplaced hospital chart on the subway. The EU Data Protection Directive comes close, regulating the activity of all data processors, a broadly defined category, but even it fails to cover those engaged "in the course of a purely personal or household activity."[361] American laws such as HIPAA, FERPA, and GLB are even narrower, limited to particular actors in particular sectors.[362]

These American laws thus act like hybrids, protecting not only privacy, but also confidentiality.[363] This distinction matters particularly in the United States, where the law of confidentiality has suffered a crabbed and narrow development, particularly compared to the way it has been expanded in the United Kingdom.[364]

If sensitive information laws are designed to mitigate the risk of privacy harm, we should reconsider the way we have restricted these laws to narrowly defined classes of covered entities. If we can establish, by applying the threat modeling framework described in Part II, that some categories of sensitive information expose data subjects to significant risk of harm while in the hands of currently unregulated entities, we should broaden those laws to meet these risks.

In other words, insofar as laws like HIPAA, FERPA, and GLB tie risk of harm to the type of entity holding the sensitive information, they reflect an older, probably now (and certainly soon) obsolete view of

361. Council Directive 95/46, art. 8, 1995 O.J. (L 281) 1 (EC).

362. 45 C.F.R. § 160.103 (2014) (HIPAA rule defining "business associate" and "covered entity"); 20 U.S.C. § 1232g (2012) (FERPA covering "applicable programs"); 15 U.S.C. § 6809 (2012) (GLB defining "financial institution").

363. Richards & Solove, *supra* note 262, at 181–82. As I am using the terms, privacy law tends to focus on the nature of the invasion and the possibility of privacy harm; confidentiality law focuses more on the special nature of the relationship between the person handling the information and the subject of the information. *Id.* at 174 ("The public disclosure tort focuses on the nature of the information being made public. By contrast, the focus of the tort of breach of confidentiality is on the nature of the relationship.").

364. *Id.* at 126–27.

information.[365] Several decades ago, high priests in narrowly defined professions were the only ones who had access to certain types of sensitive information. Doctors and nurses held health information, teachers held education records, and banks held financial records, explaining the narrow scope of HIPAA, FERPA, and GLB, respectively.[366] The times have changed, and today, these data flow to many other types of entities. Health vaults like those introduced by Google and Microsoft a few years ago encourage patients to obtain copies from their providers to upload into the cloud.[367] Many smaller companies are engaged in providing devices that monitor health vitals and upload them to the cloud.[368] Sites like Monster.com and LinkedIn encourage users to develop online resumes including education records of the sort protected by FERPA.[369] Mobile apps like Mint provide financial services like account balance tracking, budgeting, and credit reporting that once were the province of banks.[370] These trends suggest we revisit the narrow scope of sensitive information laws, expanding them to cover companies like all of these.

Thus, we might expand HIPAA, FERPA, and GLB by following the model of COPPA. COPPA applies broadly to any "operators of websites and online services,"[371] without further limitation. During the recent round of COPPA rulemaking, the Federal Trade Commission made clear that this definition expands as technology changes, covering not only traditional websites, but also mobile apps, third-party plug-ins, and third-party ad networks.[372]

365. Sharona Hoffman & Andy Podgurski, *In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information*, 48 B.C. L. REV. 331, 360 (2007) (advocating extending HIPAA to currently uncovered entities who use health information "for any business purpose").

366. *Id.*

367. Microsoft, *Microsoft Unveils Consumer Health Vision, Launches Technology Platform to Collect, Store, and Share Health Information* (Oct. 4, 2007), http://www.microsoft.com/en-us/news/press/2007/oct07/10-04healthvaultpr.aspx. Steve Lohr, *Google Offers Personal Health Records on the Web*, N.Y. TIMES (May 20, 2008), http://www.nytimes.com/2008/05/20/technology/20google.html.

368. 23 and me, *Find out What Your DNA Says About You and Your Family*, https://www.23andme.com (last visited Aug. 27, 2015); Fitbit, http://www.fitbit.com (last visited Aug. 27, 2015).

369. *It's Easy to Understand What's Going on With Your Money*, https://www.mint.com/ (last visited Aug. 27, 2015); Eric Eldon, *Mint: The Easiest Way to Manage Your Personal Finances*, VENTUREBEAT.COM (Sept. 18, 2007), http://venturebeat.com/2007/09/18/mint-the-easiest-way-to-manage-your-personal-finances/.

370. Eldon, *supra* note 369.

371. Children's Online Privacy Protection Act of 1998, 15 U.S.C. § 6501 (2012).

372. 16 C.F.R. pt. 312. I worked on these rules while employed as a Senior Policy Advisor for the Federal Trade Commission. Everything I say in this Article reflects public information.

This call to expand privacy law beyond the traditional boundaries of confidentiality builds on the work of scholars who have in recent years called for the expansion of the law of confidentiality to account for shortcomings in the law of privacy.[373] Neil Richards and Dan Solove have recommended expanding confidentiality law in the United States to resemble the expansion that has occurred in the United Kingdom.[374] Woody Hartzog has called for requiring contracts extending obligations of confidentiality to downstream recipients of information collected online[375] and for the expansion of the law of implied confidentiality.[376]

But this Article goes further, arguing that at least in the relatively narrow context of information held in large databases by businesses, the demonstrable likelihood of harm alone justifies extending protection even absent a traditionally recognized confidential relationship.[377] The likelihood of harm itself and alone deems the relationship confidential.[378] In those cases, which threat modeling helps us recognize, we should extend privacy protection regardless of the specific relationship.

2. Unstructured Yet Sensitive

Data in databases can be divided into two broad categories: structured and unstructured.[379] Structured forms of information are narrowly defined, intended to contain only one type of information—"email address" or "diagnoses" or "GPA." Unstructured forms of information are generic, and can be filled at the whim of the person doing data entry—"comments" or

373. *E.g.*, Susan M. Gilles, *Promises Betrayed: Breach of Confidence as a Remedy for Invasions of Privacy*, 43 BUFF. L. REV. 1, 20–25 (1995); Woodrow Hartzog, *Chain-Link Confidentiality*, 46 GA. L. REV. 657 (2012); Hartzog, *supra* note 18; G. Michael Harvey, Comment, *Confidentiality: A Measured Response to the Failure of Privacy*, 140 U. PA. L. REV. 2385, 2396 (1992); Richards & Solove, *supra* note 262; Alan B. Vickery, Note, *Breach of Confidence: An Emerging Tort*, 82 COLUM. L. REV. 1426, 1426 (1982).

374. Richards & Solove, *supra* note 262, at 182 ("American law has much to learn from the English confidentiality tort.").

375. Hartzog, *Chain-Link Confidentiality*, *supra* note 373, at 683 (advocating for "chain-link confidentiality" through contracts).

376. Hartzog, *supra* note 18, at 46–47 (developing a four-factor test for courts assessing claims of implied confidentiality).

377. *See* Richards & Solove, *supra* note 262, at 175 ("The breach of confidentiality tort does not contain a 'highly offensive' requirement, as it views the injury not exclusively in terms of the humiliation caused by the revelation of information but also in terms of the violation of trust between the parties.").

378. *See* Hartzog, *supra* note 18, at 29–32 (recognizing "highly personal information" as a subfactor for courts assessing confidentiality and "information exposing discloser or subject to physical harm" as another).

379. Liane Colonna, *A Taxonomy and Classification of Data Mining*, 16 SMU SCI. & TECH. L. REV. 309, 332–34 (2013).

"notes."[380]

Should privacy law regulate the use of sensitive information "hidden" in unstructured data? Today, some of our most sensitive information ends up amassed in giant, unstructured pools of information kept by tech industry giants. Google holds every search query each of us has ever entered, and many have documented how this represents our "database of intentions," perhaps the largest database of sensitive information on earth.[381] Facebook maintains a similarly sensitive database of past photos, videos, and status updates.[382] And email providers store massive repositories of sensitive messages. But every one of these providers differs from Health Vault and Mint by collecting sensitive information incidentally, not in a targeted matter. And these providers also store this sensitive information commingled with less sensitive and nonsensitive information.

If we did nothing more than expand HIPAA, FERPA, and GLB to cover sensitive information wherever it is found, then these laws might cover the unstructured databases of Google, Facebook, and email providers, subjecting these companies to a dizzying array of new regulation. Current sensitive information laws tend to focus on the type of information deemed sensitive, not the structure of the database in which the data is found. For example, HIPAA applies to "health information," defined in part as information that "[r]elates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual."[383] FERPA covers "education records," which must "contain[] information directly related to a student."[384] COPPA applies to "personal information," which means, in part, information that "permits the physical or online contacting of a specific individual."[385] None of these definitions limit the scope of their respective laws to particular field labels or column types. All three

380. To put the difference in terms familiar to any professor, consider the millions of course evaluations completed by students across the country at the end of every term. These evaluation forms include structured fields, consisting of focused spaces asking students to give the course or professor a numeric grade, and unstructured "comments" areas. The two different types of fields elicit very different forms of information. The structured, numeric grades are simple, focused, and mostly unidimensional. The open-ended comments fields can be much more unpredictable.

381. JOHN BATTELLE, THE SEARCH: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE 7 (2006).

382. James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137 (2009).

383. 42 U.S.C. § 1320d(4)(B) (2012).

384. 12 U.S.C. § 1232g(a)(4)(A)(i) (2012).

385. 15 U.S.C. § 6501(8)(F) (2012).

definitions apply even to information intermingled with other information, such as in an unstructured "comments" field or buried in a memo or letter.

Should we cover within privacy law database owners who possess vast stores of sensitive information but rendered relatively inaccessible in unstructured data? The answer would have been "no" in the recent past, but that answer may be changing.

Not too long ago, one who was bent on exploiting the contents of a structured field of data (like "symptoms") would have faced a much easier task than another trying to exploit the meaning of unstructured data (like "comments"). In most cases, the nonsensitive data found in the unstructured comments field would swamp the sensitive information, at least quantitatively, and the sensitive information could hide in the crowd.[386] Limits in computation, storage, and programming tools and techniques made it unlikely that anybody would have been able to pull the individual out of the crowd.[387] Although a very motivated adversary might have once been able to use unstructured data to cause harm, he would not have been able to systematically harm large groups of people in this way.

This situation seems to be changing rapidly. Thanks to recently-developed and still emerging Big Data techniques, technologists now can much more easily separate the wheat from the chaff in unstructured fields than they could have in the past.[388] They do this with techniques known as web scraping, natural language processing, and machine learning, each of which describes a way of imparting meaning and structure to messy data.[389]

This shift in technological possibility has inspired changes in business models and business incentives, shifts that push companies and others to try to convert unstructured blobs of information into structured fields of data.[390] And when those fields of data reveal previously hidden pockets of sensitive information, we might begin to worry about possibilities of harm in ways we did not worry about not so long ago.

This echoes what scholars have said about the privacy and the

386. This is similar to what James Grimmelmann has said about privacy and first-class objects. James Grimmelmann, *First-Class Objects*, 9 J. ON TELECOMM. & HIGH TECH. L. 421 (2011).

387. Ohm, *ISP Surveillance*, *supra* note 5, at 1422–23.

388. *See* Colonna, *supra* note 379.

389. *See id.*

390. In 2010, the Wall Street Journal reported that the Nielsen Corporation had "scraped" messages from a private online discussion board dedicated to discussions about particular diseases in order to report which pharmaceuticals were receiving the most "buzz." Julia Angwin & Steve Stecklow, *"Scrapers" Dig Deep for Data on Web*, WALL ST. J. (Oct. 12, 2010), http://www.wsj.com/articles/SB10001424052748703358504575544381288117888.

availability of information. Dan Solove has argued that the law should recognize that the digitization and searchability of records that previously existed only as paper copies can lead to new privacy harms.[391] Woody Hartzog has formalized this as the problem of "obscurity."[392] Both apply this reasoning to worry about the move to digitize and publish online court records that once could be obtained only by visiting different county court houses.[393] This is a significant problem because court records hold large amounts of sensitive information, from embarrassing tales of past misconduct, to information that may be used for identity theft, such as social security numbers. Threat modeling lets us formalize the important work of these scholars.

Although the effort required to extract sensitive information from unstructured data has decreased, it has not vanished, so we should hesitate before erasing the legal distinctions between these two categories. Thus, without more, it would probably be unwise to create a law that treated all search queries, status updates, or email messages, as sensitive health, financial, and education records. But when the surrounding context suggests that a company has the tools and incentives to extract the sensitive information, then the sensitive label should apply.

If a company affirmatively takes steps to use techniques like these to parse out sensitive information, then treating that information as sensitive and worthy of regulation makes sense. For example, Google announced in 2008 that it had developed what it called "Flu Trends," a system for monitoring search query logs for evidence of flu outbreaks in concentrated geographic areas. Some of the details of the system have been revealed in a 2009 publication.[394] The Google researchers analyzed "billions of individual searches from 5 years of Google web search logs,"[395] using ingenious machine learning techniques to distinguish flu symptoms from other types of search queries automatically.[396]

Google has thus demonstrated that it can convert billions of individual searches in the form of messy, unstructured data, into a focused database of health symptoms. This might give us a reason to start treating search queries not as unregulated unstructured data, but rather as fully-realized

391. *See* Solove, *supra* note 7.

392. *See* Hartzog & Stutzman, *supra* note 197.

393. *Id.*; Solove, *supra* note 7.

394. Jeremy Ginsberg et al., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012 (2009).

395. *Id.*

396. *Id.* at 1013.

sensitive health information.

CONCLUSION

Changes in technology have introduced new risks of significant and concrete privacy harm, risks that threaten us all. We already have a proper template for laws that protect against these risks and threats: sensitive information laws. We need to create new sensitive information laws and broaden our current laws at least to cover precise geolocation and some forms of metadata but also to go further. We need to do this to respond to a growing threat of harm stemming from advances in technology and evolving business models, forces that create a significant threat of a global database of ruin.[397]

To date, we have used folklore and anecdote to assess the sensitivity of information. It is time to find a more rigorous approach, and threat modeling for privacy seems to be a very promising candidate. By characterizing the adversary—his goals, tools, and capabilities—measuring the risk and the harm, and crafting a response, we can move beyond the anecdotal, informal patterns of discourse on which we too often rely.

397. Ohm, *Broken Promises of Privacy*, *supra* note 124, at 1776–77.