THE PII PROBLEM: PRIVACY AND A NEW CONCEPT OF PERSONALLY IDENTIFIABLE INFORMATION

PAUL M. SCHWARTZ† & DANIEL J. SOLOVE‡

Personally identifiable information (PII) is one of the most central concepts in information privacy regulation. The scope of privacy laws typically turns on whether PII is involved. The basic assumption behind the applicable laws is that if PII is not involved, then there can be no privacy harm. At the same time, there is no uniform definition of PII in information privacy law. Moreover, computer science has shown that in many circumstances non-PII can be linked to individuals, and that de-identified data can be re-identified. PII and non-PII are thus not immutable categories, and there is a risk that information deemed non-PII at one time can be transformed into PII at a later juncture. Due to the malleable nature of what constitutes PII, some commentators have even suggested that PII be abandoned as the mechanism by which to define the boundaries of privacy law.

In this Article, we argue that although the current approaches to PII are flawed, the concept of PII should not be abandoned. We develop a new approach called "PII 2.0," which accounts for PII's malleability. Based upon a standard rather than a rule, PII 2.0 utilizes a continuum of risk of identification. PII 2.0 regulates information that relates to either an "identified" or "identifiable" individual, and it establishes different requirements for each category. To illustrate this theory, we use the example of regulating behavioral marketing to adults and children. We show how existing approaches to PII impede the effective regulation of behavioral marketing, and how PII 2.0 would resolve these problems.

† Professor, University of California, Berkeley, School of Law; Director, Berkeley Center for Law & Technology.

‡ John Marshall Harlan Research Professor of Law, George Washington University Law School. We wish to thank the National Policy and Legal Analysis Network to Prevent Childhood Obesity (NPLAN) for its support for this project. We also wish to thank Shawn Curtis, Melissa DeJesus, Leah Duranti, Yan Fang, Bill Friedman, Matthew Galati, and Brian St. John, who provided research assistance. Marty Abrams, Fred Cate, Jeff Chester, Chris Hoofnagle, Paul Ohm, Jules Polonetsky, Ira Rubinstein, Peter Swire, and participants at the Fourth Annual Privacy Law Scholars Conference, Berkeley, California, provided helpful suggestions on this paper.

INTRODUCTION

Information privacy law has reached a turning point. The current debate about the topic is vigorous, and polling data reveal that Americans are extremely concerned about privacy, both on and off the Internet.[1] Moreover, the Executive Branch, independent agencies,

[1] *See Online Privacy: What Does It Mean to Parents and Kids?*, COMMONSENSE MEDIA (2010), http://www.commonsensemedia.org/sites/default/files/privacypoll.pdf (detailing concerns of adults and youth regarding online privacy); *U.S. Internet Users Ready To Limit Online Tracking for Ads*, GALLUP (Dec. 21, 2010), http://www.gallup.com/poll/145337/Internet-Users-Ready-Limit-Online-Tracking-Ads.aspx (indicating that Internet users support limiting tracking measures that may impinge on their privacy).

and Congress are all considering different paths to revitalize information privacy.[2] Yet, regardless of the nature of potential reforms, there is a deeper problem: Information privacy law rests on the currently unstable category of Personally Identifiable Information (PII). Information that falls within this category is protected, and information outside of it is not.

PII is one of the most central concepts in privacy regulation. It defines the scope and boundaries of a large range of privacy statutes and regulations. Numerous federal statutes turn on this distinction.[3] Similarly, many state statutes also rely on PII as a jurisdictional trigger.[4] These laws all share the same basic assumption—that in the absence of PII, there is no privacy harm. Thus, privacy regulation focuses on the collection, use, and disclosure of PII, and leaves non-PII unregulated.

Given PII's importance, it is surprising that information privacy law in the United States lacks a uniform definition of the term. In addition, computer science has shown that the very concept of PII is far from straightforward. Increasingly, technologists can take information that appears on its face to be non-identifiable and turn it into identifiable data.[5] A recent law review article by Paul Ohm has also challenged PII as a fatally flawed concept. In Ohm's view, privacy law must abandon its reliance on PII and find an entirely new paradigm on which to regulate information privacy.[6]

---

[2] Consider, for example, recent reports by the Department of Commerce and the Federal Trade Commission on online privacy, which suggest that both entities plan to play important and perhaps competing roles in this area. DEP'T OF COMMERCE, INTERNET POLICY TASK FORCE, COMMERCIAL DATA PRIVACY AND INNOVATION IN THE INTERNET ECONOMY: A DYNAMIC POLICY FRAMEWORK (2010); FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (2010) [hereinafter FTC, PROTECTING PRIVACY].

[3] Examples include the Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6506 (2006), the Gramm-Leach Bliley Act, 15 U.S.C. §§6801–6809, the Video Privacy Protection Act, 18 U.S.C. § 2710, and the HITECH Act, Pub. L. No. 111-5, 123 Stat. 226 (2009) (codified as amended in scattered sections of 42 U.S.C.), discussed *infra* Part I.B.

[4] Examples include California's Song-Beverly Credit Card Act, CAL. CIV. CODE § 1747 (West 2009) and the breach notification laws that have now been enacted in forty-six states, discussed *infra* Part I.B. For a discussion of the breach notification statutes, which govern disclosure procedures for data security breaches, see DANIEL J. SOLOVE & PAUL M. SCHWARTZ, PRIVACY LAW FUNDAMENTALS 135–39 (2011). For an up to date listing of these statutes, see *State Security Breach Notification Laws*, NAT'L CONFERENCE OF STATE LEGISLATURES, http://www.ncsl.org/Default.aspx?TabId=13489 (last updated Oct. 12, 2010).

[5] *See infra* Part II.B (discussing ways in which data that is non-PII can be converted to PII).

[6] Paul Ohm, *Broken Promises of Privacy*, 57 UCLA L. REV. 1701 (2010).

In contrast, this Article contends that information privacy law needs a concept of PII, and thus cannot jettison PII as one of its central dimensions. Rather, PII must be reconceptualized if privacy law is to remain effective in the future. Therefore, we develop a new conception, PII 2.0, which avoids the problems and pitfalls of current approaches. The key to our model is to build two categories of PII, "identified" and "identifiable" data, and to treat them differently.[7] This approach permits tailored legal protections built around different levels of risk to individuals. It also represents a path forward, one that avoids both the United States' reductionist view of PII, and the European Union's expansionist view. In the reductionist view, the tendency is to consider PII as only that personal data that has been specifically associated with a specific person. That model protects only identified data, and thereby leaves too much personal information without legal protections. In the expansionist approach, it is irrelevant if information has already been linked to a particular person, or might be so linked in the future; this view treats identified and identifiable data as equivalent. Rather than follow the path of these two approaches, we argue that the continuum of risk is different for these categories. The result is that the necessary legal protections should generally be different for identified and identifiable data.

This Article proceeds in four steps. In Part I, we explore the central role of PII and the grounds for its current uneasy status. The concept of PII arose only during the last fifty years and was tied to the development of the computer. Computerized record systems and techniques of digital data analysis permitted new ways to link data to people. Throughout the 1970s and 1980s, Congress struggled with questions regarding the proper organization of a set of first-generation information privacy statutes. It was only in the Cable Communications Policy Act of 1984[8] that Congress settled on the classic model for these laws: It would solely be an entity's collection or processing of PII that would obligate it to provide privacy safeguards.[9] As Part I also demonstrates, there is still neither a standard nomenclature for PII nor a standard definition of it in the United States. We explore the three basic approaches of American lawmakers to defining PII and find the current formulations of PII to be deeply unsatisfactory.

---

[7] *See infra* Part IV.D (discussing differences between "identified" and "identifiable" data).

[8] Pub. L. No. 98-549, § 2, 98 Stat. 2780 (codified as amended at 47 U.S.C. § 521 (2006)).

[9] *See infra* notes 50–55 and accompanying text (discussing the Cable Act and the new model for privacy safety statutes).

In Part II, we engage in a broader analysis of the weaknesses of PII as it is conceptualized today. First, many people believe in an "anonymity myth,"—that is, a belief that individuals remain anonymous if they have not formally used their name. This belief is especially prevalent for cyberspace activity. Yet, the growth of static IP addresses and other developments creates some built-in identifiability when one enters cyberspace. Second, information that is initially non-PII can be transformed into PII. Third, technology itself is constantly evolving, which means that the line between PII and non-PII is not fixed but rather depends upon changing technological developments. Fourth, the ability to distinguish PII from non-PII is frequently contextual. Many kinds of information are not inherently non-identifiable, or identifiable as an abstract matter.

In Part III, we use behavioral marketing, with a special emphasis on food marketing to children, as a test case for demonstrating the notable flaws in the current definitions of PII. In behavioral marketing, companies generally do not track individuals by name. Rather, they use software to construct personal profiles that exclude names but nonetheless contain a wealth of details about the individual. Companies have also tried to short-circuit the discussion of legal reforms through the simple argument that they do not collect PII.

Digital marketing is also focused on youth.[10] Due to the epidemic of obesity among minors in the United States, the targeted marketing of unhealthy food products to youth is now a highly significant public health issue. The key statute in this regard, the Children's Online Privacy Protection Act (COPPA),[11] restricts websites from gathering and using information obtained from children. Yet, COPPA has weaknesses that permit companies to argue that they are engaging in behavioral marketing without using PII, and hence that their conduct does not fall under this statute.

In its final Part, this Article develops an approach to redefining PII based on the rule-standard dichotomy. Drawing on legal scholarship that has explored this distinction in other settings, we develop a model for PII 2.0 around a standard-based approach. A standard is an open-ended decision-making tool, while a rule is a harder-edged benchmark.[12] In our revitalized standard, PII 2.0 regulates information that relates to either an "identified" or "identifiable" individual, but fixes different legal requirements for each category. We conclude

---

[10] *See infra* Part III.B (discussing the increase in food marketing to youth).

[11] 15 U.S.C. §§ 6501–6506 (2006).

[12] *See infra* Part IV.B (differentiating between rules and standards in the context of their use in legislation defining PII).

by demonstrating the merits of this new approach in the context of behavioral marketing and food marketing to youth.

I
PII's Central Role and Uneasy Status

In this Part, we examine how and why PII became a central concept in information privacy law. Due to the rise in the computer's processing of data, Congress was forced to decide the kinds of data to which information privacy law would apply. Despite legislative grappling with this issue for several decades, there is still no uniform definition of PII in the U.S. We identify three current models of PII and demonstrate why each is inadequate.

A. The Rise of PII and Its Significance

In the last fifty years, PII went from a nearly irrelevant part of privacy law to one of its most central concepts. The early jurisprudence of privacy law lacked any concept of PII as a stand-alone category. In their famous 1890 article, Samuel Warren and Louis Brandeis merely assumed that privacy regulation would always involve information identifiable to a person.[13] They conceived of privacy as a right of "personality."[14] Although the two authors did not define this concept in any detail, they drew on continental philosophy to argue that every person deserves protection against certain kinds of harms as a consequence of her status as a human.[15] The paradigmatic privacy invasion for Warren and Brandeis concerned the press intruding on the privacy of individuals by printing gossip about them.[16] Warren and Brandeis viewed such media reports as necessarily concerning information that would identify a person; otherwise, the gossip would have no sting.

---

[13] Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 Harv. L. Rev. 193 (1890).

[14] *Id.* at 205. As Warren and Brandeis wrote: "The principle which protects personal writings and all other personal productions, not against theft and physical appropriation, but against publication in any form, is in reality not the principle of private property, but that of an inviolate personality." *Id.* For a discussion of this conception of privacy as a right of personality, see Paul M. Schwartz & Karl-Nicholas Peifer, *Prosser's* Privacy *and the German Right of Personality: Are Four Privacy Torts Better than One Unitary Concept?*, 98 Calif. L. Rev. 1925, 1943–44 (2010).

[15] Warren & Brandeis, *supra* note 13, at 205. In their view, as later summarized by Robert Post, a privacy tort was necessary in order to protect each person's "emotional integrity." Robert C. Post, *Rereading Warren and Brandeis: Privacy, Property, and Appropriation*, 41 Case W. Res. L. Rev. 647, 662–63 (1991). For many years after Warren and Brandeis's article, "other authors on the subject of [tort] privacy also rallied around the notion of the right of personality as the basis for" privacy interests. Schwartz & Peifer, *supra* note 14, 1944–47.

[16] Warren & Brandeis, *supra* note 13, at 196.

They thus did not consider PII as an issue warranting any special attention or elaboration.

A later turning point in privacy law occurred in 1960 when William Prosser published his classic article organizing privacy tort law into four categories.[17] Unlike Warren and Brandeis, who built their right of privacy on concepts borrowed from European philosophy, Prosser was content to develop a series of straightforward classifications that over time were able to take on a doctrinal function.[18] Like Warren and Brandeis, however, he left unexplored the issue of PII. Prosser merely assumed that the privacy torts applied only when an identified person was involved.[19]

PII first became an issue in the 1960s with the rise of the computer, which permitted public bureaucracies and private companies to process personal data.[20] The computer did not merely increase the amount of information that entities collected—it changed how that data could be organized, accessed, and searched. A 1977 report from the Privacy Protection Study Commission, a federal blue-ribbon commission, noted that "the physical organization of the records in the database, as well as the physical organization of the items of data within the record, are ceasing to be limiting factors on the way data or records are stored or retrieved."[21] Unlike manual systems, such as telephone books, "computers [could] easily be programmed to sort or reorganize data on the basis of any particular index, attribute, or characteristic."[22] The key point, as the Commission noted, was that computers permitted information to be searched and organized by *multiple attributes* rather than simply through a single index, as, for example, a person's first and last name.[23] This capacity of computers changed the way information could be linked to an individual.

---

[17] William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383 (1960).

[18] *Id.* at 389–407. For a discussion of the doctrinal role of Prosser's concept of tort privacy, see G. EDWARD WHITE, TORT LAW IN AMERICA: AN INTELLECTUAL HISTORY 158–61 (2003).

[19] *See, e.g.*, Prosser, *supra* note 17, at 392–98 (providing examples of privacy invasions, all of which involve identified persons).

[20] Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1402 (2001). For classic early studies in American social sciences and law that trace this connection between the computer and privacy concerns, see generally ARTHUR R. MILLER, THE ASSAULT ON PRIVACY: COMPUTERS, DATA BANKS, AND DOSSIERS (1971) and ALAN WESTIN, PRIVACY AND FREEDOM (1967).

[21] PRIVACY PROT. STUDY COMM'N, PERSONAL PRIVACY IN AN INFORMATION SOCIETY app. 5: TECHNOLOGY AND PRIVACY 21 (1977) [hereinafter PRIVACY COMM'N, TECHNOLOGY & PRIVACY].

[22] *Id.*

[23] *Id.* at 21–22. In his prescient study, *The Assault on Privacy*, Miller also discusses computers' so-called retrieval capacity, which is the ability of computers to filter through

Previously, in order for information to be connected to people, it would almost invariably have to contain their name or likeness. Computerized record systems and techniques of data aggregation and analysis enabled many more pieces of personal data to become linkable to individuals.

This development obliged policymakers to explore a novel set of issues regarding the kinds of information and the nature of the linkages that should trigger the application of information privacy law. The Privacy Protection Study Commission noted that computer systems were capable of retrieving information by searches through databases that were free of indexing around an "individual identifier."[24] The Commission did not discuss the issue in terms of PII, but rather as "who and what is to be covered."[25] No longer was it possible to assume privacy could be protected solely by safeguarding information involving a person's name or likeness. The scope of information requiring privacy protection became significantly larger—and also less clear and more contestable. Thus, the development of computerized records required Congress to confront the issue of the kinds of information that should matter for information privacy law.

In its initial legislative strategy, Congress viewed the types of records at stake as determinative in triggering a statute's protections. The Fair Credit Reporting Act (FCRA) of 1970,[26] the Family Educational Rights and Privacy Act (FERPA) of 1974,[27] and the Privacy Act of 1974[28] demonstrate this approach as well as its attendant weaknesses.

FCRA was the first federal privacy law to respond to computerization and digital records. It applies to any "consumer reporting agency" (CRA) that furnishes a "consumer report."[29] A consumer report is any communication by a CRA that bears on a consumer's credit worthiness or personal characteristics when used to establish the consumer's eligibility for credit, insurance, or for a limited set of other purposes.[30] FRCA sets legal restrictions on the circumstances under which a CRA can furnish a consumer report to another party, as well as the use of these reports for purposes such as law

large amounts of information, and highlights the relevant dangers thereof. MILLER, *supra* note 20, at 39.

24 PRIVACY COMM'N, TECHNOLOGY & PRIVACY, *supra* note 21, at 45.

25 *Id.* at 44.

26 15 U.S.C. § 1681b (2006).

27 20 U.S.C. § 1232g (2006).

28 5 U.S.C. § 552a (2006).

29 15 U.S.C. § 1681b(a).

30 *Id.* § 1681a(d)(1).

enforcement and employment offers.[31] In sum, the statute focuses both on the organization of data about an individual (namely, whether that data appears in a "consumer report") and on the party who collects and uses the information (the CRA).[32] Of the two categories, the concept of the consumer report is the most important.[33] Moreover, due to FCRA's definitional approach, there are notable gaps in its coverage.[34]

Enacted four years after FCRA, FERPA focuses on student privacy. It was also the first federal statute to refer to "personally identifiable information."[35] FERPA uses the term when prohibiting educational entities from releasing or providing access to "any personally identifiable information in education records."[36] Despite its mentioning of PII, however, the statute's central concept is "education records" rather than PII.[37] FERPA defines education records as "information directly related to a student" that an educational institution itself "maintain[s]" in a file or other record.[38] Thus, the statute's

---

[31] SOLOVE & SCHWARTZ, FUNDAMENTALS, *supra* note 4, at 86–91.

[32] *Id.*; *see also* PRACTISING LAW INST., PROSKAUER ON PRIVACY 2-7 to 2-14 (Kristen J. Mathews ed., 2011) (hereinafter PROSKAUER ON PRIVACY) (defining "consumer report" and "credit reporting agency" and discussing how these terms operate under FCRA).

[33] As *Proskauer on Privacy* observes, "[g]iven that the definition of a CRA depends largely on the definition of 'consumer report,' the fact that a particular set of information is not a consumer report can prevent a person or entity from acting as a CRA for the purposes of the Act." PROSKAUER ON PRIVACY, *supra* note 32, at 2-11.

[34] The statute makes clear, for example, that it does not apply to a party, such as a bank, that furnishes financial information that goes into a consumer report. 15 U.S.C. § 1681a(d)(2)(A)(i) (2006). For case law reaching this conclusion, see Mirfasihi v. Fleet Mortg. Corp., 551 F.3d 682, 686 (7th Cir. 2008) and Smith v. First Nat'l Bank of Atlanta, 837 F.2d 1575, 1578 (11th Cir. 1988). Although such entities provide a CRA with information about consumers, they themselves are not in the business of supplying a consumer report to third parties. FCRA contains another problematic and explicit exception to its definition of consumer report. This term does not extend to the sharing of information among affiliated entities, so long as the consumer is given an opportunity to opt out of such sharing. 15 U.S.C. § 1681a(d)(2)(A)(iii).

In Congressional testimony in 2003, Joel Reidenberg pointed to the consequences of this exemption: It "means that credit report information loses protection when shared with far-flung related companies." *Affiliate Sharing Practices and Their Relationship to the Fair Credit Reporting Act: Hearing Before the S. Comm. on Banking, Hous., and Urban Affairs*, 108th Cong. 8 (2003) (statement of Joel R. Reidenberg, Professor, Fordham University School of Law).

[35] 20 U.S.C. § 1232g(b)(2) (2006). While Congress does not define PII in the statute, a federal regulation provides a broad approach to it. *See* 34 C.F.R. § 99.3 (2010) (defining PII to include a student's name, address, and social security number, as well as several other "indirect identifiers").

[36] 20 U.S.C. § 1232g(b)(2).

[37] *Id.*

[38] *Id.* § 1232g(a)(4)(A)(i)–(ii).

coverage depends on whether or not a school has first organized and then stored data in education records.[39]

Due to FERPA's limitations, schools long profited by distributing "surveys" on behalf of marketers.[40] Since the collected information went from parents and children to marketers without being "maintain[ed]" in "educational records" by schools, this practice fell outside of FERPA's coverage.[41] Congress finally responded to this practice in a limited fashion in 2005. It left FERPA unaltered, but created a limited separate statutory interest that permits parents of elementary and secondary school students to opt out of the collection of student information for commercial purposes.[42] Congress neither revisited FERPA's reliance on the concept of "educational records," nor created a more basic right to block the release of student records for commercial purposes. As for universities, they remain able to sell essential student contact information to credit card companies;[43] such data is considered "directory information," and hence not an "education[al] record[ ]."[44]

In a fashion similar to FCRA and FERPA, the Privacy Act's threshold turns on how record systems are organized rather than on whether information could be linked to an individual. The key trigger of the Privacy Act concerns how federal agencies retrieve information from a database; it applies only when information is retrieved from a "system of records."[45] The Act defines a "system of records" as "a group of any records . . . from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."[46] As a

---

[39] JAMES RAPP, 5 EDUCATION LAW § 13.04[7][a] (2010). The Supreme Court has also heard the siren call of protection based on type of record rather than individual identifiability. In 2002, in *Owasso Independent School District v. Falvo*, 534 U.S. 426 (2002), the Supreme Court went further than even FERPA's statutory language and strongly suggested in dicta that FERPA records are only those kept in a permanent file and by a "central custodian" at the school. *Id.* at 434–35.

[40] Lynn M. Daggett, *FERPA in the Twenty-First Century: Failure To Effectively Regulate Privacy for All Students*, 58 CATH. U. L. REV. 59, 100–01 (2008).

[41] 20 U.S.C. § 1232g(a)(6).

[42] Daggett, *supra* note 40, at 78–79. Regarding these changes, Congress placed modest limits on the ability of elementary and secondary schools to collect and disclose student information for commercial purposes. While schools must give parents an opportunity to opt out of such sharing, the law does not ban sharing for commercial purposes and does not require affirmative consent from parents. USA PATRIOT Act of 2001, Pub. L. No. 107-56, 115 Stat. 272, 367–68 (codified as amended at 20 U.S.C. § 1232g(j)(1) (2006)).

[43] Daggett, *supra* note 40, at 89.

[44] 20 U.S.C. § 1232g(b)(2).

[45] 5 U.S.C. § 552a(a)(5) (2006).

[46] *Id.* A record includes "any item, collection, or grouping of information about an individual that is maintained by an agency . . . and that contains his name, or the identifying

consequence, the Privacy Act only covers computer searches that identify an individual when retrieval of data is done through reference to a specific personal identifier, such as a name or Social Security Number.[47]

Like FERPA, the Privacy Act remains an antiquated law that misses the significance of the computer search revolution—namely, the ability of computers to investigate, analyze, and manipulate data sets and to find new ways to locate specific persons. As an example of an action that is *not* covered by the Privacy Act, a federal agency that examines its computer records by a search around psychiatric diagnosis, age, or other personal attributes is *not* retrieving data from a system of records by use of an identifying particular assigned to a person.[48] Within three years of the statute's enactment, the Privacy Protection Study Commission had already drawn attention to and condemned this profound flaw.[49] Nonetheless, over thirty years after enactment of the Privacy Act, Congress still has not corrected this central failing of the statute.

Finally, in 1984, with the passage of the Cable Communications Policy Act (Cable Act), Congress reached an important milestone.[50] The statute not only refers to PII, but also makes PII the trigger for the applicability of the law.[51] The innovation of the Cable Act was to tie the presence of PII to an obligation to follow Fair Information Practices (FIPs), which are the building blocks of modern information

number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph." *Id.* § 552a(a)(4).

[47] As the Department of Justice's guide to the Privacy Act summarizes, "[t]he highly technical 'system of records' definition is perhaps the single most important Privacy Act concept, because . . . it makes coverage under the Act dependent upon the method of *retrieval* of a record rather than its substantive content." DEP'T OF JUSTICE, OVERVIEW OF THE PRIVACY ACT OF 1974, at 25 (2010) (emphasis added).

[48] *See* OMB Privacy Act Implementation, 40 Fed. Reg. 28,948, 28,952 (July 9, 1975) (defining "system of records" as limited to "information [that] is retrieved by the name of the individual or by some . . . other identifying particular assigned to the individual"); PRIVACY PROT. STUDY COMM'N, THE PRIVACY ACT OF 1974: AN ASSESSMENT 6–7 (1974), *available at* http://epic.org/privacy/ppsc1977report/appendix4.html (providing an example of a Veterans' Administration search by psychiatric diagnosis that was not covered by the Act).

[49] PRIVACY PROT. STUDY COMM'N, PERSONAL PRIVACY IN AN INFORMATION SOCIETY 59–61 (1977) [hereinafter PRIVACY COMM'N, PERSONAL PRIVACY].

[50] Pub. L. No. 98-549, 98 Stat. 2779 (1984) (codified as amended in scattered sections of 47 U.S.C.).

[51] 47 U.S.C. § 551(a)(1)(A) (2006). Nevertheless, the legislative history of the Cable Act proves singularly unhelpful regarding the selection of this term as the trigger for the statute's coverage. *See* H.R. REP. NO. 98-934, at 76–79 (1984) (stating only that PII "would include specific information about the subscriber, or a list of names and addresses on which the subscriber is included, but does not include aggregate information about subscribers which does not identify particular persons").

privacy law. FIPs establish obligations for organizations that process personal information.[52] The Cable Act prohibits a cable operator using a cable system from collecting PII "concerning any subscriber without the prior written or electronic consent of the subscriber concerned."[53] It provides for subscriber access to all PII "regarding that subscriber which is collected and maintained by a cable operator."[54] It further requires that subscribers be given notice about the nature of PII "collected or to be collected with respect to the subscriber and the nature of the use of such information."[55]

There is a clear difference between the Cable Act and FCRA, FERPA, and the Privacy Act. The Cable Act does not extend its protections based on how information is assembled, whether in a credit record, as in FCRA, an educational record, as in FERPA, or a "system of records," as in the Privacy Act. Rather, the Cable Act's statutory obligations fall on a cable operator as soon as the operator collects PII.

What inspired this important shift in the law between the early 1970s and 1984? First, there was a renewed focus in society on information privacy during the latter part of the 1970s. Google Ngram provides a convincing demonstration of this development; this Google database permits statistical analysis of the frequency of use of specific words and phrases.[56] A Google Ngram search of the term "information privacy" reveals an increase in attention to the topic beginning in the late 1970s and continuing during the run-up to the enactment of the Cable Act.[57]

Part of this attention was driven, in turn, by the arrival of George Orwell's titular year, 1984. A flurry of media reports and law review

---

[52] In the United States, the Department of Health, Education and Welfare had first mentioned FIPs in an influential report in 1973. U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS COMPUTERS AND THE RIGHTS OF CITIZENS xxi, xxiii, xxix–xxx (1973). On the policy history of FIPs, see PRISCILLA M. REGAN, LEGISLATING PRIVACY 73–86 (1995). For an introduction to FIPs, see DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 655–58 (3d ed., 2009) [hereinafter SOLOVE & SCHWARTZ, IPL], and Paul M. Schwartz, *Preemption and Privacy*, 118 YALE L.J. 902, 907–08 (2009).

[53] 47 U.S.C. § 551(b)(1). Note, however, that personal information about programming choices is not protected, *see infra* notes 59–61 and accompanying text.

[54] 47 U.S.C. § 551(d).

[55] *Id.* § 551(a)(1)(A).

[56] The Ngram Viewer, a tool launched by Google Labs, creates a graphical year-by-year representation of how often a phrase has been used in books. It draws on nearly 5.2 million books from a period between 1500 and 2000 A.D., which the Google Library Project has digitalized. Patricia Cohen, *In 500 Billion Words, New Window on Culture*, N.Y. TIMES, Dec. 17, 2010, at A3.

[57] Appendix A to this Article contains the Ngram Viewer's chart for this term between 1950 and 2000. The chart also shows how this attention only intensified throughout the 1990s with the emergence of the Internet and other threats to privacy. *Infra* App. A.

articles marked this occasion with an analysis of new threats to privacy.[58] Thus, this notable literary year helped to heighten the concern about privacy in the United States.

Perhaps most importantly, however, cable operators' collection of personal information created the same kinds of issues that the Internet would later raise. Even as early as the 1980s, observers noted that cable would permit a user not only to receive information, as broadcast television long had allowed, but also to respond to information on the screen and make programming choices.[59] By collecting these data, the cable operator would be able to construct detailed profiles about viewing choices. Moreover, it was anticipated that cable would provide "videotex," which was envisioned as a two-way communication system permitting users to access information directly from their service provider's computers.[60] A "videotex explosion" would lead, in turn, to the conveying of detailed data about one's "interests, choices and views to the central computer" of the system operator.[61] Due to legislative concerns about these practices, the policy response in the Cable Act was to regulate around information, rather than around how data was organized. This regulatory insight,

---

[58] As one law review article stated, "[t]o prevent cable from turning the television set into an Orwellian nightmare, the [Cable] Act creates a framework for the protection of subscriber privacy." Michael I. Meyerson, *The Cable Communications Policy Act of 1984: A Balancing Act on the Coaxial Wires*, 19 GA. L. REV. 543, 612 (1985); *see also* John Shattuck, *In the Shadow of* 1984*: National Identification Systems, Computer-Matching, and Privacy in the United States*, 35 HASTINGS L.J. 991, 991 (1984) ("Recent actions by the federal government ha[d] brought the technology . . . invasion from the realm of science fiction into the real world of public policy."); Mindy E. Wachtel, Note, *Videotex: A Welcome New Technology or an Orwellian Threat to Privacy?*, 2 CARDOZO ARTS & ENT. L.J. 287, 311 (1983) (noting that a two-way cable system "could quickly destroy individual privacy by filtering vast quantities of intimate information to commercially exploitive enterprises, overzealous government enforcement officials or the idly curious").

For illustrative accounts of threats to privacy in the popular press in 1983 and 1984 that also discussed Orwell's famous novel, see Walter Cronkite, *Orwell's '1984'—Nearing?*, N.Y. TIMES, June 5, 1983, at E23, Thomas Ferraro, *Is an Orwellian Society Upon Us?*, L.A. TIMES, Dec. 26, 1983, at D31, and John J. Fialka, *The Time Has Come for Deciding if 1984 Will Resemble* 1984, WALL ST. J., June 7, 1983, at 1, 17.

[59] As Meyerson noted in 1985:

> [A]dvanced cable systems are able to monitor continually the viewing choices of each cable household. This capability presents a serious potential for invading the privacy of the cable subscriber. Not only can intimate information be gleaned easily by the cable operator, but an unprecedented amount and variety of information about an individual can also be inexpensively accumulated from one source—the cable system.

Meyerson, *supra* note 58, at 612.

[60] *See* Wachtel, *supra* note 58, at 287–88 (describing videotex and its potential for information sharing).

[61] *Id.* at 290.

once reached, established the model for information privacy regulation to come.

After enactment of the Cable Act, information privacy law continued to use the collection of PII as the trigger for applicability of legal protection. Congress and the states developed a series of privacy laws around the concept of PII.[62] These laws failed to settle, however, on a standard nomenclature for PII. To this day, information privacy law scholars use the alternative term, "personal information," quite frequently and sometimes interchangeably with PII.[63] Nevertheless, PII has become the preferred term of art since the mid-1990s.

Even more troublesome than the inconsistent nomenclature, information privacy law has failed to develop a coherent and workable definition of PII. Although the concept gained ascendency over the past two decades, and has become the central device for determining the scope of privacy laws, scant intellectual attention has been given to the theory behind the term. A variety of definitions of PII have arisen in privacy laws, but with little thought as to the selection of one over the others. In other words, the complexities of PII have not been adequately explored. As we will discuss in the next Section, moreover, all of these definitions are flawed.

PII is a challenging conceptual issue at the heart of any system of regulating privacy in the Information Age. If PII is defined too narrowly, then it will fail to protect privacy in light of modern technologies involving data mining and behavioral marketing. Technology will thus make privacy law irrelevant and obsolete. On the other hand, if PII is defined too broadly, then it could encompass too much information, and threaten to transform privacy law into a cumbersome and unworkable regulation of nearly all information. Privacy law must have coherent boundaries, which adequately protect privacy and which can be flexible and evolving.

But PII is complicated and hard to pin down. Computer science has shown that the concept of PII is far from straightforward. Increasingly, technologists can take information that appears on its face to be

---

[62] For illustrative laws, see Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501–6506 (2006), Video Privacy Protection Act, 18 U.S.C. § 2710 (2006), Driver's Privacy Protection Act, 18 U.S.C. §§ 2721–2725 (2006), and Information Practices Act of 1977, CAL. CIV. CODE §§ 1798.29, 1798.82, 1798.84 (2008).

[63] For two examples, compare William McGeveran, *Disclosure, Endorsement, and Identity in Social Marketing*, 2009 U. ILL. L. REV. 1105, 1135 (2009) (using "personal information" in place of PII), with Robert Sprague & Corey Ciocchetti, *Preserving Identities: Protecting Personal Identifying Information Through Enhanced Privacy Policies and Laws*, 9 ALB. L.J. SCI. & TECH. 91, 92 (2009) (using "personal information" interchangeably with PII).

non-identifiable and turn it into identifiable data.[64] Moreover, the marketing industry is involved in practices that raise privacy concerns, but that do not fall within any of the current definitions of PII.[65]

While the edifice of privacy law is built on PII, only recently has some awareness emerged about this core conceptual problem. In 2010, the FTC finally recognized the extent of the PII problem. In a major report, it acknowledged "the blurring of the distinction between personally identifiable information and supposedly anonymous or de-identified information."[66] The FTC pointed to the need to rethink PII, but did not make any headway beyond this request.[67] In the scholarly literature, moreover, there has been surprisingly scant attention paid to the issue of PII. In 1998, Jerry Kang devoted several pages in a seminal early paper about Internet privacy to a discussion of when data became "personal information."[68] More recently, Paul Ohm published a major piece devoted to arguing that we abandon the very concept of PII. For Ohm, PII is a fatally flawed concept because so much non-PII can be re-identified.[69] If the PII problem remains unresolved, then we will continue to lack a coherent approach to defining the proper boundaries of privacy regulation. Privacy law thus depends upon addressing the PII problem—it can no longer remain the unacknowledged elephant in the room.

B. The Current Typology of PII

Given the ubiquity of the concept in privacy law and the important role it plays, the definition of PII is crucial. But instead of defining PII in a coherent and consistent manner, privacy law offers multiple competing definitions, each with some significant problems and limitations. There are three predominant approaches to defining PII in various laws and regulations. We will refer to these approaches as (1) the "tautological" approach, (2) the "non-public" approach, and (3) the "specific-types" approach.

---

[64] *See infra* Part II.B (describing means by which data can be made identifiable).

[65] *See infra* Part III.A (discussing behavioral marketing and the concerns it poses for privacy).

[66] FTC, PROTECTING PRIVACY, *supra* note 2, at iv.

[67] In this report, the FTC stated that it would leave open the question of the feasibility of a proposed definition of PII centered on data that can be "reasonably linked to a specific consumer, computer, or other device." *Id.* at 43. The FTC's concern was whether such a definition was "feasible, particularly with respect to data that, while not currently considered 'linkable,' may become so in the future." *Id.*

[68] Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193, 1206–11 (1998).

[69] Ohm, *supra* note 6, at 1742.

At the start of this examination of the current definitions of PII, a brief introduction to the jurisprudence of rules and standards is in order. A standard is an open-ended yardstick for decision making, and a rule, its counterpart, is a harder-edged decision-making tool.[70] To illustrate, consider the possibilities under the rule/standard dichotomy for regulating the behavior of an automobile driver at a train crossing: (1) stop, look, and listen (the rule), or (2) proceed with reasonable caution (the standard).[71] The existing approaches to defining PII can be categorized as either rules or standards. The first two of our categories fall into the legal category of a standard, and the last one, a rule.

1. The Tautological Approach

The tautological approach is an example of a standard, and defines PII as any information that identifies a person. The Video Privacy Protection Act (VPPA) neatly demonstrates this model.[72] The VPPA, which safeguards the privacy of video sales and rentals, simply defines "personally identifiable information" as "information which identifies a person."[73] For purposes of the statute, information that identifies a person is PII and falls under the statute's jurisdiction once linked to the purchase, request, or obtaining of video material.

The virtue of the tautological approach, like that of other kinds of standards, is that it is open rather than closed in nature. As a standard, it can evolve and remain flexible in response to new developments. The problem with the tautological approach, however, is that it fails to define PII or explain how it is to be singled out. At its core, this approach simply states that PII is PII. As a result, this definition is unhelpful in distinguishing PII from non-PII.

2. The Non-Public Approach

A second approach to defining PII is to focus on non-public information. Here, too, we see the use of a standard. The non-public

---

[70] For a discussion of the distinction between rules and standards, see Carol M. Rose, *Crystals and Mud in Property Law*, 40 STAN. L. REV. 577, 592–93 (1988), and Kathleen M. Sullivan, *The Supreme Court, 1991 Term: Foreword: The Justices of Rules and Standards*, 106 HARV. L. REV. 22, 57–59 (1992).

[71] These examples follow from two Supreme Court decisions: Pokora v. Wabash Ry., 292 U.S. 98, 101–02 (1934), and Baltimore & Ohio R.R. v. Goodman, 275 U.S. 66, 70 (1927).

[72] Video Privacy Protection Act of 1988, 18 U.S.C. § 2710 (2006).

[73] *Id.* § 2710(a)(3). The VPPA prohibits "videotape service providers" from knowingly disclosing personal information, such as the titles of items rented or purchased, without the individual's written consent. It defines "videotape service providers" in a technologically neutral fashion to permit the law to be extended to DVDs. *Id.* § 2710(a)(4).

approach seeks to define PII by focusing on what it is *not*, rather than on what it is. In a sense, this approach is simply a variant of the tautological approach. Instead of saying that PII is simply that which identifies a person, the non-public approach draws on concepts of information that is publicly accessible and information that is purely statistical. This model would exclude information that falls in these two categories from PII, but the relevant legislation does not explore or develop the logic behind this approach.

The Gramm-Leach-Bliley Act (GLBA) epitomizes one aspect of this approach by defining "personally identifiable financial information" as "nonpublic personal information."[74] The statute fails to define "nonpublic," but presumably this term means information not found within the public domain.[75] In an illustration of another aspect of this approach, the Cable Act defines PII as something other than "aggregate data."[76] This statute, which protects the privacy of subscribers to cable services, views PII as excluding "any record of aggregate data which does not identify particular persons."[77] By aggregate data, the Cable Act presumably means purely statistical information that does not identify specific individuals.[78]

The problem with the non-public approach is that it does not map onto whether the information is in fact identifiable. The public or private status of data often does not match up to whether it can identify a person or not. For example, a person's name and address, which clearly identify an individual, nevertheless might be considered public information, as such information is typically listed in telephone books. In many cases, however, individuals have non-public data that they do not want matched to this allegedly public information. Yet, an approach that only protects non-public information as PII might not preclude such combinations.

---

74 Gramm-Leach-Bliley Act of 1999, 15 U.S.C. § 6809(4)(A) (2006).

75 During GLBA rulemaking proceedings, financial regulatory agencies "wrestled with the concept of nonpublic personal information" before ultimately focusing their concept of "nonpublic" on whether personal information was "publicly available." Charles M. Horn, *Financial Services Privacy at the Start of the 21st Century: A Conceptual Perspective*, 5 N.C. BANKING INST. 89, 107–08 (2001). In this context, Horn adds, "publicly available" information includes "any information that a financial institution has a 'reasonable basis' to believe is lawfully available to the general public from federal, state or local governmental records, widely distributed media (including the Internet), or disclosures to the general public required to be made by federal, state or local law." *Id.* at 107.

76 Cable Communications Policy Act of 1984, 47 U.S.C. § 551(a)(2)(A) (2006).

77 *Id.*

78 The number of Comcast customers in Virginia who subscribe to HBO is an example of aggregate data under the Cable Act.

3. The Specific-Types Approach

The third approach is to list specific types of data that constitute PII. This approach is a classic example of a rule. In the context of the specific-types approach, if information falls into an enumerated category, it becomes per se PII by operation of the statute. To illustrate three different variations on this approach, we will examine Massachusetts's breach notification statute of 2007 (officially titled the Standards for the Protection of Personal Information of Residents of the Commonwealth),[79] California's Song-Beverly Credit Card Act of 1971,[80] and the federal Children's Online Privacy Protection Act (COPPA) of 1998.[81]

The Massachusetts breach notification statute requires that individuals be notified if a defined set of their personal information is lost or leaked.[82] The Act defines PII as a person's first name and last name, or first initial and last name in combination with either a social security number, driver's license number, financial account number, or credit or debit card number.[83]

The Song-Beverly Act prohibits merchants who accept credit cards from collecting a cardholder's "personal identification information" when transacting business with her.[84] More specifically, it prohibits retailers from requesting or requiring "as a condition to accepting the credit card" that a cardholder provide "any personal identification information upon the credit card transaction form or otherwise."[85] The critical language in the Act defines PII as "information concerning the cardholder, other than information set forth on the credit card, and including, but not limited to, the cardholder's address and telephone number."[86]

Finally, COPPA, a federal statute, regulates the collection and use of children's information by Internet websites or online services.[87] Like the Massachusetts statute, it approaches the question of PII versus non-PII in a typological fashion. COPPA states that personal information is "individually identifiable information about an

---

[79] 201 MASS. CODE REGS. § 17.00 (2010).

[80] CAL. CIV. CODE § 1747 (2009).

[81] 15 U.S.C. § 6501 (2006).

[82] *E.g.*, 201 MASS. CODE REGS. § 17.04.

[83] *Id.* § 17.02.

[84] Song-Beverly Credit Card Act of 1971, CAL. CIV. CODE § 1747.08 (2009). Note that the Act uses the term "personal identification information." This language reinforces our earlier point that there is no standard nomenclature for PII. *See supra* notes 62–63 and accompanying text.

[85] CAL. CIV. CODE § 1747.08(a)(1).

[86] *Id.* § 1747.08(b).

[87] Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506 (2006).

individual collected online," including first and last name, physical address, social security number, telephone number, and e-mail address.[88] This law's definition of PII also includes "any other identifier that the [Federal Trade Commission (FTC)] determines permits the physical or online contacting of a specific individual."[89] In 2000, the FTC issued its COPPA Rule.[90] It added one element to the Act's definition of PII by extending this concept to a "persistent identifier, such as a customer number held in a cookie or a processor serial number, where such identifier is associated with individually identifiable information."[91]

An initial problem with the specific-types approach is that it can be quite restrictive in how it defines PII. The Massachusetts statute defines PII to include only a narrow set of data elements: a name plus other elements, such as a social security number, a driver's license number, or a financial account number.[92] This list is under-inclusive: There are numerous other kinds of information that, along with a person's name, would serve specifically to reveal one's identity. For example, a person's name and sensitive personal medical information would, in many cases, permit the identification of that person. Moreover, most individuals would consider such a data breach to be a significant event and one about which they would wish to be informed. Yet, this leak appears to fall outside the kind of PII that the Massachusetts breach notification statute covers. The Massachusetts version of the specific-types approach also assumes that the types of data that are identifiable to a person are static—the statute does not cover information that could potentially become personally identifiable. As we will argue later in this Article, however, this assumption is false.[93] This variant of the specific-types approach is too rigid to adequately protect personal privacy.

As for the version of the specific-types approach in the Song-Beverly Act, its text appears far less narrow than that of the Massachusetts statute. Nonetheless, a recent series of decisions regarding the Act demonstrates how easy it is for PII to be interpreted only as information exclusive to one person. Two lower courts in California have interpreted this statute as providing extremely limited

---

[88] *Id.* § 6501(8)(A)–(E).

[89] *Id.* § 6501(8)(F).

[90] 16 C.F.R. § 312.2 (2011).

[91] *Id.*

[92] 201 MASS. CODE REGS. § 17.02 (2010).

[93] *See infra* Part II (arguing that the distinction between PII and non-PII changes with context, technology, and availability of apparently anonymous data).

protection.[94] While the California Supreme Court in 2011 corrected the lower courts' interpretations of the statute, the general flaw of the specific-types approach remains even after this decision.[95]

The litigation in question, *Pineda v. Williams-Sonoma*, involved a suit for violation of the Song-Beverly Act. Plaintiff Jessica Pineda visited defendant's store in San Diego County, selected an item to purchase, and then went to the cashier to pay for it with her credit card. As the Superior Court indicated, "[t]he cashier asked her for her zip code, but did not tell her the consequences if she declined to provide the information."[96] Pineda believed that she was obliged to provide this information to complete the transaction, and she supplied it to the cashier.[97] The cashier recorded the zip code in the electronic cash register, which meant that the store now had the following information in its database: the customer's credit card number, the name on her credit card, and her zip code.[98]

As we have noted above, the critical language in the Song-Beverly Act defines PII as "information concerning the cardholder, other than information set forth on the credit card, and including, but not limited to, the cardholder's address and telephone number."[99] While this language appears broad, the appellate court in *Pineda* followed the trial court in deciding that the Song-Beverly Act defined PII only as data that was "facially specific" to the individual, such as an entire address, including the zip code, but not exclusively a zip code.[100] As the appellate court declared, the statute defined PII as data that was "specific in nature regarding an individual, rather than a group identifier such as a zip code."[101] For that court, "a zip code [was] not facially individualized information."[102] This judgment meant that information such as an entire address including the zip code would be protected, but a zip code alone would not.[103]

---

94 Trial Order, Pineda v. Williams-Sonoma Stores, Inc., No. 37-2008-00086061-CU-BT-CTL, 2008 WL 7414542 (Cal. App. Dep't Super. Ct. Oct. 29, 2008); Pineda v. Williams-Sonoma Stores, Inc., 100 Cal. Rptr. 3d 458 (Cal. Ct. App. 2009), *rev'd*, 246 P.3d 612 (Cal. 2011).

95 Pineda v. Williams-Sonoma Stores, Inc., 246 P.3d 612 (Cal. 2011).

96 *Pineda*, 100 Cal. Rptr. 3d at 460.

97 *Id.*

98 *Id.*

99 CAL. CIV. CODE § 1747.08(b) (2009).

100 *Pineda*, 100 Cal. Rptr. 3d at 461.

101 *Id.* (quoting Party City Corp. v. Superior Court, 86 Cal. Rptr. 3d 721, 738 (Ct. App. 2008)).

102 *Pineda*, 100 Cal. Rptr. 3d at 461 (citing *Party City*, 86 Cal. Rptr. 3d at 736).

103 *Id.* As the appellate court declared, the statute defined PII as data that were "specific in nature regarding an individual, rather than a group identifier such as a zip code." *Id.* (quoting *Party City*, 86 Cal. Rptr. 3d at 738).

The California Supreme Court reversed this verdict, but it did so on the narrowest possible grounds. The Court analyzed the statutory language and legislative history of the statute and found that both supported finding a legislative intent to include a zip code as part of the "cardholder's address."[104] In other words, that statutory category included "not only a complete address, but also its components."[105] Yet, the California Supreme Court had only tweaked a subcategory within the specific-types approach. It did not reach the broader conclusion that the Act's specific categories reflected a policy to prevent retailers from collecting identification indices that would permit a definitive linkage between a customer and her address.

A more accurate reading of the law would be that it prohibits merchants from collecting information that is specific enough to allow the identification of a unique person. Under such a reading, a zip code would be covered by the statute. Regardless of the fact that numerous individuals may share the same zip code, this piece of information, coupled with an individual's name, enables retailers to link the individual customer with a wealth of PII about her.[106] As the state supreme court itself observed, once the Williams-Sonoma store had Pineda's zip code, it drew on a licensed proprietary database to perform a "reverse search[ ]" that allowed it to identify the customer's address and other information about her.[107] In fact, the store had created this database to market products to its customers as well as to have the possibility of selling "the information it ha[d] compiled to other businesses."[108]

As for COPPA, our third example of the specific-types approach, this federal statute has an advantage that the Massachusetts and California laws lack. COPPA explicitly references FTC rulemaking as a way to expand and adapt its definition of PII.[109] The FTC has indeed acted to expand the definition of PII in the statute; its COPPA rule added one element to the statutory concept of PII, namely, the idea of "a persistent identifier," such as a cookie.[110] However, the FTC's ability to alter the definition of PII is limited by a requirement that information covered by the statute must permit the "contacting of a

---

104 Pineda v. Williams-Sonoma Stores, Inc., 246 P.3d at 616, 619–20 (quoting CAL. CIV. CODE § 1747.08(b)).

105 *Id.*

106 *Id.* at 615.

107 *Id*. Such searches were run through "databases that contain millions of names, e-mail addresses, residential telephone numbers and residential addresses . . . ." *Pineda*, 100 Cal. Rptr. 3d at 460.

108 *Pineda*, 246 P.3d at 615.

109 15 U.S.C. § 6501(8)(F) (2006).

110 16 C.F.R. § 312.2 (2011).

specific individual."[111] As we will discuss in more detail later, there are indications that this agency is unlikely to define "contacting" to include serving specific ads to a person.[112]

A final difficulty with COPPA, and one that is typical of rules, is that it requires that PII be defined in advance.[113] The COPPA twist is to permit the statutory listing to be expanded through agency rulemaking.[114] Nonetheless, the risk remains that new technology will develop too quickly for this approach to be effective. For example, the FTC's COPPA Rule has not been revisited since it was issued in 2000. Indeed, the FTC's own wavering line regarding new privacy legislation serves as an illustration of internal gridlock in a regulatory agency.[115] In his study of co-regulatory privacy approaches, Ira Rubinstein traces a long cycle, from 1995 to 2010, in which "the FTC's embrace of self-regulatory solutions has waxed and waned over the years, and once again appears to be ascendant at least as to online behavioral advertising."[116]

* * *

Despite the importance of the concept of PII to privacy law and regulation, there remains a lack of consensus in the United States about how to define it. All current legal models for this concept are flawed. The tautological approach merely begs the question. The non-public approach seeks to define what PII is not, but its focus on the public or private nature of the data is ultimately a different issue than whether or not data are identifiable to a person. Finally, the specific-types approach fails to offer a definition—it merely lists examples of PII, but supplies no concept or method by which to determine whether a type of information belongs on or off the list.

As we have also seen, the PII issue only emerged in the late 1960s with the widespread use of the computer. It was due to this device's ability to change the means of accessing and searching information that the line between PII and non-PII became less certain. Today, that line is not merely uncertain: Professor Ohm questions whether

---

111 15 U.S.C. § 6501(8)(F).

112 *See infra* Part IV.C.I (arguing that a cookie used only to send targeted ads would not fall within the FTC's definition of PII).

113 *See* 15 U.S.C. § 6501(8) (defining "personal information" and giving a representative list of types of information that fall within that category).

114 *Id.* § 6501(8)(F).

115 *See* Ira Rubinstein, *Privacy and Regulatory Innovation: Moving Behind Voluntary Codes* (NYU Sch. of Law, Pub. Law Research Paper No. 10-16, 2011), http://www.is-journal.org/hotworks/rubinstein.php (last visited Oct. 31, 2011) (describing shortcomings in the FTC's approach to monitoring online profiling).

116 *Id.*

maintaining a distinction between PII and non-PII is even possible.[117] Thus, privacy law and scholarship must confront the PII problem.

II
The Problems with PII

PII remains a central concept in privacy regulation. It strikes many as common sense that a person's privacy can be harmed only when PII is collected, used, or disclosed. In this Part, we explain why PII, as currently defined, is a troubled concept for framing privacy regulation. As we contend, the current distinction between PII and non-PII proves difficult to maintain. Indeed, whether information is identifiable to a person will depend upon context and cannot be determined *a priori*.

In this Section, we proceed through four steps to show defects in the existing distinction between PII and non-PII. First, we discuss a widely held misunderstanding about anonymity on the Internet. Many people believe that if they do not formally use their name when operating in cyberspace that they are anonymous. Due to the growth of static IP addresses, however, there is a basic level of built-in identifiability as soon as a computer connects to the Internet. Second, we show how information that is initially non-PII can be transformed into PII. Technology increasingly enables marketers and others to combine various pieces of non-PII to produce PII, or otherwise to forge a link between some data and a specific person. In fact, the permanent de-identification of information is difficult because so much data about individuals exists in so many places, and some of these data are linked to specific identities. Third, technology itself is constantly changing. As a result, the line between PII and non-PII is not fixed but depends upon changing technological developments. Fourth, the ability to distinguish PII from non-PII frequently depends on context. For example, whether or not a search query is PII cannot be determined in the abstract.[118]

A. The Anonymity Myth and the IP Address

There is a common myth about anonymity on the Internet. Many people likely believe that anonymity exists for most situations when one surfs the Web or engages in behavior in cyberspace. The "anonymity myth," as we will call it, is the incorrect assumption that as long as one does not explicitly do something under one's actual name

[117] *See* Ohm, *supra* note 6, at 1742 (positing that PII is an "ever-expanding category" that "will never stop growing until it includes everything").

[118] *See infra* notes 142–45 and accompanying text.

on the Internet, there will be safety from identification. In other words, there is a false belief that the default for most Internet situations is anonymity. By extension, many believe that if one does not provide specific identification when posting a comment to a blog or social network website, or if one relies on a pseudonym, anonymity has been secured for such behavior. Despite the fact that it appears so easy to act anonymously online, this anonymity offers no more protection than a veil over one's face that can readily be lifted.

At its most basic level, the anonymity myth stems from a mistaken conflation between momentary anonymity with actual untraceability. It is easy to communicate online or surf the Web without immediately revealing one's identity, but it is much more difficult to be non-traceable. Whenever one is online, a potential for traceability exists. In this section, we explore a threshold issue at the entry to cyberspace that contributes significantly to traceability: the internet protocol (IP) address. In later sections, we will discuss a number of other factors that contribute to such traceability on the Internet.

An IP address is a unique identifier that is assigned to every computer connected to the Internet.[119] Due to the shift from dial-up to static IP addresses, Internet service providers (ISPs) now have logs that link IP addresses with particular computers and, in many cases, eventually to specific users.[120] To understand why these links exist, it will be useful to trace the shift in usage from dial-up Internet service to broadband.

Like the Sony Walkman and cassette tapes, dial-up service is a cultural relic of fading significance. Dial-up is a form of Internet access that uses the facilities of the public-switched telephone network to establish a connection to an ISP. According to a 2010 report from the Pew Research Center, only five percent of Americans continue to use dial up service to access the Internet.[121] A pro-anonymity aspect of dial up Internet service is its dynamic assignment of a new IP address to a customer's computer every time she connects to the Internet.[122] As a consequence, many customers share a single IP address at different times over the course of a single day. Moreover, ISPs typically do not retain records about dynamic IP use for long

---

[119] GARY BAHADUR ET AL., PRIVACY DEFENDED 192 (2002).

[120] *Id.* at 194.

[121] AARON SMITH, PEW INTERNET & AM. LIVE PROJECT, HOME BROADBAND 2010, at 6 (Aug. 11, 2010), http://pewinternet.org/Reports/2010/Home-Broadband-2010.aspx.

[122] BAHADUR ET AL., *supra* note 119, at 194.

periods of time.[123] The result is that identification of any specific person through an IP address is relatively unlikely.

Starting in the last decade, however, the majority of Internet users began to access it through high-speed services, such as cable or DSL.[124] The benefit of such broadbrand access is that it enables a wide range of activities in cyberspace, including multimedia and virtual worlds. These experiences would be impossible at dial-up's glacial rate of Internet access. On the negative side, broadband connections generally are based on static IP addresses, which do not change. That is, a long-standing DSL or cable account will have the same IP address for years.[125] In the current age of broadband, where an IP address is statically assigned to a particular computer, the overall capability for identification of users is greatly enhanced. The threshold of cyberspace is now marked in a new fashion.

The identification of a seemingly anonymous Internet user can easily follow from an IP address. Connection to a website normally reveals a user's IP address to the host website, and look-up tools available on the Internet permit certain information to be revealed about an IP address.[126] Such details include the hostname and a map indicating its general location.[127] With such access to a user's IP address, a third party need only have the user's ISP match the relevant account information to the IP address assigned to that user's computer in order to personally identify the account holder.

To be sure, IP addresses do not directly identify a particular person. Instead, an IP address is assigned to a specific computer or internet device in order to allow it access to the Internet. Therefore, identification does not follow automatically from access to an IP address alone. For example, a computer may be used by multiple members of a household. Not surprisingly then, some companies have

---

123 For this reason, the Electronic Frontier Foundation finds the concern about linking a person to her searches lessened in cases where the ISP assigns dynamic IP addresses. *Six Tips To Protect Your Search Privacy*, ELEC. FRONTIER FOUND. (Sept. 2006), https://www.eff.org/wp/six-tips-protect-your-search-privacy.

124 John B. Horrigan, *Home Broadband 2008*, PEW INTERNET & AM. LIFE PROJECT 2 (Jul. 2, 2008), http://pewinternet.org/Reports/2008/Home-Broadband-2008.aspx.

125 BAHADUR ET AL., *supra* note 119, at 194.

126 For a selection of these websites, see IP-LOOKUP, http://ip-lookup.net/ (last visited Oct. 31, 2011); NETWORK-TOOLS.COM, http://network-tools.com/ (last visited Oct. 31, 2011); *IP Lookup*, WHATISMYIPADDRESS.COM, http://whatismyipaddress.com/ip-lookup (last visited Oct. 31, 2011).

127 For a website offering this information, see *IP Lookup*, WHATISMYIPADDRESS.COM, http://whatismyipaddress.com/ip-lookup (last visited Oct. 31, 2011).

argued that an IP address is non-PII.[128] Yet this argument is misleading.[129] In the case of the IP address, various other clues can readily be used to identify particular individuals. These clues include analysis of the websites that a person visited during a particular session of web surfing. For example, a family member may check her work webmail and use a unique password to do so. In this fashion, it will be possible to distinguish one member of the family from another.

IP addresses can also be readily linked to individuals who post information online. In one notable example, an anonymous person wrote defamatory information in a Wikipedia entry for John Seigenthaler, who had been an assistant to Attorney General Bobby Kennedy during the Kennedy Administration. The anonymous person wrote that Seigenthaler "was thought to have been directly involved in the Kennedy assassinations of both John, and his brother, Bobby. Nothing was ever proven."[130] The incident gathered national attention when Seigenthaler wrote an editorial in *USA Today* condemning the defamation.[131]

As it turned out, Wikipedia had kept a record of the IP address listed for the person who posted the contested information in the Seigenthaler biography.[132] A third party who read about the incident was able to obtain the IP address from Wikipedia and use IP lookup software to trace it to the address of a delivery company in Nashville. A *New York Times* reporter then called the company, and this additional publicity, as well as the likelihood of an internal company investigation, prompted the person who wrote about Seigenthaler to confess, apologize, and resign from his job.[133]

While Seigenthaler did not wish to file a lawsuit against the ISP to unmask the identity of the user in question, so-called "John Doe"

---

[128] For examples of companies successfully making this argument in litigation, see Klimas v. Comcast Cable Commc'ns, Inc., 465 F.3d 271, 276 n.2 (6th Cir. 2006); Johnson v. Microsoft Corp., No. C06-0900RAJ, 2009 U.S. Dist. LEXIS 58174, at *13 (W.D. Wash. June 23, 2009); Columbia Pictures Indus. v. Bunnell, No. CV 06-1093 FMCJCX, 2007 WL 2080419, at *3 n.10 (C.D. Cal. May 29, 2007).

[129] *Cf.* Pineda v. Williams-Sonoma Stores, Inc., 100 Cal. Rptr. 3d 458, 461 (discussing the potential for zip codes to be PII).

[130] John Seigenthaler, *A False Wikipedia "Biography,"* USA TODAY, Nov. 30, 2005, at A11.

[131] Katharine Q. Seelye, *A Little Sleuthing Unmasks Writer of Wikipedia Prank*, N.Y. TIMES, Dec. 11, 2005, at 1.51.

[132] Seigenthaler, *supra* note 130, at A11. BellSouth, the ISP for the account, refused to reveal the account information of the account holder without a court order, and Seigenthaler declined to file a so-called "John Doe" lawsuit to unmask the identity of that person. *Id.*

[133] Seelye, *supra* note 131.

cases are now common.[134] Although case law is far from settled, ISPs generally require that an entity seeking account information for an IP address provide them with a subpoena.[135] The standard for obtaining such subpoenas is relatively lenient—courts require the party seeking the data to show that the identity is needed as a key element in a case, and that this identity information is not otherwise available to the party seeking it.[136] The Recording Industry Association of America (RIAA) has recently made active use of John Doe actions to unmask individuals who are engaged in file-sharing of copyrighted works.[137] Revealing IP address information has made legal actions possible against tens of thousands of patrons of sites such as BitTorrent.[138]

Finally, IP addresses can lead to identification of a person even without account information from an ISP. Three computer scientists have demonstrated a way to identify a person based on a "trail of seemingly anonymous and homogenous data left across different locations."[139] Their paper provides the example of "an online consumer [who] visits websites, leaving the IP address of his computer logged at each site visited."[140] This consumer can be identified without a John Doe lawsuit because at other sites "he may also provide explicitly identifying information; for example, his name and address are provided to complete a purchase."[141] As the authors explain, "[b]y examining the trails of which IP addresses appeared at which locations in the de-identified data and matching those visit patterns to which customers appeared in the identified customer lists, IP addresses can be related to names and addresses."[142]

---

134 For a discussion of John Doe suits, see Patrick Fogarty, *Major Record Labels and the RIAA*, 9 HOUS. BUS. & TAX L.J. 140, 156–58 (2009) (explaining how John Doe suits, which are usually granted *ex parte*, permit the RIAA to discover the identities of ISP customers) and Julie E. Cohen, *Pervasively Distributed Copyright Enforcement*, 95 GEO. L.J. 1, 16–17 (2006) (describing how identified John Doe defendants usually settle outside of court and for "relatively small monetary settlement[s]").

135 Cohen, *supra* note 134, at 16–17.

136 Fogarty, *supra* note 134, at 156–57.

137 Paul Roberts, *RIAA Sues 532 'John Does*,' PCWORLD (Jan. 21, 2004), *available at* http://www.pcworld.com/article/114387/riaa_sues_532_john_does.html.

138 Casey J. Dickinson, *Movie Industry Seeks Cornell Pirate*, 19 BUS. J. CENT. N.Y., Dec. 9, 2005, at 1. As Cohen notes: "Most defendants quickly settle for an amount reported to be in the $3000–$6000 range. Because these lawsuits typically have low filing and overhead costs, the civil settlement program has become a profit center for the industry." Cohen, *supra* note 134, at 17.

139 Bradley Malin, Latanya Sweeney & Elaine Newton, *Trail Re-identification* 1 (Carnegie Mellon Univ., Sch. of Computer Sci., Data Privacy Lab., Tech. Report No. LIDAP-WP12), *available at* http://dataprivacylab.org/dataprivacy/projects/trails/index3.html.

140 *Id.* at 2.

141 *Id.*

142 *Id.*

Like the Wikipedia user who wrote about Seigenthaler, people may be surprised when linked to something they presumed that they had said or done anonymously. They may be dismayed when a notice of a lawsuit arrives from the RIAA after they visit BitTorrent anonymously. They may be amazed that even scattered visits to websites can enable their identity to be linked to their IP address. These initial examples, all centered around IP addresses, demonstrate only some of the ways in which online anonymity is often a mirage. In today's information age, it is increasingly difficult for data to remain unidentified. We now explore additional dimensions of this problem.

B. The Re-Identification of Data: Goodbye Non-PII?

Technology is now posing a considerable challenge to the non-PII side of the dichotomy. Computer scientists are finding ever more inventive ways to combine various pieces of non-PII to make them PII. This trend shows up, for example, in some remarkable demonstrations of how supposedly de-identified information can be re-personalized. America Online's (AOL) release of search queries and research by Latanya Sweeney both illustrate this point.

In 2006, AOL released twenty million search queries for the benefit of researchers.[143] These queries were considered to be fully anonymized. Yet, reporters from the *New York Times* quickly demonstrated that at least some of this information could easily be re-personalized. The reporters showed how they were able to identify one person based on her search queries—User No. 4417749.[144] According to the article:

> [S]earch by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."
>
> It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.[145]

---

[143] Michael Barbaro & Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, at A1.

[144] *Id.*

[145] *Id.*

AOL ultimately apologized for the disclosure. It recognized that it had violated the privacy of its users despite its attempts to anonymize the data.[146]

The privacy debacle at AOL demonstrates a major problem with defining some types of information as non-PII: Technology increasingly enables the combination of various pieces of non-PII to produce PII. According to a study done by computer science professor Latanya Sweeney, the combination of a ZIP code, birth date, and gender will be sufficient to identify 87% of individuals in the United States.[147] These pieces of data are all generally considered to be non-PII. Moreover, they are not intimate, embarrassing, or particularly sensitive. Nevertheless, combining them will identify the vast majority of Americans. According to Sweeney, "for much of the adult population in the United States, local census information can be used to re-identify de-identified data since other personal characteristics, such as gender, date of birth, and ZIP code, often combine uniquely to identify individuals."[148] As a further example, during the 1960s, the United States government began to sell census data to marketers that consisted only of addresses without names.[149] Marketing companies, however, were able to link names to these addresses by drawing on data in telephone books and voter registration lists.[150]

A further problem with non-PII is the wide availability of so much information about people. This phenomenon of data availability heightens the ability to turn non-PII into PII. This aspect of the re-personalization problem stems from a privacy problem that we will call "aggregation."[151] Aggregation involves the combination of various pieces of data.

We have already seen an example of this phenomenon involving IP addresses and the identification of individuals, even without a John Doe lawsuit. Visitation patterns can permit the use of an IP address to

---

146 Anick Jesdanun, *AOL: Breach of Privacy Was a Mistake*, WASH. POST (Aug. 7, 2006), http://washingtonpost.com/wp-dyn/content/article/2006/08/07/AR2006080700790.html.

147 Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* 1 (Carnegie Mellon Univ., Sch. of Computer Sci., Data Privacy Lab., Working Paper No. 3, 2000).

148 Latanya Sweeney, *Maintaining Patient Confidentiality When Sharing Medical Data Requires a Symbiotic Relationship Between Technology and Policy* 5 (Artificial Intelligence Lab., Mass. Inst. of Tech., Working Paper No. AIWP-WP344, 1997), *available at* http://privacy.cs.cmu.edu/dataprivacy/projects/law/aiwp.pdf.

149 ERIK LARSON, THE NAKED CONSUMER: HOW OUR PRIVATE LIVES BECOME PUBLIC COMMODITIES 41, 52 (1992).

150 *Id.* at 218–19.

151 *See* DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 117–21 (2008) (explaining the mechanics of aggregation).

link de-identified data to names and addresses.[152] Additionally, a person who thinks she is anonymous while using certain sites may provide explicitly identifying information, as when completing a purchase.[153] A further example involves a study of Netflix movie rentals by two computer scientists, Arvind Narayanan and Vitaly Shmatikov. The Narayanan-Shmatikov research demonstrated that at least some people in a supposedly anonymous data set could be identified based on how they rated movies on a publicly available website.[154] This example is worth exploring in some detail.

Netflix is a popular online movie rental service, which made a supposedly de-identified database of ratings publicly available as part of a contest to improve the predictive capabilities of its movie recommending software. Narayanan and Shmatikov found a way to link this data with the movie ratings that participating individuals gave to films in the Internet Movie Database (IMDb), a popular website with information and ratings about movies.[155] They concluded: "Given a user's *public* IMDb ratings, which the user posted voluntarily to selectively reveal *some* of his . . . movie likes and dislikes, we discover *all* the ratings that he entered *privately* into the Netflix system, presumably expecting that they will remain private."[156] As this study demonstrates, a single piece of non-PII does not exist alone. Rather, such data form only part of a shifting landscape in which extensive information is available about almost every individual. This rich tableau of available information poses significant concerns for information privacy.

The more information about a person that is known, the more likely it becomes that this information can be used to identify that person or to determine further data about her. When aggregated, information has a way of producing more information, such that de-identification of data becomes more difficult. Thus, it becomes possible to look for overlap in the data and then to link up different bodies of data.

---

[152] Malin, Sweeney & Newton, *supra* note 139, at 2.

[153] *Id.*

[154] Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets* (2008 IEEE Symp. on Sec. and Privacy 111, Feb. 5, 2008), *available at* http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.

[155] INTERNET MOVIE DATABASE, http://www.imdb.com (last visited Oct. 31, 2011).

[156] Narayanan & Shmatikov, *supra* note 154, at 16. The authors concede that the results did not "imply anything about the percentage of IMDb users who can be identified in the Netflix Prize dataset." *Id.* For an insightful technical analysis of the limits of the Netflix study and how it is has been misunderstood, see Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J.L. & TECH., manuscript at 25–26 (forthcoming Dec. 2011–Jan. 2012).

This discussion is far from hypothetical—data miners and marketers currently draw on these techniques. For example, suppose the following anonymous record exists about an individual:

Name: Unique alpha-numerical identifier
Age: 13
Favorite Toy: Legos
Favorite Movie: Batman
Favorite Candy: Snickers
Favorite Restaurant: McDonald's
Zip Code: 20052

In a world without other sources of data, this information would likely remain anonymous. But in today's world, there are countless other data sources. This seemingly anonymous child might have a profile at a social network website, such as Facebook:

Name: Billy Doe
Age: 13
Location: I live in Washington, DC
Narrative: I love to build things with Legos. I love Snickers bars. I recently saw the Batman movie and thought it was the coolest movie ever!

Another database might have the following information:

Name: William Doe
Date of Birth: 04-04-1996
Address: 2000 H Street, NW, Washington, DC 20052

Piecing together these pieces of information, one can link the anonymized record to William Doe.

In *Northwestern Memorial Hospital v. Ashcroft*, Judge Richard Posner aptly recognized that de-identified data can readily be re-identified.[157] The government had subpoenaed patient records of women who had undergone partial birth abortions. The records were to be redacted so that the identities of the women would not be disclosed.[158] Despite the redaction, the court quashed the subpoena, concluding that de-identified patient records still violated the patients' right to privacy.[159] As Judge Posner reasoned:

> Some of these women will be afraid that when their redacted records are made a part of the trial record in New York, persons of their acquaintance, or skillful "Googlers," sifting the information contained in the medical records concerning each patient's medical

[157] 362 F.3d 923, 929 (7th Cir. 2004).
[158] *Id.* at 925.
[159] *Id.* at 932–33.

> and sex history, will put two and two together, "out" the 45 women, and thereby expose them to threats, humiliation, and obloquy.[160]

Through his concept of "skillful 'Googlers,'" Judge Posner identified only one of the many powerful tools that now exist for retrieving de-identified information, analyzing it, and linking it to other information in order to re-personalize it.[161]

Research by computer scientists indicates the legitimacy of Posner's concern about the unmasking of information that is considered non-PII. For example, Professor Sweeney notes that in many health care data sets, there will be unique data about people that can be used to identify them even when they are not explicitly identified in the data set. As she proposes, medical data stripped of identifying information such as names, addresses, phone numbers, and social security numbers, is not really anonymized because "the remaining data can be used to re-identify individuals by linking or matching the data to other databases or by looking at unique characteristics found in the fields and records of the database itself."[162] Further, she observes that in medical facilities, "[n]urses, clerks and other hospital personnel will often remember unusual cases and in interviews may provide additional details that help identify [a] patient."[163] In another study, Sweeney and co-author Bradley Malin demonstrate that "genomic data can often be re-identified in a distributed health environment."[164] Finally, as we have already noted, Ohm has brought the complexities of anonymization to the attention of the legal academy.[165]

C. The Problem of Changing Technology and Information-Sharing Practices

The technical difficulties of de-identifying data raise a challenge to current concepts of PII. Yet, as we have demonstrated in Part I, it is precisely this idea that serves a gatekeeping function at present in information privacy law. A further challenge to current concepts of PII is that technology is constantly changing. As early as 1977, the Privacy Protection Study Commission observed that "[a] major problem created by the widespread adoption of computer and tele-

---

160 *Id.* at 929.

161 For an example of an earlier court that recognized these same issues, see *Parkson v. Cent. DuPage Hosp.*, 435 N.E.2d 140, 144 (Ill. App. 1982).

162 Sweeney, *supra* note 148, at 6.

163 *Id.* at 1.

164 Bradley Malin & Latanya Sweeney, *How (not) To Protect Genomic Data Privacy in a Distributed Network: Using Trail Re-identification To Evaluate and Design Privacy Protection Systems*, 37 J. BIOMED. INFORMATICS 179, 191 (2004).

165 Ohm, *supra* note 6, at 1716–31.

communications technology to personal-data record keeping is the inability to anticipate and control future use of information."[166] The Commission noted that systems were developed and then modified with an eye only to immediate and specific needs. There was a lack of consideration of the long-term implications of the computerization of other areas of record-keeping, and such developments were, at any rate, difficult to predict.[167]

The same problem exists for the distinction between PII and non-PII. The line between PII and non-PII is not fixed, but depends upon technology. Thus, today's non-PII might be tomorrow's PII. New and surprising discoveries are constantly being made about ways of combining data to reveal other data. For example, a recent study by Alessandro Acquisti and Ralph Gross demonstrates that people's social security numbers can be predicted based on other pieces of data such as birth date and birth location.[168] As they state, " it is possible to predict, entirely from public data, narrow ranges of values wherein individual social security numbers are likely to fall."[169] The implications of this study are dramatic, as Acquisti and Gross state, "Unless mitigating strategies are implemented, the predictability of social security numbers exposes them to risks of identity theft on mass scales."[170]

In addition to new technological abilities that permit the re-identification of data, another important factor that facilitates re-identification of data is the proliferation of personal information online and in offline record systems. In particular, corporate practices now play an important role in shaping the amount and kinds of information that are available online. To illustrate, we can consider the Facebook Beacon system and Google Buzz.

In 2007, Facebook introduced the Beacon online ad system, which tracked users' online activities on third-party websites. Without initial warning to Facebook users, this ad system shared collected information from the third-party sites not only with Facebook, but with a user's Facebook friends.[171] Thus, activities such as purchasing a product, signing up for a new service, or placing an item on a wish list would lead to personal information flowing to one's friends and to

---

166 PRIVACY COMM'N, TECHNOLOGY & PRIVACY, *supra* note 21, at 26.

167 *Id.*

168 Alessandro Acquisti & Ralph Gross, *Predicting Social Security Numbers from Public Data*, 106 PNAS 10975 (2009).

169 *Id.* at 10975.

170 *Id.*

171 Caroline McCarthy, *Facebook's Zuckerburg: "We Simply Did a Bad Job" Handling Beacon*, CNET (Dec. 5, 2007, 11:41 AM), http://news.cnet.com/8301-13577_3-9829526-36.html.

Facebook.[172] As a further example, in 2010 Google introduced Buzz, its own social networking platform in a fashion that also led to a widespread proliferation of users' personal information. Buzz permits users to share updates, comments, photographs, videos, and other information through posts, which are called "buzzes."[173] However, as the FTC noted in its complaint against Google, "Without prior notice or the opportunity to consent, Gmail users were, in many instances, automatically set up with 'followers' (people following the user). In addition, after enrolling in Buzz, Gmail users were automatically set up to 'follow' other users."[174]

In sum, whether information can be re-identified depends on technology and corporate practices that permit the linking of de-identified data with already-identified data.[175] Moreover, as additional pieces of identified data become available, it becomes easier to link them to de-identified data because there are likely to be more data elements in common.

D. The Ability To Identify Depends on Context

In many cases, a determination of whether some data are PII as opposed to non-PII is complex because information does not readily fit into one of these two categories. As noted above, identifiability is a complex concept because of the changing landscape of technology, as well as social and corporate practices. Abstract determinations of whether a given piece of information is PII are insufficient because the ability to identify information is driven by context.

Consider Internet search queries that are anonymized. A search query is the information that a person types into a search engine like Google.[176] In the abstract, if anonymized, search queries appear to be non-PII. Recall, though, that AOL mistakenly believed such information was anonymous when it released its search-query data.[177] Yet it is

---

[172] *Id.* The social networking site also structured the Beacon system such that these data were initially transmitted without a user being able to opt out of the program. *Id.* In 2010, a federal judge approved a $9.5 million settlement of a class action lawsuit concerning this matter. Settlement Agreement Exhibit 2 at 2, Lane v. Facebook Inc., No. 5:08-cv-03845-RS (N.D. Cal., Aug. 12, 2008).

[173] Complaint ¶ 7, In re Google Inc., File No. 102-3136, 2011 WL 1321658 (F.T.C. Mar. 30, 2011).

[174] *Id.* The program automatically shared user information even if a Gmail user selected the "Nah, go to my inbox" choice from the initial Buzz screen. *Id.* ¶ 8.

[175] *See* FTC, PROTECTING PRIVACY, *supra* note 2, at 37–38 (highlighting examples of AOL's and Netflix's data collection practices to demonstrate the potential identifiability of de-identified data).

[176] For a general discussion of privacy at Google and some of the international implications of its privacy policies, see JOHN BATTELLE, THE SEARCH 189–210 (2005).

[177] *See supra* notes 143–46 and accompanying text.

not possible to make an abstract judgment of whether or not a search query can become PII. It depends upon the nature of the search in which the subject person had been engaged. If the only data is a single search query for something general (such as a search for "poodles") then identifying a specific user might be difficult. But if the user has engaged in a highly specific search, or multiple searches, she becomes more identifiable. At some point, a search allows a person to be readily identifiable.

In *Gonzales v. Google*, the government had sought to obtain from Google a sample of user search queries.[178] The district court quashed the subpoena on privacy grounds and reasoned that

> [a]lthough the Government has only requested the text strings entered . . . basic identifiable information may be found in the text strings when users search for personal information such as their social security numbers or credit card numbers. . . . The Court is also aware of so-called "vanity searches," where a user queries his or her own name perhaps with other information.[179]

The court's example of the "vanity search" is an excellent one. A search for one's own name combined with just a few other searches will readily allow the de-masking of the data subject.

Thus, the question of whether search queries are PII cannot be answered in the abstract. Trying to classify search queries as PII or non-PII in order to fit them into the binary system of much current privacy regulation is futile. The consequences of search queries will depend upon the context, such as the specific things searched for, as well as what other information is already available about a user. Nonetheless, the distinction between PII and non-PII is almost always made in the abstract in privacy regulation. As Part IV demonstrates, our concept of PII 2.0 responds to this situation by requiring context-based evaluations around a standard-based definition of PII.

III
Behavioral Marketing and the Surprising Irrelevance of PII

The problems with the current approach to PII are most dramatically illustrated by looking at the burgeoning practice of behavioral marketing. This practice—sometimes referred to as targeted marketing—involves examining the behavioral patterns of consumers in order to target advertisements to them based on their presumed preferences. Public interest groups, scholars, and government regulatory

178 234 F.R.D. 674 (N.D. Cal. 2006).
179 *Id.* at 687.

agencies such as the FTC have examined this practice and raised objections to it on privacy grounds.[180] As we demonstrate in this Part, behavioral marketing is conducted in ways that challenge traditional conceptions of PII. In particular, we explore behavioral marketing in the context of selling food products to children, an issue with profound implications because of the growing health crisis of obesity among minors.

A. From Mass Marketing to Behavioral Marketing

In the past, companies engaged in mass marketing. They targeted their audience using the general demographical information regarding those watching particular TV shows or reading particular periodicals. Today, companies instead direct offerings to a specific individual based on information collected about the particular characteristics, preferences, and behavior of this person. The holy grail of modern advertising is "one-to-one" marketing.[181] The result of such marketing is to create "advertising crafted to uniquely engage" each individual.[182] This technique is called behavioral marketing; the idea is for advertisers to record a person's behavior, analyze it, and shape the kinds of offers directed to that party based on the patterns that emerge from this collected data.

The key recent development has been, moreover, the ability of companies to engage in behavioral marketing without using PII—at least as this term is traditionally understood. In this section, we trace the transformation from the epoch of mass marketing to the contem-

---

[180] For the views of an NGO, see the insightful reports by Jeff Chester and Kathryn Montgomery under the sponsorship of the Berkeley Media Studies Group (BMSG). These include JEFF CHESTER ET AL., ALCOHOL MARKETING IN THE DIGITAL AGE 2–12 (2010) (noting alcohol brands' use of digital marketing, including "interactive virtual universe[s]" and behavioral marketing, to advertise to youth); JEFF CHESTER & KATHRYN MONTGOMERY, INTERACTIVE FOOD & BEVERAGE MARKETING: TARGETING CHILDREN AND YOUTH IN THE DIGITAL AGE 31–36 (2007) [hereinafter CHESTER & MONTGOMERY, INTERACTIVE MARKETING] (discussing "behavioral profiling" as one of the ways that food and beverage advertisers reach children in "the new digital marketing landscape").

The FTC reported on behavioral marketing in depth in 2009. *See* FED. TRADE COMM'N, SELF-REGULATORY PRINCIPLES FOR ONLINE BEHAVIORAL ADVERTISING (2009) [hereinafter FTC, SELF-REGULATORY PRINCIPLES]. For an example of international attention to the topic, see Article 29 Data Protection Working Party, *Opinion 2/2010 on Online Behavioural Advertising*, 00909/10/EN/WP 171 (June 22, 2010).

[181] *See* DON PEPPERS & MARTHA ROGERS, ENTERPRISE ONE TO ONE 30–78 (1997) (providing background on one-to-one marketing and noting that targeting customers individually can vastly improve marketing results); DON PEPPERS & MARTHA ROGERS, THE ONE TO ONE FUTURE 138–72 (1993) (indicating that firms that engage in one-to-one marketing will be more successful than those that engage in mass marketing).

[182] JEFF CHESTER & KATHRYN MONTGOMERY, INTERACTIVE FOOD & BEVERAGE MARKETING: AN UPDATE 2 (2008).

porary age of one-to-one marketing. We then explore modern information exchanges, and the way in which companies increasingly structure this process to be free of the collection of PII as it is defined in law.

1. Modern One-to-One Marketing

The age of merchandizing on a mass scale began in the 1850s with department stores displaying goods that were marked with uniform prices for all to see.[183] Such mass marketing was a global phenomenon—for example, Émile Zola devoted a brilliant novel, *Au Bonheur des Dames*, to the events in a department store, a new kind of industrial organization that fascinated him.[184] Throughout the Western world, the mass-merchandizing approach proved stable for over a century. In Joseph Turow's words, the result was "a fairly egalitarian and transparent marketplace, with products and prices that all could see."[185] Along with mass merchandising came mass marketing, as advertisers and other "hidden persuaders" during this period exploited broad patterns drawn from demographic data to effectively target groups of consumers.[186]

In contrast, contemporary behavioral marketing targets individuals—by drawing on digital information about their past behavior, as well as on knowledge about how other parties similarly situated have behaved. Already in 1971, Arthur Miller warned that computerization would permit "simulation activities involving the prediction of an individual's or a group's behavior."[187] Miller was worried about the possibility of future "attempts at human manipulation" by organizations using computers to affect and shape their customers' behavior.[188] Digital technology and the Internet have now made Miller's prediction a daily occurrence; the goal of modern marketing is for a targeted tracking of individuals to customize products, services, and prices.

In the twenty-first century, targeted marketing now occurs both online and offline in highly sophisticated and potent ways. As Jeffrey

---

183 JOSEPH TUROW, NICHE ENVY: MARKET DISCRIMINATION IN THE DIGITAL AGE 23–24 (2006).

184 ÉMILE ZOLA, AU BONHEUR DES DAMES (Robin Buss ed. & trans., Penguin Group 2001) (1883).

185 TUROW, *supra* note 183, at 180. In this same vein, Turow also writes of a "democratization of shopping." *Id.* at 179.

186 For a popular account of how advertisers enlisted the aid of social scientists in the 1950s, see VANCE PACKARD, THE HIDDEN PERSUADERS (1957). As Packard described, advertisers and other "symbol manipulators" were "sitting at the feet of psychiatrists and social scientists (particularly psychologists and sociologists) who ha[d] been hiring themselves out as 'practical'consultants or setting up their own research firms." *Id*. at 7.

187 MILLER, *supra* note 20, at 42.

188 *Id.* at 42–43.

Chester warned in 2007, "[a]dvertisers are developing increasingly sophisticated technologies designed to track, analyze, and persuade us in the Internet era."[189] Marketers draw on extensive databases, which sometimes combine people's online and offline behavior. They are able to cross-reference online activity with offline records including home ownership, family income, marital status, zip code, and a host of other information, such as one's recent purchases as well as favorite restaurants, movies, and TV shows.[190]

Individuals can now be tracked across different websites or digital media. Moreover, online advertising networks follow people around the web.[191] In this new paradigm, an advertising network first places a tracking file on a user's computer, which allows the company to gather information about a person's behavior and preferences as she surfs the Internet.[192] In this tracking process, the advertising industry relies on diverse technology to conduct such tracking, including basic "cookies," "flash cookies," and "beacons."[193] Some technology, particularly the beacon, or "Web bug," permits real-time observation of a user's activity on an Internet page, including where one's mouse moved and the information that one typed, such as search queries or personal information that an individual filled into a form.[194] The cutting edge of this technology continues to advance, with some ISPs starting to monitor the content of their customers' internet activity through a practice known as deep-packet inspection.[195]

Marketers today engage in a pinpoint process that focuses on ever smaller groups of people.[196] Instead of companies selling ads for specific websites, advertisers now seek to buy access to individuals

---

189 JEFF CHESTER, DIGITAL DESTINY 128 (2007).

190 Emily Steel & Julia Angwin, *On the Web's Cutting Edge, Anonymity in Name Only*, WALL ST. J., Aug. 4, 2010, at A1; *see also* Julia Angwin, *The Web's New Gold Mine: Your Secrets*, WALL ST. J., July 31, 2010, at W1 (noting that marketers have access to a consumer's favorite movies, television shows, and news preferences); Jessica E. Vascellaro, *Google Agonizes on Privacy as Ad World Vaults Ahead*, WALL ST. J., Aug. 10, 2010, at A1 (describing Google's access to a "vast trove of data," including one user's recent purchase of a TV).

191 Vascellaro, *supra* note 190.

192 Angwin, *supra* note 190.

193 *Id.*

194 *Id.*

195 *See Deep Packet Inspection and Privacy*, ELECTRONIC PRIVACY INFORMATION CENTER, http://epic.org/privacy/dpi (last visited Oct. 31, 2011) (describing deep-packet inspection as involving inspection of the contents of data transmitted across the Internet at the "packet" level, that is, through examination of the individual packages of bytes in which all information is sent over the Internet thereby allowing the determination of all contents of unencrypted data).

196 TUROW, *supra* note 183, at 1–3, 8.

who fit a certain profile.[197] Information collected about consumers is packaged into profiles, which are sold on new kinds of "stock-market-like exchanges."[198] The company that buys the underlying information can then use it to serve targeted ads. As an investigatory series in the *Wall Street Journal* states, "[i]nformation about people's moment-to-moment thoughts and actions, as revealed by their online activity, can change hands quickly. Within seconds of visiting eBay.com or Expedia .com, information detailing a Web surfer's activity there is likely to be auctioned on [a] data exchange."[199] Information about an individual's browsing habits sells for as little as a tenth of a cent online.[200] All those slivers of a cent nonetheless add up—marketing online is a billion dollar industry and remains a growth field.[201]

Behavioral marketing also depends on so-called analytics to decide how to approach customers.[202] Analytics provide a way for organizations to draw on the great quantities of information in their control or available from third parties and to use the data to make better decisions and to create new products and services.[203] In the definition of Thomas Davenport and Jeanne Harris, two leading authorities on this technology, "analytics" refers to "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions."[204] The idea is to take the information that entities have or to which they can gain access, and to convert it to actionable knowledge.[205] This approach is now popular in the corporate world—as a blogger on the website of the *Harvard Business Review* concisely observed in September 2010, "[a]nalytics are now king."[206]

---

197 *Id.* at 8.

198 Angwin, *supra* note 190.

199 *Id.*

200 *Id.*

201 According to one estimate, online advertising is a $23 billion a year industry. INTERACTIVE ADVER. BUREAU, INTERNET ADVERTISING REVENUE REPORT 3 (2009), *available at* http://www.iab.net/media/file/IAB_PwC_2008_full_year.pdf.

202 THOMAS H. DAVENPORT & JEANNE G. HARRIS, COMPETING ON ANALYTICS 86–91 (2007).

203 Thomas H. Davenport, *Competing on Analytics*, HARV. BUS. REV., Jan. 2006, at 98, 101, 104, 106–07.

204 DAVENPORT & HARRIS, *supra* note 202, at 7.

205 As Thomas Davenport and co-authors explain, "[t]he analytic process makes knowledge from data." Thomas H. Davenport et al., *Data to Knowledge to Results*, CAL. MGT. REV., Winter 2001, at 117, 128. Corporations in the information age are drowning in data, but without drawing on the right technology, strategies, and "analytic resources," the data will not be turned into the "sort of knowledge that can inform business decisions and create positive results." *Id.* at 117–22.

206 Michael Fertig, *Hire Great Guessers*, HARV. BUS. REV. BLOG NETWORK (Sept. 2, 2010, 8:30 AM), http://blogs.hbr.org/cs/2010/09/hire_great_guessers.html.

Behavioral marketing has also been controversial. Much of the reaction, quite understandably, has been at a visceral level. For example, newspapers have talked of "creepy" and secret practices.[207] At the same time, and as a general matter when directed toward adults, advertising is an accepted and inescapable part of life, and on occasion, Americans even look forward to it.

There are two core objections to behavioral advertising when directed toward adults. The first has to do with transparency and the second with money. As for transparency, behavioral marketing takes place today in a multi-channel process about which individuals generally receive scant information about the data that organizations collect about them or how that information is used to shape interactions with them. As the *Wall Street Journal* observes, "the tracking of consumers has grown both far more pervasive and far more intrusive than is realized by all but a handful of people in the vanguard of the industry."[208] This new kind of tracking largely takes place in the shadows, and Americans, not surprisingly, have responded with deep unease.[209] The instinct of many people is to view these practices, at least in absence of knowledge as to how they take place, as deceptive, otherwise unfair, or even as a force capable of chilling their free behavior.[210] Moreover, the very complexity of the marketing ecosystem heightens the general ignorance of these corporate techniques, and reduces the value of the tools that some companies are making available to users.[211]

---

[207] *E.g.* Jeff Gelles, *When 'Behavioral Marketing' Turns Creepy*, PHILADELPHIA INQUIRER: INQUIRING CONSUMER (Feb. 21, 2011, 2:52 PM), http://www.philly.com/philly/business/When_behavioral_marketing_turns_creepy.html; *see also* Julia Angwin & Emily Steel, *Web's Hot New Commodity: Privacy*, WALL ST. J., Feb. 28, 2011, at A1 (indicating that individuals find targeted ads to be "spooky").

[208] Angwin, *supra* note 190.

[209] *See* JOSEPH TUROW ET AL., AMERICANS REJECT TAILORED ADVERTISING 3 (2009) ("[M]ost Americans (66%) do not want marketers to tailor advertisements to their interests."); Lymari Morales, *U.S. Internet Users Ready To Limit Online Tracking for Ads*, GALLUP (Dec. 21, 2010), http://www.gallup.com/poll/145337/Internet-Users-Ready-Limit-Online-Tracking-Ads.aspx ("Internet users are overwhelmingly negative about whether it is OK for advertisers to use their online browsing history to target ads to them . . . .").

[210] *See, e.g.*, Angwin & Steel, *supra* note 207 (quoting former brand marketer who suggested that "[p]eople feel targeted online ads are 'spooky'"); Nicholas Carr, *Tracking is an Assault on Liberty, With Real Dangers*, WALL ST. J., Aug. 6, 2010, at W1 ("Personalization's evil twin is manipulation.").

[211] Some organizations offer users individual controls, such as the ability to opt out from some tracking and to set preferences about the kinds of information that are collected. Google, Microsoft, and Mozilla are among the companies offering such privacy tools. Byron Acohido, *Google Chrome Will Join Other Browsers with Privacy Tools*, U.S.A. TODAY (Jan. 26, 2011, 10:58 AM), http://www.usatoday.com/money/industries/technology/2011-01-26-privacy26_ST_N.htm.

Regarding money, as we have noted, marketing online is a billion-dollar growth industry. Even more specifically, targeted advertisements command a considerable premium in the marketplace.[212] As the FTC noted in December 2010, the more that is known about someone, the more that advertisers will pay to send her an advertisement.[213] Here, the question is how different parties should share in the wealth that the trade in personal information creates. Ideally, a market economy would permit the free price mechanism to set a price for the data.[214] Put less abstractly, *Money* magazine once summed up the matter in these terms: "It's your data, after all; these guys just figured out how to sell it."[215] Yet the lack of transparency regarding practices of data collection and tracking creates an asymmetry of knowledge about existing information collection practices between consumers and the organizations that collect information about them. This information assymetry places consumers at a profound disadvantage in negotiations, such as they may exist, with those who collect their information.[216] In sum, consumer objections to behavioral advertising are real and deserve a policy response. At the same time, and as the next section discusses, these tracking technologies do not rely on PII as the law generally defines it today. This twist complicates the matter of the appropriate policy response.

2. Marketing, Legal Enforcement, and the Question of Adults' PII

In behavioral marketing, companies generally do not track individuals by name. Instead, they use software to build personal profiles that exclude this item but that contain a wealth of details about each individual.[217] In lieu of a name, these personal profiles are associated with a single alphanumerical code that is placed on an individual's computer to track their activity. In one reported case, for example, the

---

[212] It is also possible to combine information that is collected offline with information collected online and use the data to tailor advertisements to specific individuals. CHESTER & MONTGOMERY, INTERACTIVE MARKETING, *supra* note 180, at 33–34. A firm distinction between online and offline marketing no longer exists. Instead, the relevant category is "*digital marketing*," which occurs through multiple channels and different platforms. *Id*. at 3.

[213] FTC, PROTECTING PRIVACY, *supra* note 2, at 37.

[214] *See* Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2055, 2069–76 (2004) (describing a model for propertization of personal data and proposing solutions to combat the alleged market failure in the trade of personal information).

[215] Pat Regnier, *The ID Theft Protection Racket*, MONEY, Sept. 2005, at 112, 116.

[216] *See* Schwartz, *supra* note 214, at 2078–81 ("Consumer ignorance leads to a data market in which one set of parties does not even know that 'negotiating' is taking place.").

[217] Angwin, *supra* note 190.

tracking file consisted of this string: "4c812db292272995e5416a323e79 bd37."[218]

These codes are used to decide which advertisements people see, as well as the kinds of products that are offered to them. For example, Capital One Financial Corporation relies on [x+1], an advertising network, to decide instantaneously the specific type of credit card to show first-time visitors to its website.[219] It uses the ad network's information about people to suggest products to individuals and to steer them toward one card and not another.[220] As [x+1] explains, however, it does not gather the names of the individuals whose information it collects and analyzes.[221] Thus, behavioral marketing occurs without identifying, in the traditional sense, any specific individual.

While advertising networks may not know peoples' names, identification of individuals is nonetheless possible in many cases. As we have seen in Part II.A, this result follows for a number of reasons. For example, enough pieces of information linked to a single person, even in the absence of a name, social security number, or financial information, will permit identification of the individual. Such identification of seemingly non-PII is, moreover, a genuine possibility.

Nonetheless, online companies have attempted to short circuit the discussion of privacy harms and necessary legal reforms by simply asserting that they do not collect PII.[222] The *Wall Street Journal* reports, "[t]he ad industry says tracking doesn't violate anyone's privacy because the data sold doesn't identify people by name, and the tracking activity is disclosed in privacy policies."[223] Or, as the FTC describes the matter, the position of advertisers is that "there is a reduced privacy interest in, and risk of harm from, non-PII."[224]

This defense suggests what may be a further barrier to adequate privacy protection in the future: The online marketing industry may develop a strategic compromise around a PII-based regulatory regime that employs a narrow definition of PII. Currently, behavioral marketing is regulated only to a limited extent, but legal rules in this area

---

218 *Id.*

219 Steel & Angwin, *supra* note 190.

220 *Id.*

221 *Id.*

222 *E.g.*, Steve Stecklow, *On the Web, Children Face Intensive Tracking*, WALL ST. J., Sept. 17, 2010, at A1 (indicating marketers' defense that using tracking devices to collect information from children's websites does not amount to collection of PII).

223 Julia Angwin & Tom McGinty, *Sites Feed Personal Details to New Tracking Industry*, WALL ST. J., July 30, 2010, http://online.wsj.com/article/SB10001424052748703977004575393173432219064.html.

224 FTC, SELF-REGULATORY PRINCIPLES, *supra* note 180, at 31.

may soon increase.[225] Marketers may thus wish to preemptively influence the nature of any forthcoming legal reforms. To grasp the implications of this potential industry strategy, therefore, it is first necessary to understand the current legal landscape.

At present, there is no specific federal statute regulating these marketing practices. Some privacy protection is provided through the oversight of the FTC, which brings actions against companies that violate their own privacy policies.[226] In addition to policing the privacy promises of organizations, the FTC also guards against inadequate security and promotes transparency. We examine each of these roles in turn.

On numerous occasions, the FTC has interpreted a company's behavior that breaches its stated privacy policy as an "unfair or deceptive act" under the Federal Trade Commission Act of 1914.[227] Still, the impact of such FTC regulation is limited: The agency merely ensures that companies live up to their promises, and companies need not promise much.[228] Moreover, studies have shown that few consumers read privacy policies, and that those who do frequently fail to understand them.[229] In fact, consumers commonly and falsely believe that a website with a posted "privacy policy" necessarily provides a positive level of substantive protection.[230]

In addition to the FTC's actions enforcing privacy policies, the agency has taken actions against companies that fail to provide adequate data security.[231] The FTC can bring such enforcement actions

---

225 As the *New York Times* has reported: "The Federal Trade Commission had some sharp words for Internet advertising companies . . . saying that they simply are not disclosing how they collect information about users well enough. And the agency threatened that the industry had better get its act together—or else." Saul Hansell, *The F.T.C. Talks Tough on Internet Privacy*, NYTIMES.COM (Feb. 12, 2009, 3:53 PM), http://bits.blogs.nytimes.com/2009/02/12/the-ftc-talks-tough-on-internet-privacy.

226 For an overview, see PRIVACY AND DATA SECURITY LAW DESKBOOK 16.01 (Lisa J. Sotto ed., 2010) [hereinafter SOTTO, DESKBOOK]; SOLOVE & SCHWARTZ, IPL, *supra* note 52, at 776–87.

227 15 U.S.C. § 45(b) (2006).

228 For examples of enforcement actions, see Vision I Props., LLC, FTC Docket No. C-4135 (Apr. 19, 2005); Bonzi Software, Inc., FTC Docket No. C-4126 (Oct. 7, 2004); Gateway Learning Corp., FTC Docket No. C-4120 (Sept. 10, 2004).

229 Thus, the FTC has spoken of "long, incomprehensible privacy policies that consumers typically do not read, let alone understand." FTC, PROTECTING PRIVACY, *supra* note 2, at iii.

230 *Id.* at 26.

231 *See, e.g.*, ACRAnet, Inc., F.T.C. File No. 092-3088 (Feb. 3, 2011) (settling FTC charges against a credit report reseller for failing to protect its internet portals and thereby furnishing credit reports to hackers, through a consent order barring the reseller from future violations); Twitter, Inc., FTC File No. 092-3093 (June 24, 2010) (settling FTC charges against social networking site Twitter for consumer deception and inadequate data security measures that enabled hackers to obtain unauthorized administrative control of

even in the absence of a data breach, though more typically it acts only once a data spill has occured.[232] As part of these enforcement actions, such as that against Eli Lilly in 2002, the agency also seeks to sanction companies for failing to train their employees about adequate privacy and data security practices.[233]

In the *Eli Lilly* case, the company had sent out an e-mail message to customers who had registered for reminders via its Prozac.com website.[234] When Eli Lilly decided to terminate this service, called the Medi-messenger, it made the mistake of listing the e-mail address of every person who had registered for the reminders in the "to" line of the notification e-mail.[235] As the FTC noted, "[b]y including the email addresses of all Medi-messenger subscribers within the . . . email message, respondent unintentionally disclosed personal information provided to it by consumers in connection with their use of the Prozac .com Web site."[236] The FTC's complaint faulted the company for its "failure to maintain or implement internal measures appropriate under the circumstances to protect sensitive consumer information," including its lack of "appropriate training for its employees regarding consumer privacy and information security" and lack of "appropriate oversight and . . . checks and controls on the process . . . ."[237] In its settlement order, the FTC required Eli Lilly to "establish and maintain an information security program for the protection of personally identifiable information collected from or about consumers in connection with the advertising, marketing, offering for sale, or sale of any

---

Twitter, including access to non-public user information and tweets that consumers had designated as private); Eli Lilly & Co., F.T.C. Docket No. C-4047 (May 8, 2002) (settling FTC charges against a pharmaceutical company for unintentional release of the names and e-mail addresses of Prozac consumers through a consent order necessitating implementation of privacy and security protections and prohibiting any future false or misleading privacy statements).

232 *See, e.g.*, United States v. Am. United Mortg. Co., No. 07C-7064 (N.D. Ill. 2007) (requiring a mortgage company to pay a $50,000 civil penalty for improper disposal of loan documents containing consumers' personal and financial information in an unsecured dumpster); Superior Mortgage Corp., F.T.C. Docket No. C-4153 (Dec. 14, 2005) (settling FTC charges against a lender for failing to provide reasonable Internet security for sensitive customer data and falsely claiming that it encrypted data submitted online); Sunbelt Lending Servs., Inc., F.T.C. Docket No. C-4129 (Jan. 3, 2005) (requiring a company's information security program to be certified by an FTC-chosen expert following the company's failure to implement safeguards to protect its customers' financial information and social security numbers, oversee its service providers, and supervise its loan officers working in remote offices).

233 Complaint ¶ 7, Eli Lilly & Co., F.T.C. Docket No. C-4047 (May 8, 2002).

234 *Id.* ¶ 3–4.

235 *Id.* ¶ 6.

236 *Id.*

237 *Id.* ¶ 7.

pharmaceutical, medical, or other health-related product or service . . . ."[238]

Finally, the FTC is beginning to take a broader approach to privacy based on a concept of transparency. Its policing of privacy notices and enforcement of adequate data security already moved in this direction. More broadly than its "broken promises" approach, however, the agency has begun to take a more substantive approach to disclosure of company behaviors. Thus, in an enforcement action against Sears, which was settled in 2009, the FTC alleged that Sears had engaged in an unfair practice by failing to adequately disclose the extent to which it tracked customers who were paid to use a program that would record their Internet browsing.[239] The FTC acted even though Sears had provided users with a license agreement that, albeit with obscure language, arguably informed users of the tracking.[240] The FTC charged that Sears's failure to provide adequate disclosure of the scope of the data collection was a deceptive act.[241] Its settlement order required Sears to provide clear and prominent disclosure of "the types of data that the [software] will monitor, record, or transmit . . . ."[242]

The FTC's enforcement of transparency continued in 2010 with a settlement against EchoMetrix.[243] In that case, "parental controls" software, marketed as a way for parents to track their children's online activities, also secretly collected data about children's computer activity and fed the resulting information to marketers.[244] The FTC's theory of the case was that the company provided inadequate disclosure of its tracking.[245] Finally, in a 2010 Report, the FTC explicitly emphasized companies' obligation to increase the transparency of their data practices.[246]

With this legal landscape as a backdrop, we now turn to the possible industry strategy of compromise. With Congress appearing eager to enact legislation in this area, affected companies might accept some kind of PII-based regulation while insisting on a restricted definition of PII. This strategy would serve as a kind of red herring to regulators.

---

238 Decision and Order at II, Eli Lilly & Co., F.T.C. Docket No. C-4047 (May 8, 2002).

239 Complaint ¶¶ 13–14, Sears Holdings Mgmt. Corp., F.T.C. Docket No. C-4264 (Aug. 31, 2009).

240 *Id.* ¶ 8.

241 *Id.* ¶ 14.

242 Decision and Order at IA, Sears Holdings Mgmt. Corp., F.T.C. Docket No. C-4264 (Aug. 31, 2009).

243 Stipulated Final Order for Permanent Injunction and Other Equitable Relief, FTC v. EchoMetrix, Inc., No. CV10-5516 (E.D.N.Y. Nov. 30, 2010).

244 Complaint for Permanent Injunction and Other Equitable Relief, FTC v. EchoMetrix, Inc., No. CV10-5516 ¶¶ 8–14 (E.D.N.Y. Nov. 30, 2010).

245 *Id.* ¶¶ 16–18.

246 FTC, PROTECTING PRIVACY, *supra* note 2, at 69–79.

It would allow marketers to achieve the same goals using the same tools of behavioral marketing. Thus, the online marketing industry may be willing to make seemingly large compromises on PII-based privacy regulation because it will still be able to influence consumer behavior that falls outside the definition of PII in ways that many would view as troublesome.

This strategy is well captured by a quotation from a newspaper story about U.S. websites installing as many as one hundred tracking tools at a single time on the computers of people visiting their sites. The reporters stated that "[t]he ad industry says tracking doesn't violate anyone's privacy because the data sold doesn't identify people by name, and the tracking activity is disclosed in privacy policies."[247] The strategy encapsulated in this quotation is two-pronged; it proposes that (1) as non-PII, the information collected and the tracking that it enables fall outside of a PII-based regulatory regime, and (2) as long as they are described in a privacy notice, these same practices fall outside the FTC's oversight of unfair and deceptive practices. In light of these challenges to PII-based regulation, this Article seeks to revisit the current paradigm of PII.

B. Food Marketing to Youth

As the last section demonstrated, marketing has changed greatly over the last century and is now conducted in highly sophisticated and potent ways. Marketing using personal information is focused not only on adults, but also on youth, a term that public health experts define to include children and adolescents. This group's access to large amounts of disposable income and its incompletely-developed tastes and interests make them appealing targets.

Marketing to youth and the impact of legal definitions of PII raise distinct issues in part because American law generally views youth as deserving of special protection. Policymakers are also highly concerned about food marketing to children and eager to act to assist parents in promoting healthy diets. In short, policymakers consider youth, and especially children, to be particularly susceptible to advertising. The fear is of marketing's strong role in influencing youth to consume high-calorie and low-nutrient foods.[248] Therefore, this Article will separately analyze marketing issues that concern youth.

---

[247] Angwin & McGinty, *supra* note 223, at 1.

[248] For a summary of the available research, see INST. OF MED., FOOD MARKETING TO CHILDREN AND YOUTH: THREAT OR OPPORTUNITY? 226–318 (J. Michael McGinnis et al. eds., 2006) [hereinafter FOOD MARKETING TO CHILDREN AND YOUTH].

1. Digital Marketing and the "Net Generation"

In 1998, Don Tapscott announced that "[t]he Net Generation has arrived!"[249] Tapscott identified a new age cohort, the first to grow up surrounded by digital media, and predicted that this generation would be more interested in and affected by interactive digital media than traditional broadcast media such as television.[250] In addition, the commercialization of the web and associated digital devices occurred all but simultaneously with the emergence of these new technologies. As Jeff Chester and Kathryn Montgomery observe, "[t]he rapid growth of the Internet and proliferation of digital media are fundamentally transforming how corporations do business with young people in the twenty-first century."[251] Corporations have actively sought to shape the experiences of minors using these new media.

Digital marketing is also directed more intensively toward youth than adults. In September 2010, the *Wall Street Journal* reported that the fifty most popular websites aimed at children install more tracking devices on personal computers than do the top sites for adults.[252] Digital marketing now occurs around many kinds of products and services, including food products, to young consumers using quite precise information collected about a data subject's characteristics, interests, and hobbies.[253] With access to enormous amounts of disposable income, young people will continue to be an attractive audience for marketers. Due to the public health crisis in the United States around youth obesity, we will now focus on marketing activities that involve food products.

Over the past three decades, the extent of obesity among minors has risen dramatically throughout the United States. In 2007, over one-third of children and adolescents in the United States were obese or overweight.[254] This number represents triple the rate in 1980.[255] A

---

[249] DON TAPSCOTT, GROWING UP DIGITAL: THE RISE OF THE NET GENERATION 1 (1998).

[250] *Id.* at 2–6. Tapscott returned to this generational topic a decade later and found that among other impacts of the digital age, young people "expect speed" in all interactions, "not just in video games." DON TAPSCOTT, GROWN UP DIGITAL: HOW THE NET GENERATION IS CHANGING YOUR WORLD 93 (2009).

[251] CHESTER & MONTGOMERY, INTERACTIVE MARKETING, *supra* note 180, at 13.

[252] Stecklow, *supra* note 222, at 1. The fifty websites most popular with minors placed 4123 pieces of tracking technologies on the newspaper's test computers, which was thirty percent higher than the fifty most popular general-audience U.S. websites. *Id.*

[253] *See* Angwin, *supra* note 190 (detailing the precision with which individuals' preferences can be targeted using data tracking tools); Stecklow, *supra* note 222 (indicating that such data tracking is heavily focused on children).

[254] TRUST FOR AMERICA'S HEALTH, F AS IN FAT: HOW OBESITY THREATENS AMERICA'S FUTURE 7 (2011), *available at* http://healthyamericans.org/reports/obesity2011/Obesity2011Report.pdf.

different study from 2005 raised the possibility that diet-related diseases will cause today's children to be the first generation in the United States to have a shorter life span than their parents.[256] The stakes are high; as the Institute of Medicine has declared, "[p]revention of obesity in children and youth should be a national public health priority."[257]

Experts view the public health crisis of obesity among youth as having multiple roots. Nonetheless, experts agree about the detrimental effect of the marketing of food products to minors. As the Institute of Medicine concisely declared, "[m]arketing works."[258] In more detail, but to the same effect, a review in 2009 of the relevant psychological research on food marketing stated, "[y]outh marketing is powerfully effective, occurs in massive amounts, and is done in forms that thwart cognitive defenses and subvert parents' ability to monitor what their children see and ultimately their ability to provide their children a healthy food environment."[259]

Children and adolescents are highly vulnerable to food marketing. For example, psychologists have shown that "marketing effects occur even in the absence of conscious awareness of marketing stimuli."[260] The net result is summarized by three psychologists: "Marketing practices that promote calorie-dense, nutrient-poor food directly to children and adolescents present significant public health risk."[261]

In light of the migration of youth to the Internet and other digital environments and the power of marketing on decisions about food consumption, it is hardly surprising that the food industry has actively embraced behavioral marketing to minors. As one corporate executive explained, his company moved away from traditional TV advertising into new forms of digital marketing because "[t]he eyeballs have

---

255 *Id.* at 11.

256 S.J. Olshansky et al., *A Potential Decline in Life Expectancy in the United States in the 21st Century*, 352 NEW ENG. J. MED. 1138, 1139, 1141 (2005).

257 INST. OF MED., PREVENTING CHILDHOOD OBESTITY: HEALTH IN THE BALANCE 5 (Jeffrey P. Koplan et al. eds., 2005). The *New York Times* has observed that the current public health focus appears to be shifting from the effort against tobacco to obesity. Duff Wilson, *A Shift Toward Fighting Fat*, N.Y. TIMES, July 28, 2010, at B1.

258 FOOD MARKETING TO CHILDREN AND YOUTH, *supra* note 248, at xiii; *see also* ELIZABETH S. MOORE, IT'S CHILD'S PLAY: ADVERGAMING AND THE ONLINE MARKETING OF FOOD TO CHILDREN 1–2 (2006), *available at* http://www.kff.org/entmedia/upload/7536.pdf (discussing marketing to children).

259 Jennifer L. Harris et al., *The Food Marketing Defense Model: Integrating Psychological Research To Protect Youth and Inform Public Policy*, 3 SOC. ISSUES & POL'Y REV. 211, 255 (2009).

260 *Id.* at 224.

261 *Id.* at 211.

moved."[262] Food and beverage companies are now among the leaders of the new one-to-one digital marketing system.[263]

Sometimes as part of behavioral marketing and sometimes distinctly, advertisers use other advertising techniques to sell food to youth. Among these practices are viral marketing, "advergaming," and individual targeting through social media platforms such as Facebook.[264] The cutting edge of marketing now involves the use of "command centers" in which corporations' staff members directly interact with select consumers. As an example, Gatorade has developed a "Mission Control" center that tracks developments in the social media ether.[265] In the center, Gatorade employees monitor social media posts twenty-four hours a day.[266] Once someone mentions Gatorade in Twitter or other online media, the staff can weigh in and interact with the consumer.[267]

As a final point, some empirical evidence suggests "an especially damaging potential role of targeted food marketing on at-risk minority youth."[268] The available evidence is far from conclusive, however, due to a relative lack of research focusing on minority populations and food marketing. Nonetheless, there are indications that "targeted food marketing efforts that focus on minorities' social identity" heighten "the unhealthy influence of these messages."[269]

2. *Marketing, Legal Enforcement, and the Question of Children's PII*

The same basic issues concerning PII in the context of digital marketing arise for youth as well as for adults. Companies now track young people through personal profiles that exclude names, but contain a wealth of details about the individual. The law responds differ-

---

262 CHESTER & MONTGOMERY, INTERACTIVE MARKETING, *supra* note 180, at 13–14.

263 *Id.* at 61.

264 On viral marketing and other techniques, see CHESTER & MONTGOMERY, INTERACTIVE MARKETING, *supra* note 180, at 2 (2008). On Facebook's use of the "like" button, to allow "effective word-of-mouth marketing on a large, global scale," see FACEBOOK, BUILDING YOUR BRAND ON FACEBOOK 16–17 (2010), http://ads.ak.facebook.com/ads/FacebookAds/Facebook_MediaKit_2010_US.pdf.
Advergaming is a form of "branded entertainment" in which a brand, such as M&M or Oscar Mayer Lunchables is placed within a digital entertainment property. MOORE, *supra* note 258, at 1.

265 Valerie Bauerlein, *Gatorade's 'Mission': Sell More Drinks*, WALL ST. J., Sept. 14, 2010, at B6.

266 *Id.*

267 *Id.*

268 Harris et al., *supra* note 259, at 245.

269 *Id. See also* INST. OF MED., *supra* note 257, at 58–61 (examining relevant information about socioeconomic and ethnic make up of high-risk groups for childhood obesity).

ently, however, when this practice is directed toward youth. While there is no federal statute that regulates digital marketing to adults, the Children's Online Privacy Protection Act (COPPA) establishes certain rules for marketing to young children.[270] COPPA seeks to protect children under the age of thirteen,[271] and it mandates that a covered website have a posted privacy policy and obtain parental consent before collecting, using, and disclosing children's information.[272] The statute also grants the FTC an enforcement role for its mandates.[273] The FTC has responded vigorously with sixteen enforcement actions, and over $6 million levied in fines pursuant to settlements.[274] Its most recent COPPA enforcement action, in May of 2011, led to a settlement that included a $3 million fine against an operator of online "virtual worlds."[275]

COPPA has several notable weaknesses. First, it applies only to children under thirteen.[276] Advertisers and marketers can therefore ply their trade with teenagers without being affected by the statute. Yet, teenagers may be even more vulnerable to targeted marketing than are younger children.[277] Second, COPPA extends only to a "website or online service," and thus does not regulate new digital plat-

---

270 Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506 (2006).

271 *Id.* § 6501(1).

272 *Id.* § 6502(b)(1)(A)(ii).

273 *Id.* § 6501(8)(F).

274 United States v. Playdom, Inc., No. 11-00724 (C.D. Cal. May 24, 2011); United States v. Iconix Brand Grp., Inc., No. 09-8864 (S.D.N.Y. Nov. 5, 2009); United States v. Sony BMG Music Entm't, No. 08-10730 (S.D.N.Y. Dec. 15, 2008); United States v. Industrious Kid, Inc., No. 08-0639 (N.D. Cal. Mar. 6, 2008); United States v. Xanga.com, Inc., No. 06-6853 (S.D.N.Y. Sept. 12, 2006); United States v. UMG Recordings, Inc., No. 04-1050 (C.D. Cal. Feb. 25, 2004); United States v. Mrs. Fields Famous Brands, Inc., No. 2:03-205 (D. Utah Apr. 4, 2003); United States v. Hershey Foods Corp., No. 4: 03-350 (M.D. Pa. Mar. 6, 2003); United States v. Ohio Art Co., F.T.C. File No. 022-3028 (N.D. Ohio Apr. 22, 2002); United States v. Am. Pop Corn Co., No. 02-4008 (N.D. Iowa Feb. 28, 2002); United States v. Lisa Frank, Inc., No. 01-1516 (E.D. Va. Oct. 3, 2001); United States v. Monarch Serv., Inc., No. 01-1165 (D. Md. Apr. 20, 2001); United States v. Looksmart, Ltd., No. 01-606 (E.D. Va. Apr. 19, 2001); United States v. Bigmailbox.com, Inc., No. 01-605 (E.D. Va. Apr. 19, 2001); FTC v. Toysmart.com, LLC, No. 00-11341 (D. Mass. Sept. 29, 2000); Bonzi Software, Inc., F.T.C. Docket No. C-4126 (Oct. 7, 2004). For a concise discussion of COPPA enforcement actions, see SOLOVE & SCHWARTZ, *supra* note 4, at 110–11.

275 Consent Decree and Order for Civil Penalties, Injunction and Other Relief ¶ 19, United States v. Playdom, Inc., No. 11-00724 (C.D. Cal. May 24, 2011) (levying $3 million fine); Complaint for Civil Penalties, Injunction and Other Relief ¶ 11, United States v. Playdom, Inc., No. 11-00724 (C.D. Cal. May 11, 2011) (describing the defendant's online activities).

276 15 U.S.C. § 6501(1).

277 *See* Harris et al., *supra* note 259, at 236–37 (indicating that social science has shown that the impact of "[m]edia, including marketing messages" is especially strong for "older children and adolescents . . . as they focus more on the world beyond their families and actively develop their independent identities").

forms that are independent of the Internet, such as cell phones.[278] Third, COPPA only regulates a website or online service if it is "directed to children," or if the operator of the website "has actual knowledge that it is collecting personal information from a child."[279] It is relatively easy for website operators to avoid acquiring actual knowledge that they are collecting information from a child.[280]

Even if these problems with COPPA are addressed, the statute still suffers from a fundamental flaw: its concept of PII. COPPA defines PII through the specific-types paradigm,[281] but it employs this approach with a twist. In addition to setting out a traditional list of types of PII (first and last name, social security number, e-mail address, and other elements), it provides an authorization for the FTC to add additional factors to that list.[282] Although the FTC made limited use of this power in its COPPA Rule in 2000, adding "a persistent identifier" used to track a person to the list,[283] the statute's key concept remains whether or not the "identifier" will permit "the physical or online contacting of a specific individual."[284] The meaning of "contacting of a specific individual" remains unresolved. Marketers will argue, and the FTC is likely to agree, that it is not a "contacting of a specific individual" when targeted ads are served to children. Support for this proposition is found in the FTC's definition of "online contact information" as "an e-mail address or any other substantially similar identifier that permits direct contact with a person online."[285]

---

278 15 U.S.C. § 6501(2)(A). Indeed, the FTC in 2007 had already noted that children's access to the Internet was increasingly taking place on mobile devices rather than personal computers. FTC, IMPLEMENTING THE CHILDREN'S ONLINE PRIVACY PROTECTION ACT 2 (2007) [hereinafter FTC, COPPA REPORT]. In this report, the FTC also identified challenges to COPPA in social- networking sites and the convergence of wireless and landline communications with the Internet. *Id.* at 25–27.

279 15 U.S.C. § 6502(b)(1)(A) (2006). General-audience websites that have a special section for children are also subject to COPPA, as the statute explicitly applies to "that portion of a commercial website or online service that is targeted to children." *Id.* § 6501(10)(A).

280 As a result of this requirement, many websites that might otherwise fall under COPPA have a simple way of avoiding its reach: using drop-down age menus, which require a user to indicate their age before being allowed access to the site. Of course, it is not especially difficult for children to determine the appropriate birthday that will allow them to access a website.

281 *See supra* Part I.B.3 (describing the specific-types paradigm).

282 15 U.S.C. § 6501(8)(A)–(F).

283 16 C.F.R. § 312.2 (2011).

284 15 U.S.C. § 6501(8)(F).

285 16 C.F.R. § 312.2.

IV
PII 2.0

The existing definitions of PII have proven problematic. Nonetheless, we reject the idea that privacy law should abandon the concept of PII. If it did so, privacy law would be left without a means for establishing coherent boundaries on necessary regulation. Therefore, we reconceptualize the current standard by introducing PII 2.0, compare the new model to existing approaches in the United States and the European Union, and defend this new approach against possible objections. Finally, we apply this proposal to behavioral marketing to adults and targeted food marketing to children.

A. Should Privacy Law Abandon the Concept of PII?

The PII problem appears daunting, and a dramatic solution would be to abandon PII as a central concept in information privacy law. Indeed, Paul Ohm argues that the concept of PII is unworkable and unfixable. According to Ohm, "No matter how effectively regulators follow the latest re-identification research, folding newly identified data fields into new laws and regulations, researchers will always find more data field types they have not yet covered. The list of potential PII categories will never stop growing until it includes everything."[286] In Ohm's analogy, the attempt to define PII is as futile as the classic carnival game of "whack-a-mole." As he explains it, "As soon as you whack one mole, another will pop right up."[287]

In fairness to Ohm, his primary focus is not on abandoning PII, but on alerting the legal academy and policy makers to the problem of new means for re-identification of data.[288] This effort is a valuable and meritorious one. Nevertheless, his argument goes further when it suggests that privacy law be reoriented around a different concept than PII. In place of PII, Ohm proposes that regulators seek to "prevent privacy harm by squeezing and reducing the flow of information in society, even though in doing so they may need to sacrifice, at least a little, important counter values like innovation, free speech, and security."[289] He would thus replace the current reliance on PII as a gatekeeper for privacy law with a cost-benefit analysis for *all* data

286 Ohm, *supra* note 6, at 1742.

287 *Id.*

288 *See id.* at 1703–04 ("This Article is the first to comprehensively incorporate an important new subspecialty of computer science, reidentification science, into legal scholarship.").

289 *Id.* at 1706.

processing and data collection of any kind.[290] Ohm proposes that privacy regulation "should weigh the benefits of unfettered information flow against the costs of privacy harms."[291]

Abandoning PII is problematic, however, because the concept serves a crucial function: it establishes the boundaries of privacy regulation. Without some concept of PII, there would be no limits on the scope of privacy law. In a world overflowing with information, the law cannot possibly regulate all of it. Yet, without adequate boundaries on regulation, privacy rights would expand to protect a nearly infinite array of information, including practically every piece of statistical or demographic data. The law would encompass nearly every fact about human behavior, no matter how generalized. Moreover, Ohm's proposal to assess the costs and benefits of every collection and release of data would be tremendously difficult because all costs and benefits are rarely known in advance. Ohm suggests that when in doubt, the law should limit the release or even the creation of large data sets.[292] Such data sets, however, play an important role in research, health care, data security, and the dissemination of knowledge generally.

In health care research, an important distinction is now drawn between clinical trials, the traditional form of health care research, and new "information based" forms of inquiry.[293] In clinical trials, patients volunteer or are paid to participate in specific studies that test new medical interventions. In contrast, in information-based research, there is "analysis of data and biological samples that were initially collected for diagnostic, treatment, or billing purposes, or that were collected as part of other research projects."[294] The Institute of Medicine has noted that such information-driven research has "led to significant discoveries, the development of new therapies, and a remarkable improvement in health care and public health."[295]

These benefits have included the development of the cancer treatment therapy Herceptin, the use of an "open source" approach to Alzheimer's research, and increased medical database research on children's health. As an initial example, through analysis of the records of a cohort of 9000 breast cancer patients, scientists were able

---

290 *See id.* at 1768 (proposing that regulators restrict data processing and collection when the costs to privacy outweigh the benefits of the information).

291 *Id.* at 1759.

292 *See id.* at 1766–68 (suggesting that the size of data sets should be a regulatory consideration, and that regulation should restrict the creation and release of data sets).

293 *See* INST. OF MED., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 112 (Sharyl J. Nass et al. eds., 2009) ("[A]n increasingly large portion of health research is now information based.").

294 *Id.*

295 *Id.* at 113.

to identify the HER-2 oncogene and develop a targeted therapy for treating women with breast cancer who fall into this genetic category.[296] In another major research effort, one that started in 2003, universities, the drug and medical-imaging industries, and nonprofit groups joined in a collaborative effort to find biological markers that reveal the progression of Alzheimer's disease in the human brain.[297] As the *New York Times* summarized, "[t]he key to the Alzheimer's project was an agreement as ambitious as its goal: . . . to share all the data, making every single finding public immediately, available to anyone with a computer anywhere in the world."[298] There already have been more than 3200 downloads of the entire data set, and almost a million downloads of the database that contains images from brain scans.[299] As a final example, medical database research has improved children's health. Leading examples of this research are the discovery that supplementing folic acid during pregnancy can prevent neural tube birth defects, and the identification of the negative effects of intrauterine DES exposure.[300]

Analytics also play an important role in data security. For example, a multi-institutional response is necessary to combat data security breaches.[301] One of the most important requirements of such a response is the sharing of information about security attacks among different entities to minimize harm and to increase the relevant knowledge among private organizations, governmental entities, and the public.[302] Elements of such a coordinated response are beginning to emerge. Companies in the private sector now offer services that draw on information from multiple organizations to spot data anomalies that can identify malicious activities.[303]

---

296 *Id.* at 114.

297 Gina Kolata, *Rare Sharing of Data Led to Results on Alzheimer's*, N.Y. TIMES, Aug. 13, 2010, at A1. The data are posted online. ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE, http://adni.loni.ucla.edu/ (last visited Oct. 31, 2011).

298 Kolata, *supra* note 297, at A1.

299 *Id.*

300 INST. OF MED., *supra* note 293, at 114. In a recent study that used database analysis, Flaura Winston and other researchers drew on "child-focused crash surveillance information" reported to the State Farm Insurance Companies in fifteen states and the District of Columbia and then shared with the Partners for Child Passenger Safety. Flaura K. Winston et al., *The Danger of Premature Graduation to Seat Belts for Young Children*, 105 PEDIATRICS 1179, 1179–80 (2000).

301 *See* Paul M. Schwartz & Edward J. Janger, *Notification of Data Security Breaches*, 105 MICH. L. REV. 913, 918, 959–70 (2007) (arguing that effectively addressing data breaches would require a "multi-institutional, coordinated response" to mitigate harm to consumers and to improve security to prevent future breaches).

302 *Id.* at 962.

303 For example, ID Analytics draws on information about 2.6 million frauds and 1.4 billion consumer transactions in its national, cross-industry compilation of identity infor-

Analytics are also used in creating new products and services for direct use by individuals. For example, Google Flu Trends is a free service that furthers early detection of influenza epidemics throughout the world.[304] Epidemics of seasonal influenza are a major public health issue. They cause between 250,000 and 500,000 deaths worldwide annually as well as tens of millions of cases of respiratory illness.[305] There is also growing concern about the possibility of a future pandemic causing millions of possible fatalities worldwide if a new strain of influenza virus emerges.[306] Scientists at Google and the Centers for Disease Control and Prevention have developed a method of analyzing large numbers of Google search queries to track influenza-like illnesses in different parts of the world. The technique monitors health-seeking behavior, specifically the online search queries that millions of individuals submit to the Google search engine each day.[307]

Although analytics have great benefits, they can also implicate information privacy concerns. Still, an approach in which the first step is to restrict the flow of information is a move in the wrong direction, especially because new technology is increasing the benefits from analysis of large data sets in ways we might not be able to predict in advance. The general approach to information flow in the United States is a "Schillerian" one. As Friedrich Schiller wrote in *Wallensteins Lager*: "*Was nicht verboten ist, ist erlaubt*" ("What is not forbidden is allowed").[308] Applying his insight to privacy law, one might say that all information collection and processing that is not specifically forbidden by law is permitted.[309] This approach wisely encourages the flow of information and the benefits it brings, while building in restrictions where it can cause problems. Shifting to a regime where the full benefits and costs must be weighed in advance might prevent the discovery of new benefits and overly constrain information flow. Moreover, an ex ante cost-benefit analysis would be

---

mation. *Technology: ID Network*, IDANALYTICS.COM, http://www.idanalytics.com/technology/index.php#id-network (last visited Oct. 31, 2011).

304 Jeremy Ginsberg et al., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012, 1014 (2009).

305 *Influenza (Seasonal), Fact Sheet N°211*, WORLD HEALTH ORG. (Apr. 2009), http://www.who.int/mediacentre/factsheets/fs211/en/.

306 Ginsberg et al., *supra* note 304, at 1012.

307 *Id.* at 1014.

308 FRIEDRICH SCHILLER, *Wallensteins Lager*, *in* 4 FRIEDRICH SCHILLER WERKE UND BRIEFE, act 1, sc. 6 (von Otto Dann et al. eds., Deutscher Lassiker Verlag 2000) (1798).

309 *See* Paul M. Schwartz, *Preemption and Privacy*, 118 YALE L.J. 902, 908-16 (2009) (contrasting a European Union approach to information privacy centered on prevention of harm with the United States approach based on "regulatory parsimony," or avoiding unnecessary regulation of information flow).

so speculative in nature that its accuracy and usefulness would be questionable. And, any kind of presumptive rejection of the collection and dissemination of large data sets will constitute a major sacrifice of potential benefits.

Privacy rights should attach when data pertain to particular people. The disclosure that there are nine million people living in New York City does not create a privacy harm for any specific New Yorker. Of course, certain types of aggregate data can be used in harmful ways. For example, banks might draw on a statistical indication that a certain demographic group has a much higher default rate to deny loans to members of this group or to charge them higher rates. In addition, actuarial data from insurance companies affects coverage and rate decisions. These decisions can be harmful, and information does play a role in these harms. Nonetheless, this category of harm is far broader than the category of information privacy harms. As a policy matter, these issues raise questions that predominately sound in civil rights, discrimination, and insurance law.[310] At least as far as the analysis of aggregate data is concerned, the critical issues in these areas are not those of information privacy law.

When a privacy harm is created, it is because the data disclosed or used pertains to specific individuals. Disclosing that 10,000 copies of a particular book were sold does not implicate privacy; this is just a piece of information. Disclosing that a particular book was sold to a particular person does implicate information privacy.[311] The privacy harm, or the potential for it, is created by linking the information to an individual. This result does not mean that the harm following upon a linkage of data to a particular person is exclusively an individual one—indeed, the resulting harm can affect all of society.[312] Thus, the individual capacity for self-determination can be harmed by informa-

---

310 For a discussion of antidiscrimination law, see generally TIMOTHY P. GLYNN ET AL., EMPLOYMENT LAW 515–43 (2007). Regarding civil rights and discrimination based on information about one's neighborhood, Congress enacted the Community Reinvestment Act in 1977 to prevent lenders from discriminatory credit practices against inhabitants of low-income neighborhoods, a practice known as redlining. Community Reinvestment Act, Pub. L. No. 95-128, 91 Stat. 1147 (1977).

311 Whether or not the United States provides enough protections in privacy law for such information is, of course, another matter. For a discussion of the inadequacies of American privacy law, see SOLOVE & SCHWARTZ, IPL, *supra* note 52, at 565–72.

312 For example, the disclosure of a person's membership in an organization can limit the freedom of association of groups seeking to affect social change. *See* Bates v. City of Little Rock, 361 U.S. 516, 527 (1960) (striking down state ordinance requiring the disclosure of NAACP's members and contributors); NAACP v. Alabama, 347 U.S. 449, 461–63 (1958) (finding that the Constitution, by protecting the freedom of association, prevents a state from requiring disclosure of lists of members of lawful associations absent some overriding valid interest of the state).

tion surveillance, which, in turn, will have a negative effect on the maintenance of a democratic order.[313]

B. A Standard for PII

In devising an approach to conceptualize PII, the first step is to determine whether it should be defined as a rule or a standard. As we have noted earlier, a standard is an open-ended decision making yardstick, and a rule is a harder-edged decision making tool.[314] The discussion of PII in the law has not yet considered the issue of whether PII ought to be a rule or a standard, but this issue is of paramount importance and an essential first step.

We can now revisit the current state of play concerning PII. The first model, the tautological approach, ultimately rests on the circular notion that PII is personal information. The second model, the non-public approach, defines PII as that data which is not publicly available. Finally, the third model, the specific-types approach, expressly lists the kinds of data that are PII. As previously discussed, the first two are standards, and the third is a rule. The distinction between rules and standards also provides a window into the current failure of all three approaches.

Standards, as we have seen in the tautological and non-public approaches, permit broad discretion and allow the decision maker to take into account relevant factors.[315] To illustrate, consider the VPPA's definition of PII as "information which identifies a person," or the Cable Act's explanation of this same concept as anything other than "aggregate data."[316] In both statutes, the use of a standard permits freedom in deciding which factors to take into account. The decision maker can identify these factors based on the original policy.[317] The result is a better fit between policy and the facts at hand.

Yet, the two kinds of standards used in information policy law have not led to good decision making in identifying PII. Due to any standard's inevitable generality, the problematic tendency has been to interpret a specific definition of PII as applying only to information

---

313 *See* Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609, 1653–58 (1999) (discussing how privacy harms can undermine a Civic Republican understanding of self-determination from a lack of sufficient protections).

314 *See supra* notes 70–71 and accompanying text.

315 Sullivan, *supra* note 70, at 58–59 ("Standards allow the decisionmaker to take into account all relevant factors or the totality of the circumstances.").

316 Video Privacy Protection Act of 1988, 18 U.S.C. § 2710(a)(3) (2006); Cable Communications Policy Act of 1984, 47 U.S.C. § 551(a)(2)(A) (2006).

317 *See* Sullivan, *supra* note 70, at 58 ("A legal directive is 'standard'-like when it tends to collapse decisionmaking back into the direct application of the background principle or policy to a fact situation.").

that taken in isolation, in that single case, actually identifies a specific individual. We will call this viewpoint the "reductionist reading" of PII. Such an interpretation ignores the dangers of re-identification and other issues that we discussed in Part III. The reductionist tendency pervades U.S. law at present, and we attempt to overcome this problem in our definition of PII. To be sure, a second risk also exists—that too much information could be considered PII. We associate such an "expansionist tendency" with the European Union and respond to it as well in crafting our definition.

As for the third category, the specific-types approach lists certain kinds of data that fall within the category of PII. The resulting attempts at a rule, however, prove either too narrow, as in the Massachusetts breach notification statute,[318] or outdated, as in the COPPA Rule.[319] As Kathleen Sullivan pointed out in 1991, one problem with rules is that they "tend toward obsolescence."[320] Indeed, while COPPA permits the FTC to add to the definition of PII, this authorization has languished unused since 2000.[321] Here, we can draw on the insight of Louis Kaplow, who noted that rules require the legal system to expend more work ex ante, and standards require it to engage in more work ex post.[322] As Kaplow observed, "[w]hen the government promulgates a rule, it gathers information before individuals act and announces its findings."[323] The difficulty for rules is that the government entity designated to revise them may be unable or unwilling to expend the necessary resources to do so.[324] As an illustration, the FTC has been gridlocked around marketing to children and has not changed the COPPA Rule for over a decade.

In sum, we view the current condition of PII, whether defined in terms of either standards or rules, as deeply unsatisfactory. In moving forward, we opt for a reconceptualization of a standard for PII and not a rule. We do so for three reasons.

First, standards are generally the superior choice for dealing with situations of rapid change, because, as mentioned, rules can become

---

318 Massachusetts Security Breach Law, 201 MASS. CODE REGS. 17.00 (2011). For discussion of the deficiencies in the Massachusetts statute's definition of PII, see *supra* notes 92–93 and accompanying text.

319 Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506 (2006). For discussion of the deficiencies in the COPPA definition of PII, see *supra* notes 113–16 and accompanying text.

320 Sullivan, *supra* note 70, at 66.

321 *See* Children's Online Privacy Protection Rule, 16 C.F.R. § 312.1 (2011) (showing last changes from over a decade ago).

322 Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 559–63 (1992).

323 *Id.* at 585.

324 *See id.* at 623.

obsolete.[325] Indeed, rules function best when an area of social and technological development has reached a fairly settled state. Sullivan observed that a rule reflects an area of "epistemological maturity."[326] The many routes to the creation of PII do not fit into a set of neat categories, and the technology of tracking and the science of re-identification will continue to develop in ways that legal decision makers are unlikely to anticipate.

A second ground to prefer defining PII as a standard is the heterogeneous nature of the behavior to be regulated. As this Article has demonstrated,[327] the means to track individuals and re-identify information are diverse. Numerous scholars, including Isaac Ehrlich and Richard Posner, have demonstrated that rules are quite poor for handling situations involving heterogeneous types of behavior that should be treated differently.[328] Capturing these behaviors in a rule, or a series of rules, is only possible through a highly detailed codification, and such extensive statutory detail often fails to adapt well to technological change.

At the same time, and as a third benefit, a standard for PII has the further merit of being capable of identifying discrete areas in which rules will be more useful for defining data as PII or non-PII. If developments in technology and society around a certain subcategory of data use become settled, it may become possible to formulate harder-edged rules to supplement a broad standard. In this way, there can be a "back-and-forth pattern" between standards and rules over time.[329] These many factors suggest that the best starting point for information privacy law, at least under present conditions, is to conceive of PII as a standard. The question then becomes what the nature of this standard is. In the next section, we consider two existing models: (1) the United States' reductionist approach to PII, and (2) the European Union's expansionist approach.

C. Reductionism, Expansionism, and PII 2.0

Information privacy law is now divided between reductionist and expansionist regulation of PII. The United States and the European Union offer examples of the former and the latter, respectively. Both

---

325 Sullivan, *supra* note 70, at 66.

326 *Id.* at 62.

327 *See supra* Part III.

328 *See, e.g.*, Isaac Ehrlich & Richard A. Posner, *An Economic Analysis of Legal Rulemaking*, 3 J. LEG. STUD. 257, 268 (1974) (describing "the necessarily imperfect fit between the coverage of a rule and the conduct sought to be regulated").

329 Rose, *supra* note 70, at 580.

approaches are flawed. In this section we develop a different concept, which we term "PII 2.0."

1. Reductionism in the United States

In the United States, as we have seen, the law often engages in a reductionist reading of PII. This tendency manifests itself when statutes, judges, or policy makers consider PII to be only information that refers to a currently identified person. Although computer scientists and data security experts in the United States recognize the category of identifiable information, the law has by and large failed to understand this concept. Identified information already refers to a specific person, while identifiability suggests that such a connection has not yet occurred, but is possible. To be sure, the first "I" in the acronym PII is supposed to represent *identifiable*, but most legal definitions of PII only focus on *identified* individuals.

As an example of this interpretation of PII, consider the FTC's view of a "persistent identifier," such as a cookie. As we have argued above, evidence suggests that the FTC views the applicable statutory language and its own COPPA rule as regulating this technology only when there is information about an "identified" person.[330] An activity that falls within the COPPA rule would be a company gathering information about a person and then using it to send her an e-mail.[331] However, when a company engages in the same information gathering, only to send the same person a targeted ad based on cookies placed on her computer, it is likely to fall outside the COPPA rule.

Another example of the reductionist tendency in the United States involves the Privacy Act's definition of a "system of records," which turns on whether federal agency records involve an "identified" person.[332] The Privacy Act does *not* apply to data processing if a person is identifiable within a federal agency's database, but is not located through a unique identifier.

2. Expansionism in the European Union

In comparison, the European Union takes an expansionist approach to PII. For example, the E.U. Data Protection Directive defines "personal data" as "information relating to an identified or

---

[330] *See supra* Part I.B.3.

[331] *See* FTC, COPPA REPORT, *supra* note 278, at 8 (discussing applicability of the COPPA Rule to e-mail communications).

[332] Privacy Act of 1974, 5 U.S.C. § 552a (2006). *See supra* notes 45–47 and accompanying text (describing the approaches of federal laws and the Privacy Act specifically to PII).

identifiable natural person."[333] This accord sets out common rules for data protection in European Union Member States and requires these countries to enact legislation that follows its standards.[334] Through this supranational agreement, a definition of PII extending to identifiable individuals has been fixed deep in the DNA of European Union information privacy law. The E.U. Data Protection Directive defines "an identifiable" person as "one who can be identified, directly, or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural, or social identity."[335] Of some additional definitional assistance, the Directive in its Recital 26 explains that in determining whether a person is identifiable, "account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person."[336]

In the European Union, moreover, information that refers to an identifiable person is treated in the same fashion as that which refers to an identified person. The treatment in privacy of identified and identifiable as equivalents is a German innovation. The German Federal Data Protection Act (*Bundesdatenschutzgesetz*, or BDSG) of 1977 defines "personal data" information as data relating to both "identified" and "identifiable" individuals.[337] Whether the data are identified or identifiable proves, however, to be irrelevant. As Ulrich Dammann writes in the leading treatise on the Federal Data Protection Law statute, there is "personal data" if "the reference person is identifiable."[338] He adds, "[i]t is irrelevant for the BDSG's application whether the person is identified or identifiable."[339]

To be sure, the early concern in European Union law for the risks of identifiable data has proven prescient. In 1978, Dammann had already zeroed in on a threat that he called "re-individualization" (*Re-Individualisierung*) of data.[340] He observed that "[w]here the

---

[333] Council Directive 95/46, on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, art. 2(a), 1995 O.J. (L 281) 31, 38 [hereinafter E.U. Data Protection Directive]. For background on the Directive, see Paul M. Schwartz, *European Data Protection Law and Restrictions on International Data Flows*, 80 IOWA L. REV. 471, 480–83 (1995).

[334] Schwartz, *supra* note 333, at 484.

[335] E.U. Data Protection Directive, *supra* note 333, art. 2(a).

[336] *Id.* at Recital 26.

[337] Bundesdatenschutzgesetz [BDSG] [Federal Data Protection Act], Jan. 14, 2003, BGBL. I at 66, last amended Aug. 14, 2009, BGBL. I at 2814 (Ger.).

[338] ULRICH DAMMANN, KOMMENTAR ZUM BUNDESDATENSCHUTZGESETZ § 3, marginal no. 22 (Spiros Simitis ed., 6th ed. 2006).

[339] *Id.* § 3, marginal no. 23.

[340] ULRICH DAMMANN, KOMMENTAR ZUM BUNDESDATENSCHUTZGESETZ § 2, marginal no. 25 (Spiros Simitis ed., 1st ed. 1978).

layperson sees only statistical tables, the mathematician, thanks to sophisticated 'snooping technologies,' can pry columns of individual data sets out of the computer and frequently within a short time."[341] According to Dammann, the critical question concerning the nature of PII turns on the availability of "additional knowledge" (*Zusatzwissen*) about the concerned individual.[342]

The European Union's expansionist approach to PII is more in tune with technology than is the United States' reductionist approach. It also has exercised significant international influence. In 1980, the Privacy Guidelines of the Organization for Economic Cooperation and Development (OECD) followed the recently enacted first federal data protection law of Germany.[343] These guidelines define personal data as "any information relating to an identified or identifiable individual (data subject)."[344] The OECD Guidelines apply eight privacy principles to all PII, and, in doing so, demonstrate the European Union's expansionist approach.[345] Once there is PII, the OECD principles are to be applied in full force. Like the OECD Guidelines, the Privacy Framework of the Asia-Pacific Economic Cooperation of 2004 defines PII as "any information about an identified or identifiable individual."[346]

Finally, Canadian privacy law reflects the influence of the European Union's approach, but goes even further in dropping the concept of "identified" in its approach to PII. Canada's federal privacy law, the Personal Information Protection and Electronic Documents Act (PIPEDA), regulates the collection, use, and transfer of personal information by private organizations.[347] Enacted in 2000, PIPEDA defines PII simply as "identifiable" information with the limited exceptions of "the name, title[,] or business address or telephone number of an employee of an organization."[348] As a leading treatise

---

341 *Id.*

342 *Id.* § 2, marginal no. 26.

343 *See* Organization for Economic Cooperation and Development, *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, OECD Doc. C(80)58 Final (Sep. 23, 1980) [hereinafter OECD Guidelines]. The OECD is a group of leading industrial countries, including the United States, and the OECD Guidelines provide a non-binding framework for member nations. For a discussion of the OECD Guidelines, see SOLOVE & SCHWARTZ, IPL, *supra* note 52, at 997–98.

344 OECD Guidelines, *supra* note 343, § 1(b).

345 *Id.* §§ 7–14.

346 Asia-Pacific Economic Cooperation, *APEC Privacy Framework*, at ii-9, APEC Doc. 205.SO-01.2 (2005).

347 Personal Information Protection and Electronic Documents Act, S.C. 2011, c. 5 § 3 (Can.) [hereinafter PIPEDA]. PIPEDA also regulates the use of personal information by federal organizations and data flows between Canadian provinces. *Id.* §§ 23(1)–23(3).

348 *Id.* § 2(1).

on Canadian privacy law summarizes the result, "[i]n essence, almost any information in any form that can be attributed to an identified individual is caught by this expansive definition."[349] The federal Privacy Commissioner plays a key role in deciding whether information is identifiable, and the general tendency has been expansionist. As the Privacy Commissioner stated in his annual report to the 2001–2002 Parliament, "[t]he definition is deliberately broad, and in my findings I have tended to interpret it as broadly as possible. . . . I am inclined to regard information as personal even if there is the smallest potential for it to be about an identifiable individual."[350]

Notwithstanding its widespread adoption in other international documents, the European Union's expansionist approach is flawed because it treats data about identifiable and identified persons as conceptually equivalent. The difficulty is that there is a broad continuum of identifiable information that includes different kinds of anonymous or pseudonymous information. Different levels of effort will be required to identify information, and varying risks are associated with the possible identification of data. To place all such data into the same conceptual category as data that currently relate to an identified person is a blunt approach.

More specifically, this approach would lead to a hard trigger for information privacy law. Consider merely two elements of the basic toolkit of Fair Information Practices (FIPs):[351] (1) notice, access, and correction rights for the individual, and (2) security for personal data.[352] For information that merely relates to an *identifiable* person, the law should *not* generally require that the entity that processes

---

[349] BARBARA MCISAAK, Q.C. ET AL., THE LAW OF PRIVACY IN CANADA 4–7 (2011). *See* 1 PRIVACY LAW IN THE PRIVATE SECTOR: AN ANNOTATION OF THE LEGISLATION IN CANADA PIP-15 (Jeffrey A. Kaufman ed., 2007) ("It is, therefore, important to note at the outset that the definition of 'personal information' [in PIPEDA] is extremely broad."); STEPHANIE PERRIN ET AL., THE PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS ACT: AN ANNOTATED GUIDE 54 (2001) ("The definition in the Act is limitless in terms of what can be information about an identifiable individual.").

[350] GEORGE RADWANSKI, OFFICE OF THE PRIVACY COMMISSIONER OF CANADA, ANNUAL REPORT TO PARLIAMENT 2001–2002, at 56 (2003). The term "identifiable" in PIPEDA has been interpreted in case law. *See* Gordon v. Canada (Minister of Health), 2008 F.C. 258, para. 43 (Can.) ("[D]isclosure of the province field [of a form] would substantially increase the possibility that information about an indentifiable individual . . . would fall into the hands of persons seeking . . . to identify 'particular' individuals."); Rousseau v. Wyndowe, 2008 F.C.A. 39, para. 44–45 (Can.) (holding that notes taken by a doctor in the course of filling out a form during an independent medical examination are "personal medical information," a "subset of personal information").

[351] *See supra* note 52 and accompanying text (defining and describing the origins of FIPs).

[352] *See* FED. TRADE COMM'N, PRIVACY ONLINE: A REPORT TO CONGRESS 7–11 (providing definitions and discussions of these two core principles of information privacy law).

information provide a full panoply of notice, access, and correction rights. Indeed, to do so might be counterproductive in many circumstances. It would require the data processing entity to first associate merely identifiable data with an identified person, which would thereby render it identified data.

3. The Benefits of PII 2.0

A benefit of having two different categories of PII, identified and identifiable data, is to permit an assessment of the optimal nature of legal protections. Rather than a hard "on-off" switch, this approach allows for legal safeguards for both identified and identifiable information—safeguards that permit tailored FIPs built around varying levels of risk to individuals. Our model places information on a continuum that begins with no risk of identification at one end, and ends with identified individuals at the other. We divide this spectrum into three categories, each with its own regulatory regime: under the PII 2.0 model, information can be about an (1) identified, (2) identifiable, or (3) non-identifiable person. Our three categories divide up this spectrum and provide different regimes of regulation for each. Because these categories do not have hard boundaries, we define them in terms of standards.

Information refers to an *identified* person when it singles out a specific individual from others. Put differently, a person has been identified when her identity is ascertained. There is general international agreement about the content of this category, albeit not of the implications of being placed in it. For example, in the United States, the General Accounting Office, Office of Management and Budget, and National Institute of Standards and Technology associate this concept with information that distinguishes or traces a specific individual's identity.[353] In Europe, the Article 29 Group states that a person is identified "when, within a group of persons, he or she is 'distinguished' from all other members of the group."[354] In German data protection law, as Dammann explains, "[t]he person is identified

---

[353] ERIKA MCCALLISTER ET AL., NAT'L INST. OF STANDARDS AND TECH., GUIDE TO PROTECTING THE CONFIDENTIALITY OF PERSONALLY IDENTIFIABLE INFORMATION (PII) 2-1 (2010); U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-536, PRIVACY: ALTERNATIVES EXIST FOR ENHANCING PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION 1 n.1 (2010); CLAY JOHNSON III, OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, M-07-16, SAFEGUARDING AGAINST AND RESPONDING TO THE BREACH OF PERSONALLY IDENTIFIABLE INFORMATION 1 n.1 (2007).

[354] Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concept of Personal Data* at 12, 01248/07/EN/WP 136 (June 20, 2007).

when it is clear that the information refers to this person and not to another."[355]

Information in the middle of the risk continuum relates to an *identifiable* individual when specific identification, while possible, is not a significantly probable event. In other words, an individual is identifiable when there is some non-remote possibility of future identification. The risk level for such information is low to moderate. Information of this sort should be treated differently from an important subcategory of nominally identifiable information, in which linkage to a specific person has not yet been made, but where such a connection is more likely. As we shall explain, such nominally identifiable data should be treated the same as identified data.

At the other end of the risk continuum, *non-identifiable* information carries only a remote risk of identification. Such data cannot be said to be relatable to a person, taking account of the means reasonably likely to be used for identification. In certain kinds of data sets, for example, the original sample is so large that other information will not enable the identification of individuals. An example would be high-level information about the populations of the United States, China, and Japan, and their relative access to telecommunications.[356]

There are certain instances where identifiable information should be treated like information referring to an identified person. Information that brings a substantial risk of identification of an individual should also be treated as referring to an identified person. In other words, identifiable data should be shifted to the *identified* category when there is a significant probability that a party will make the linkage or linkages necessary to identify a person. This essential subcategory requires assessment of the means likely to be used by parties with current or probable access to the information, as well as the additional data upon which they can draw. This test, like those for the other categories, is a contextual one. It should consider factors such as the lifetime for which information is to be stored, the likelihood of future development of relevant technology, and parties' incentives to link identifiable data to a specific person.[357]

---

355 DAMMANN, *supra* note 338, § 2.3, ¶ 21.

356 The CIA's World Factbook provides online access to such information. CENT. INTELLIGENCE AGENCY, THE WORLD FACTBOOK, https://www.cia.gov/library/publications/the-world-factbook (last visited Oct. 31, 2011).

357 *See* Article 29 Data Protection Working Party, *supra* note 354, at 15 (discussing the need to take into account "all the means likely . . . to be used" by a data processor "to single out" the individual and identify them (quoting Council Directive 95/46, ¶ 26, 1995 O.J. (L 281) 31 (EC))).

Practical tools also exist for assessing the risk of identification. In fact, computer scientists have developed metrics for assessing the risk of identifiability of information. For example, Khaled El Emam has identified benchmarks for assessing the likelihood that de-identified information can be linked to a specific person—that is, can be made identifiable.[358] The critical axes in El Emam's work concern the "mitigating controls" available to parties in possession of information, and the likely motives and capacity of outsiders who might seek to tie that information to a person.[359] In addition, computer scientists' ongoing work in developing more secure software offers useful lessons. The relevant need is to focus on: (1) the nature of internal and external threats to a data asset, and (2) the effectiveness of possible countermeasures to those threats.[360]

In our next section, we will discuss how FIPs apply to the three categories of PII 2.0 and how this model will encourage companies to keep information in the least identifiable form possible. We then deal with possible objections to PII 2.0 and conclude by applying our model to behavioral marketing and digital marketing to children.

D. PII 2.0 and Fair Information Practices (FIPs)

In our reconceptualized notion of PII, the key is to think about identification in terms of risk level. PII 2.0 conceives of identifiability as a continuum of risk rather than as a simple dichotomy. A clear way to demonstrate the functioning of this new approach is by considering the applicability of FIPs.

The basic toolkit of FIPs includes the following:

---

358 *See* Khaled El Emam, *Heuristics for De-Identifying Data*, SECURITY & PRIVACY, July/Aug. 2008, at 58; Khaled El Emam, *Risk-Based De-Identification of Health Data*, SECURITY & PRIVACY, May/June 2010, at 64 [hereinafter El Emam, *Risk-Based De-Identification*].

359 El Emam, *Risk-Based De-Identification*, *supra* note 358, at 66. *See also* ANN CAVOUKIAN & KHALED EL EMAM, DISPELLING THE MYTHS SURROUNDING DE-IDENTIFICATION: ANONYMIZATION REMAINS A STRONG TOOL FOR PRIVACY 13 (2011) ("There are many tools available that mitigate the risk of re-identification of de-identified information."); Cynthia Dwork, *A Firm Foundation for Privacy Data Analysis*, 54 COMMC'NS OF THE ACM 86, 91–94 (2011) (developing the concept of "differential privacy," a privacy guarantee that "revolves around hiding the presence or absence of a[ny] single individual" or small group, thereby equalizing privacy risks for those included and those not included in a database).

360 *See* MICHAEL HOWARD & STEVE LIPNER, THE SECURITY DEVELOPMENT LIFECYCLE (2006) (discussing techniques for engineers to develop more secure software). Moreover, Adam Shostack and Andrew Stewart have proposed that data security experts should analyze objective information about data breaches, draw on other fields, such as economics and psychology, and use the scientific method in testing hypotheses. ADAM SHOSTACK & ANDREW STEWART, THE NEW SCHOOL OF INFORMATION SECURITY 144–52 (2008).

(1) limits on information use; (2) limits on data collection, also termed data minimization; (3) limits on disclosure of personal information; (4) collection and use only of information that is accurate, relevant, and up-to-date (data quality principle); (5) notice, access, and correction rights for the individual; (6) the creation of processing systems that the concerned individual can know about and understand (transparent processing systems); and (7) security for personal data.[361]

When information refers to an *identified* person, all of the FIPs generally should apply.

To be sure, no single information privacy statute contains all these principles in the same fashion or form. The precise content of the resulting obligations will often differ based on the context of data processing, the nature of the information collected, and the specific legislative, regulatory, and organizational environment in which the rules are formulated.[362] Nonetheless, the basic idea is that all of the FIPs should generally be available once a party processes information that singles out a specific individual.

As for the category of *identifiable*, it is not appropriate to treat such information as fully equivalent to identified. The information does not yet refer to a specific person and may never do so. Nonetheless, some protections are in order because there is a risk of linkage to a specific individual. The question then becomes, which of the FIPs should apply?

Full notice, access, and correction rights should *not* be granted to an affected individual simply because identifiable data about her are processed. For one thing, if the law created such interests, these obligations would decrease rather than increase privacy by requiring that all such data be associated with a specific person. This connection would be necessary in order to allow that individual to exercise her rights of notice, access, and correction. In this fashion, the law would create a vicious circle that could transform identifiable data into identified data. Moreover, limits on information use, data minimalization, and restrictions on information disclosure should not be applied across the board to identifiable information. Such limits would be disproportionate to risks from data use and also would cripple socially productive uses of analytics that do not raise significant risks of individual privacy harms.[363]

---

[361] Schwartz, *supra* note 52, at 907.

[362] On the development of privacy legislation in the United States, the classic study remains REGAN, *supra* note 52, at 174–211.

[363] At the Article 29 Working Party of the European Union, there recently has been openness to a concept of proportionality in the use of information privacy law. *See* Article 29 Data Protection Working Party, *Opinion 3/2010 on the Principle of Accountability* at 3,

At the same time, some FIPs should apply to identifiable data. The key FIPs are those that concern data security, transparency, and data quality. Data security refers to the obligation to "protect against unauthorized access to and use, destruction, modification, or disclosure of personal information."[364] Identifiable information should be subject to data security principles. Recall that identifiable data are those for which a specific identification, while possible, is not a significantly probable event. Yet these data, unlike non-identifiable information, might be relatable to a person. Data security for identifiable information, as for identified information, should be commensurate with the nature of the information and the likely risks of disclosure.

As for transparency, this FIP calls for the creation of data processing systems that are open and understandable to affected individuals. Transparency also means that tracking or surveillance should not be done secretly. This FIP is important for identifiable data for two reasons. First, openness about information use allows for improved policies and law. As Louis Brandeis famously stated, "Sunlight is said to be the best of disinfectants; electric light the most efficient policeman."[365] Brandeis was also concerned about privacy, as reflected first in his famous 1890 article with Samuel Warren, and later in his opinions as a Supreme Court Justice.[366] Yet, Brandeis's attention to privacy for individuals was accompanied by his interest in open flows of information about "social and industrial diseases."[367] Characteristic is his argument about the need for "publicity as a remedy" for abusive practices among financial institutions and bankers in the early twentieth century.[368] In an analogous fashion, behavioral marketing and food marketing to children are controversial today, and there is a need for transparency about these emerging practices.[369]

00062/10/EN/WP 173 (July 13, 2010). The question remains as to how successful this concept will be in a system that treats identified and identifiable data as equivalents.

364 SOTTO, DESKBOOK, *supra* note 226, at 14-3.

365 LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 92 (1914).

366 Warren & Brandeis, *supra* note 13, at 193. The most famous of his opinions about privacy as a Supreme Court Justice is his dissent in *Olmstead v. United States*, 277 U.S. 438 (1928), in which he discusses the need for an expansive principle of privacy that adapts to technological innovation.

367 BRANDEIS, *supra* note 365, at 92.

368 *Id.* at 101.

369 For more on Brandeis as a progressive advocate and his belief in public advocacy and in shaping opinion, see generally Neil M. Richards, *The Puzzle of Brandeis: Privacy and Speech*, 63 VAND. L. REV. 1295 (2010), which describes Brandeis's commitment to the idea that the public disclosure of wrongdoing is in the public service.

Second, identifiable information can have great value. As we have discussed, "stock-market-like exchanges" now exist around information that is collected online.[370] Some of this information may fall into our category of identifiable data for which there is a substantial risk of identification of a specific individual. Other data may be merely identifiable. Transparency about the collection of identifiable information will serve to heighten awareness about data flows among all parties, both consumers and corporations. It will also improve the position of consumers who have preferences about the collection and further use of their data.

Finally, data quality is a FIP that requires organizations to engage in good practices of information handling. This requirement depends on the purpose for which information is to be processed. In the context of *identified* data, for example, it means that the greater the potential harm to individuals, the more precise that the data and its processing must be. Some decisions matter more than others, however, and the stakes are low when the issue is whether or not one receives a coupon for a dollar discount on a case of seltzer. More precision is required in a data system that decides whether or not one receives a mortgage, and determines the interest rate associated with it. In contexts where the decision to be made about a person based on identified data is more important, or the harm to the person potentially greater, there must be higher requirements for data quality.

In the context of *identifiable* information, data quality also requires good practices of information handling. In particular, it requires that companies pay attention to the handling of identifiable information by third parties. If information is non-identifiable, a company can publicly release it or permit third parties access to it without further obligations. We have used the example of comparative telecommunications statistics for the U.S., China, and Japan. Another example of non-identifiable information would be the information presented in Google Flu Trends. As we have noted,[371] Google Flu Trends furthers early detection of influenza epidemics throughout the world by monitoring health-seeking behavior, specifically the online web search queries that millions of individuals submit to the Google search engine each day.[372] When one clicks on Google Flu Trends, there is only high-level information that is safely aggregated.

---

370 *See* Angwin, *supra* note 190 (discussing the market for consumer profiles of Internet users); *supra* notes 198–201 and accompanying text (discussing the billon-dollar industry of behavioral marketing as practiced by popular websites).

371 *See supra* notes 304–07 and accompanying text.

372 *See* Ginsberg, *supra* note 304, at 1014.

Identifiable information is capable of identification, even if this risk is not significantly probable. Thus, companies cannot merely release it or allow unmonitored access to it. Depending on the potential harm to individuals and the likely threat model, companies should also be required to use a "track and audit" model for some identifiable information. An example would be information used in health care research.[373] Access to such data should be accompanied by obligations that travel with the information. Companies that handle identifiable information can structure these obligations by associating metadata, or information about information, with data sets.[374]

Thus, one benefit of PII 2.0 is that it tailors FIPs to whether information is identified or identifiable. A further benefit of PII 2.0 is that it creates an incentive for companies to keep information in the least identifiable form. If we abandon PII, or treat identified and identifiable information as equivalents, companies will be less willing to expend resources to keep data in the most de-identifiable state practicable. As an illustration of such a disincentive in action, the European Union's Article 29 Group, an independent advisory body on privacy, has articulated an "absolute certainty" test for ISPs and search engine operators.[375] Under that test, unless a company in this category can demonstrate "with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all . . . information as personal data, to be on the safe side."[376] The absolute certainty test is not linked to a sense of proportionality regarding the risks associated with re-identification of seemingly non-identifiable information, or of linking identifiable information to a specific person. In contrast, PII 2.0 is more likely to motivate a company to invest resources in maintaining information in either identifiable or non-identifiable form. Companies can thereby benefit from FIPs that become easier to comply with as they move along this continuum *away* from identified information.

E. Possible Objections

What then are the possible objections to PII 2.0? From Ohm's perspective, the difficulty might be that the concept of PII is doomed

---

373 For a similar risk-based assessment of the threat of re-identification, see CAVOUKIAN & EL EMAM, *supra* note 359, at 13.

374 *See* Schwartz, *supra* note 214, at 2077 (calling for the association of propertized personal information with non-personal metadata).

375 Article 29 Data Protection Working Party, *supra* note 354, at 17; Article 29 Data Protection Working Party, *Opinion 1/2008 on Data Protection Issues Related to Search Engines* at 8, 00737/EN/WP 148 (Apr. 4, 2008) [hereinafter Article 29 Data Protection Working Party, *Opinion 1/2008*].

376 Article 29 Data Protection Working Party, *Opinion 1/2008*, *supra* note 375, at 8.

because the risk of identification can never be eliminated. From the European Union's perspective, the problem is that treating identifiable information as subject to a different level of protection than identified might open a back door for significant privacy violations. We deal with each set of objections in turn and contrast them with PII 2.0.

As we have noted, Ohm views an attempt to define PII as being as useless as expecting a successful outcome to the game of whack-a-mole.[377] Potential PII is everywhere, and attempts to predict where it will appear, or in his metaphor, "pop right up," are pointless.[378] In our view, however, computer science is developing metrics that are suitable for just this task. Where Ohm sees only chaos and whack-a-mole, we think that a standard-based approach can be made operational and predictable. It certainly will be as workable as the law's recourse to standards in other areas, such as the concept of "reasonable" behavior in negligence law, or that of "access or acquisition of information" in data breach notification law.

In tort law, the concept of reasonableness functions as a way for juries to sift through an otherwise unordered universe of facts that are potentially relevant each time an accident occurs.[379] Only "unreasonable" behavior can be said to be negligent, and a jury uses this standard, in focusing on various circumstances, as well as its shared sense of the kinds of behavior that each person owes another.[380] As a further matter, tort law also demonstrates the capacity of standards to generate rules in areas where consensus has emerged regarding the required behavior. For example, insurance companies resolve responsibility for most garden variety traffic accidents through use of bright-line rules. In an analogous fashion, standards regarding the risks associated with different processing of identifiable data may prove capable of generating rules for more settled areas.

To shift from the common law to modern statutory regulation of a high tech issue, we can consider data breach notification laws, which have been enacted in forty-five states.[381] These statutes typically require that an individual be notified when there is evidence that an

---

377 *See supra* notes 286–87 and accompanying text.

378 Ohm, *supra* note 6, at 1742.

379 On the role of the jury in tort law as an institution for sifting and selecting the facts that matter, see LAWRENCE ROSEN, LAW AS CULTURE 68–130 (2006).

380 For an introduction to the variable factors in concepts of reasonable and unreasonable behavior in negligence determinations, see RICHARD A. EPSTEIN, CASES AND MATERIALS ON TORTS 169–285 (9th ed. 2008).

381 SOLOVE & SCHWARTZ, *supra* note 4, at 136–38. For leading data breach notification laws, see CAL. CIV. CODE § 1798.82 (West 2006) and 201 MASS. CODE REGS. 17.00–05 (2010).

outside party has gained access to or acquired personal data.[382] These laws do not require a showing that a third party actually acquired or gained control of the information.[383] This standard has led to the development of contextual benchmarks regarding relevant indices of "access or acquisition" of information.[384] No more is required for PII 2.0; here too, there is a need to develop norms that permit a tailored response to a wide range of situations.

If "whack-a-mole" is ultimately not a convincing objection, Ohm does develop a more successful critique of the technique that he terms "release-and-forget." He writes, "As the name suggests, when a data administrator practices these techniques, she releases records—either publicly, privately to a third party, or internally within her own organization—and then she forgets, meaning she makes no attempt to track what happens to the records after release."[385] Ohm points out that "release-and-forget" can be a prescription for privacy disaster. In our view, it can be especially problematic in the absence of risk assessment. Nonetheless, and unlike Ohm, we do not think that the current arms race necessarily favors re-identification. Computer scientists continue to seek to develop new and seemingly promising methods of anonymizing data sets for research purposes.[386] At the same time, we also think that the sheer rate of technological change in this area counsels introduction of a "track-and- audit" approach, as set out above.

The European Union's objection to PII 2.0 would be that it will open a back door to privacy violations. In the words of the Article 29 Working Party, the goal must be to avoid "unduly restricting the interpretation of the concept of personal data."[387] The fear is that any other definition would narrow the jurisdictional sweep of the law. The Article 29 Working Party has also conceded that "[t]he scope of the

---

[382] *See* SOLOVE & SCHWARTZ, *supra* note 4, at 136 (providing examples of various states' data breach notification statutes).

[383] *Id.*

[384] As an example of such benchmarks, see CAL. OFFICE OF PRIVACY PROTECTION, RECOMMENDED PRACTICES ON NOTICE OF SECURITY BREACHES INVOLVING PERSONAL INFORMATION 12–13 (2009).

[385] Ohm, *supra* note 6, at 1711–12. For a technical critique of ad hoc de-identification and its shortcomings, see Dwork, *supra* note 359, at 86–89.

[386] For an example of the different attempts to develop strong statistically-based methods of de-identification, see Arvind Narayanan & Vitaly Shmatikov, *Privacy and Security: Myths and Fallacies of "Personally Identifiable Information*," 53 COMM. ACM 24 (2010). For an argument about how policymakers and legal scholars err by ignoring the likelihood of an actual threat of re-identification of data as opposed to concentrating on "the opportunities and motivations for the hypothetical adversary," see Yakowitz, *supra* note 156, manuscript at 22, 35–37.

[387] Article 29 Data Protection Working Party, *supra* note 354, at 5.

data protection rules should not be overstretched."[388] Nonetheless, it seeks to claim as much information as possible to be "personal data"—as demonstrated, for example, in its "absolute certainty" test.[389] Once information is swept into this category, the Working Party concedes the need for "a substantial degree of flexibility . . . between protection of the data subject's rights on the one side, and on the other side the legitimate interests of data controllers, third parties and the public interest."[390] The evidence is at best mixed regarding whether such flexibility has, in fact, been forthcoming.[391]

In PII 2.0, by contrast, flexibility follows from a general association of *different* FIPs with identified or identifiable information. An additional safeguard is provided by treating identifiable information with a substantial risk of being identified as a form of identified information. At this point, the risk of being identified has grown too high. Such an approach prevents tactical attempts to use readily-identifiable data in lieu of identified data in order to avoid regulation and responsibility. PII 2.0 addresses the European Union's concern that regulation might be skirted by drawing the boundaries too narrowly, because it builds no hard line between identified and identifiable data, and because this regulatory regime is not all-or-nothing.

F. Applying the New Concept

In this final section, we wish to apply our definitions of PII to the two areas on which this Article focuses: behavioral marketing to adults and food marketing to youth. Regarding the former, PII 2.0

---

388 *Id.* at 5.

389 *Id.* at 17. Regarding ISPs, the Article 29 Working Party has stated that "unless the Internet Service Provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data to be on the safe side." *Id.*

390 *Id.* at 4–5.

391 On the Article 29 Working Party's sweeping definition of PII in the use of Radio Frequency ID (RFID) tags and highly detailed follow up requirements for Privacy Impact Assessments for all uses of RFID, see Article 29 Data Protection Working Party, *Opinion 9/2011 on the Revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications*, 00327/11/EN/WP 180 (Feb. 11, 2011); Article 29 Data Protection Working Party, *Opinion 5/2010 on the Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications*, 00066/10/EN/WP 175 (July 13, 2010); Article 29 Data Protection Working Party, *Working Document on Data Protection Issues Related to RFID Technology*, 10107/05/EN/WP 105 (Jan. 19, 2005); E.U. Data Protection Directive, *supra* note 333. There have been complaints in the European Union that the broad definition of personal data has led to restrictive policies and procedures that have limited medical and social science research. For a recent objection along these lines in a paper that is part of the "Data Protection and the Open Society Project" in the United Kingdom, see David Erdos, *Stuck in the Thicket? Social Research Under the First Data Protection Principle*, 19 INT'L J.L. & INF. TECH. 133 (2011).

leads to a contextual analysis of the data used in behavioral marketing. In many instances, the information now being gathered is in fact identified, not merely identifiable. Since falling into this category will bring regulatory burdens with it, companies will seek to invest in technologies that truly make identification of personal data less likely. The PII 2.0 model also will promote heightened disclosure of commercial practices and demonstrate limits in both the FTC's current approach and the current legal regime. Finally, we point to flawed definitions of PII in the current policy debate about "Do Not Track" and a general privacy statute. We also argue that PII 2.0 will encourage data trade on terms more favorable to those consumers who wish to participate in it.

Regarding food marketing to young persons, PII 2.0 will be relevant to COPPA and beyond. In cases where behavioral marketing involves collection of information with significant risk of identification of a specific child, the full protections of COPPA will apply. Thus, a benefit of PII 2.0 would be to block marketing companies from collecting identified information from young children in the absence of parental consent. Due to certain limitations on COPPA, however, the FTC's transparency jurisprudence will be needed to close significant regulatory gaps.

1. Behavioral Marketing to Adults

As we have shown, behavioral marketing companies now track individuals across different websites or digital media. These efforts involve the use of tracking files being placed on a user's computer and, in some instances, include the sale of individuals' information on data exchanges. Companies have argued that they are not processing PII because they associate their data with a unique identifier that is not immediately associated with information such as a name, address, or social security number.[392]

The information at the heart of targeted marketing is not *non-identifiable* data. Indeed, the promise of these new forms of marketing is to go beyond advertising's past reliance on crude demographical categories and to personalize marketing strategies down to the individual level. Therefore, the critical issue will be whether behavioral marketing implicates identified or identifiable data.

The necessary analysis in PII 2.0 should be contextual. *Identified* information is present when a person's identity has been ascertained, or when there is a substantial risk of identification of a specific indi-

[392] *See supra* Part III (explaining how marketing firms examine online behavior patterns in order to target advertisements towards specific users).

vidual. In contrast, *identifiable* information exists when such a specific identification, while possible, is not significantly probable. Put differently, the question becomes whether the gathering of information pursuant to behavioral marketing, in a specific application, makes an individual reasonably capable of being singled out from others and linked to her identity. In such cases, the law should treat this information as identified. In other cases, the information that is processed may only be identifiable.

Under many circumstances, information gathered through cookies or web beacons can easily be correlated through registration data, connection with static IP addresses, or links with explicitly identifying information at other websites.[393] Since falling into the category of "identified" data traditionally brings greater regulatory scrutiny and at least some enhanced legal burdens with it, PII 2.0 will encourage companies to invest in technologies that reduce the risk of identification of personal data. The goal would be to structure data operations so that the possibility of identifying specific individuals becomes truly remote, which will then lower the risk of data collection and processing.

Beyond the benefits of PII 2.0's contextual determination regarding identified and identifiable information, we think that this approach suggests four insights into the current privacy law landscape. First, when behavioral marketing carries a significant risk of identification of specific individuals, the same level of heightened disclosure of a company's practices should be required as in other circumstances involving the collection of personal data. Since 2009, the FTC has been developing a jurisprudence of transparency that finds companies' failures to adequately disclose processing practices to be deceptive, and hence legally actionable.[394] Milestones in the development of this concept are the FTC's settlements in *Sears* (2009) and *EchoMetrix* (2010), and its Consumer Privacy white paper (2010).[395] The absence of such transparency should be viewed as falling under the FTC's enforcement of unfair and deceptive practices.[396]

Second, PII 2.0 demonstrates certain limits in the FTC's approach and weaknesses in the current legal regime. PII 2.0 will likely lead to

---

393 *See* Angwin, *supra* note 190 (describing tracking companies' ability to create robust consumer profiles by combining information about users' activities across different websites); *supra* Part II (describing how markets can combine multiples pieces of non-PII to create an identifiable profile).

394 For a discussion of the FTC's role, see FTC, PROTECTING PRIVACY, *supra* note 2, at 69–78.

395 *See supra* notes 239–46 and accompanying text.

396 FTC, PROTECTING PRIVACY, *supra* note 2, at 12–13.

at least some behavioral marketing being classified as involving identified information. Moreover, traditional privacy FIPs extend far beyond providing only transparency to consumers.[397] Yet, as the FTC itself conceded in 2010, "the emphasis on notice and choice alone has not sufficiently accounted for other widely recognized fair information practices, such as access, collection limitation, purpose specification, and assuring data quality and integrity."[398] The lack of a general online privacy statute or a specific behavioral marketing statute leaves questionable practices free of effective regulation. The new classifications of PII 2.0 thereby provide support for additional sectoral privacy laws, a number of which have been introduced in Congress.[399]

Third, PII 2.0 also proves to be a useful concept in the current debates around potential legislation. The discussion involves two kinds of legislation: one concerns so-called Do Not Track, and the other concerns a general privacy statute. As for Do Not Track, Congress is now considering legislation that would give individuals the ability to prevent the collection and use of data about their online activities. While the potential for privacy protection is great, the proposed legislation—the Rush Bill—adopts the flawed specific-types approach to PII.

The Rush Bill adopts a rule-based approach to PII. It sets up an initial category of "covered information," including name, postal address, and telephone number. To this category, it adds a second one, which consists of "any other information" that is collected, used, or disclosed "in connection" with these data.[400] This approach turns on a rule that excludes the current practices of behavioral marketing from its reach. As we have discussed, targeted marketing does not collect information "in connection" with name, postal address, and telephone number.[401] Yet, these practices can nonetheless collect identified information in circumstances when there is a significant risk of linking the data collected with a specific individual. Thus, from our viewpoint, this Bill contains a significant flaw in the way that it sets its regulatory threshold.

As for the possibility of a general privacy statute in the United States, the leading candidate at present is the Commercial Privacy Bill

---

397 *See* Schwartz, *supra* note 52, at 907–08 (explaining that transparency to consumers is only one of a handful of different types of obligations traditionally included in FIPs).

398 FTC, PROTECTING PRIVACY, *supra* note 2, at 20.

399 The latest such draft legislation concerns an online privacy bill of rights, as of yet circulating only in draft form, that Senators John Kerry and John McCain are cosponsoring. Julia Angwin, *Proposed Bill Would Put Curbs on Data Gathering*, WALL ST. J., Mar. 10, 2011, at B1.

400 H.R. 5777, 111th Cong. § 2(4)(A)(viii) (2010).

401 *See supra* Part III.

of Rights Act of 2011, which Senators John Kerry and John McCain have co-sponsored.[402] The Bill employs the specific-types approach, but its list of PII is extremely broad, including a catchall category of "[a]ny other information concerning an individual that may reasonably be used by the party using, collecting, or storing that information to identify that individual."[403] Through such language, this Bill resembles the European Union's expansionist approach.

Fourth, for *identifiable* information, this Article has proposed obligations concerning data security, transparency, and data quality. Under PII 2.0, companies will not be able to evade duties associated with information collection and processing by rote arguments that the data are not PII.[404] In particular, we think greater transparency about behavioral marketing, even when exclusively identifiable data are involved, will stimulate data trade on terms more favorable to those consumers who wish to participate in it. As an international example of a related proposal, the Cabinet Office in the United Kingdom is leading a consumer empowerment effort that includes its "mydata" initiative.[405] Its goal is to provide consumers with greater knowledge about how organizations use their personal data. This knowledge is envisioned as "an important stepping stone towards a world where consumers make decisions on the basis of accurate information of their past usage of a service and competitive offers made by sellers."[406] As we have already argued, such increased transparency will go far toward correcting the asymmetry of knowledge between consumers and the companies that track their online behavior.[407]

---

402 Commercial Privacy Bill of Rights Act of 2011, S. 799, 112th Cong. (2011).

403 *Id.* § 3(5)(A)(vii).

404 *See supra* Part III.A.2 (describing how many companies have avoided regulation by arguing that the information they collect does not violate privacy because it cannot be linked to an individual's name).

405 DEP'T FOR BUS. INNOVATION & SKILLS, BETTER CHOICES: BETTER DEALS 17–20 (2011).

406 *Id.* at 17.

407 Schwartz, *supra* note 214, at 2076–80. The treatment of both identified and identifiable information alike will heighten consumer awareness of behavioral marketing at a critical moment. The introduction of the concept of PII 2.0 occurs at a time when consumers know little about behavioral marketing, but are also predisposed to be skeptical towards it.

Regarding the Americans' skepticism toward industry practices, a 2009 survey by Joseph Turow and associates revealed that a majority of Americans do *not* want marketers to tailor advertisements to their interests. Joseph Turow et al., *Americans Reject Tailored Advertisement and Three Activities that Enable It*, University of Pennsylvania Scholarly Commons, 1–4 (2009), http://repository.upenn.edu/asc_papers/137/. The results of the Turow study suggest that greater transparency in this area will promote greater options for consumers and more informed choice.

2. *Food Marketing to Youth*

PII 2.0 will also have significant implications for food marketing to youth. Under PII 2.0, whenever a marketing technique makes an individual reasonably capable of being "singled out" from others and linked to her identity, the law should treat this information as identified. In other cases, the information that is processed may only be identifiable.

Consider the use of "digital command centers" in which staff members of corporations interact with select consumers. Recall the example of Gatorade's "mission control center," in which this company monitors social-media posts twenty-four hours a day.[408] Although we do not know the precise nature of Gatorade's activities, it is not hard to imagine that this company and others are mining social-media posts for information about the youth. Even if companies gather the data in a way that does not involve children's names, many of these new digital command centers would fall within the scope of PII 2.0 as involving identifiable information.

When behavioral marketing is directed toward children under age thirteen, COPPA should fully apply in situations where either identified or identifiable information is involved. Congress's purpose in enacting COPPA was to provide a mechanism for parental consent before companies could collect children's personal information on the Internet.[409] Consistent with Congress's intention, PII 2.0 would update this policy and apply it to the new way of collecting data from children, namely through behavioral tracking that follows Internet activity. As a result, companies would no longer be able to argue that they were not collecting PII about children because they did not have access to a name. The FTC's previous enforcement of COPPA against those who fail to obtain parental consent has been vigorous, and its history of large fines against parties who violate this statute will ensure industry attention once it asserts jurisdiction over behavioral marketing.

At the same time, however, PII 2.0 alone will not overcome certain shortcomings in COPPA. We have noted these above,[410] and now will merely summarize the statute's weaknesses. First, the Act only

---

408 Bauerlein, *supra* note 265; *supra* notes 265–67 and accompanying text.

409 *See Children's Online Privacy Protection Act of 1998: Hearing Before the Subcomm. on Commc'ns of the S. Comm. on Commerce, Sci., and Transp.*, 105th Cong. 2 (1998) (opening statement of Sen. Conrad Burns, Chairman, Subcomm. on Commc'ns of the S. Comm. on Commerce, Sci., and Transp.) (observing that as of June 1998, eighty-nine percent of children's websites collected personal information while only ten percent of the sites provided parental control over collection and use of the information).

410 *See supra* notes 109–16 and accompanying text.

applies to children under thirteen.[411] Second, COPPA extends only to a "website or online service," and third, it regulates these entities only when "directed" to children, or where the operator of the website has "actual knowledge" that it is collecting personal data from children.[412] These restrictions, and the limited vision of technology that they imply, mean that COPPA—like grunge music and Beanie Babies—remains entrenched in the 1990s.

Due to these limitations, the FTC's transparency jurisprudence will again have a role to play. Without overselling the benefits of such heightened disclosure, we do wish to disagree with stereotypes concerning the Facebook generation's lack of concern about privacy. Indeed, in a 2010 survey, Christopher Hoofnagle and co-authors actually found a high level of concern about this topic among young people.[413] A large majority of young people also believe that an individual should have legal rights to know the information that websites have about her and to require them to delete all such stored information.[414]

PII 2.0 would invoke greater transparency about marketing to youth. The FTC's *EchoMetrix* settlement points a way forward. Recall that the FTC complaint in that case concerned a company's secret use of parental control software to collect data about children's computer activity and to feed it to marketers.[415] The FTC did not bring an enforcement action under COPPA; its theory of the case was one of "inadequate disclosure"—a theory that it advanced even though the company supplied language to its customers that arguably covered the underlying activity.[416] Expansion of *EchoMetrix* to the larger digital tracking environment through PII 2.0 will force companies to provide greater information to consumers about the scope and nature of these activities.

---

411 15 U.S.C. § 6501(1) (2006).

412 *Id.* §§ 6501(2)(A), 6502(a)(1).

413 Chris Hoofnagle et al., *How Different Are Young Adults from Older Adults when It Comes to Information Privacy Attitudes & Policies* (Apr. 14, 2010), www.ftc.gov/os/comments/copparulerev2010/547597-00005-54505.pdf.

414 *Id.* at 15–16. Finally, Hoofnagle and his co-authors argue, "the savvy that many attribute to younger individuals about the online environment doesn't appear to translate to privacy knowledge." *Id.* at 17. The survey found that higher proportions of young adults than older ones "believe incorrectly that the law protects their privacy online and offline more than it actually does." *Id.* at 4.

415 *See supra* notes 243–45 and accompanying text; Complaint for Permanent Injunction and Other Equitable Relief, FTC v. EchoMetrix, Inc., No. CV10-5516 ¶¶ 8–14 (E.D.N.Y. Nov. 30, 2010).

416 *Id.* ¶ 12 (E.D.N.Y. Nov. 30, 2010). The critical information was both buried in a Terms of Service notice and obscure in its phrasing. *Id.*

Thus, one aspect of a response to food marketing to children should rely on a transparency approach. The need is for greater information to be granted to youth and parents about how companies gather PII in the new digital marketing landscape. At the same time, however, transparency alone does not represent a full range of FIPs as they are traditionally understood, and marketing campaigns directed toward youth and adults will sometimes occur without collecting PII, even under our new definition.

On a concluding note, we wish to observe that information privacy law cannot solve all the social issues associated with food marketing to children. There is a need, for example, to draw on consumer protection law and public health law. While an examination of this topic is beyond the scope of this Article, we can simply observe that, fortunately, a broad and multi-pronged public policy effort is now being directed toward this issue. The highest profile participant in the debate is First Lady Michelle Obama, who is directing the nation's attention to the many dimensions of this public health crisis.[417] At the federal interagency level, a working group is developing nutrition principles to guide industry when it markets foods to children ages twelve to seventeen years old.[418] The FTC, Food and Drug Administration, Centers for Disease Control and Prevention, and United States Department of Agriculture are the agencies involved in this effort "to improve the nutritional profile of foods marketed to children."[419]

Conclusion

Personally identifiable information (PII) is one of the central concepts in information privacy regulation. The basic assumption behind relevant statutes is that their applicability will turn on the presence or absence of PII. At the same time, and surprisingly, there is no uniform definition of PII in information privacy law. Moreover, the definitions that do exist are unsatisfactory.

---

417 Sheryl Gay Stolberg & William Neuman, *Restaurant Nutrition Draws Focus of First Lady*, N.Y. Times, Feb. 7, 2011, at A11.

418 Interagency Working Group on Food Marketed to Children (Apr. 2011), *available at* http://www.ftc.gov/os/2011/04/110428foodmarketproposedguide.pdf.

419 *Id.* at 1. As the *New York Times* summarized the new guidelines, "[r]egulators are asking food makers and restaurant companies to make a choice: make your products healthier or stop advertising them to youngsters." William Neuman, *U.S. Seeks New Limits on Food Ads for Children*, N.Y. Times, Apr. 28, 2011, at B1. Public health advocates have also developed a sound methodology of possible regulatory approaches. In one of the most useful, that of the Berkeley Media Studies Group, policy strategies are targeted along the concepts of the "four P's": products, places, promotions, and price. Berkeley Media Studies Grp., Fighting Junk Food Marketing for Kids 18 (2006).

In response, this Article has developed a new concept of PII. Its model of PII 2.0 protects information that relates either to an identified or identifiable person, and associates different legal interests with each category. This flexible approach provides the safeguard of treating identifiable information that has a substantial risk of being identified as a form of identified data. Additionally, such an approach has the merit of preventing tactical attempts to use readily identifiable data in lieu of identified data in order to avoid regulation and responsibility.

PII 2.0 represents a way beyond the reductionist reading of PII in the United States and the expansionist reading in the European Union. Its use would represent a significant step forward in responding to the privacy implications of behavioral marketing generally, as well as the marketing of unhealthy food products to youth, specifically. In this Article we have argued that PII cannot be abandoned, as this concept is essential to defining regulatory boundaries. At the same time, however, information privacy law faces limits in its policy reach. Other kinds of law and additional policy initiatives are needed as part of a response in public policy to the negative implications of the food industry's marketing techniques.

APPENDIX

Google Books Ngram Viewer,
References to "Information Privacy" 1950–2000