07/08/11
PII ARTICLE FINAL FPF

Forthcoming NYU LAW REVIEW (2011)

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

by

Paul M. Schwartz
&
Daniel J. Solove

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

By
Paul M. Schwartz*
Daniel J. Solove**

---

* Professor, Berkeley Law School, U.C. Berkeley; 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 of this project. We also wish to thank Leah Duranti, Yan Fang, Bill Friedman, Matthew Galati, Melissa DeJesus, and Shawn Curtis, who provided research assistance. Marty Abrams, Jules Polonetsky, Chris Hoofnagle, and participants at the Fourth Annual Privacy Law Scholars Conference, Berkeley, California, provided helpful suggestions on this paper.

1. Reductionism in the U.S.
2. Expansionism in the EU
3. The Benefits of PII 2.0
D. PII 2.0 and Fair Information Practices (FIPs)
E. Possible Objections
F. Applying the New Concept
1. Behavioral Marketing to Adults
2. Food Marketing to Youth
Conclusion

INTRODUCTION

Information privacy law has reached a turning point. The debate about the topic is vigorous at present, and polling data reveal that Americans are highly concerned about privacy on and off the Internet.[1] Moreover, the Executive Branch, independent agencies, and Congress are considering different paths to revitalizing information privacy.[2] At the same time, regardless of the nature of any 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; information outside of it is not.

Thus, the concept of PII is one of the most central in privacy regulation. It defines the scope and boundaries of a large range of privacy statutes and regulations. Federal statutes that turn on this distinction include the Children's Online Privacy Protection Act, the Gramm-Leach Bliley Act, the HITECH Act, and the Video Privacy Protection Act.[3] Moreover, state statutes that rely on PII as a jurisdictional trigger include California's Song-Beverly Credit Card Act and the forty-six state breach notification laws.[4] These laws all share the same basic

---

[1] Commonsense Media, *Online Privacy: What Does It Mean to Parents and Kids?* (2010), at http://www.commonsensemedia.org/sites/default/files/privacypoll.pdf; Gallup, *U.S. Internet Users Ready to Limit Online Tracking for Ads*, (2010), at http://www.gallup.com/poll/145337/Internet-Users-Ready-Limit-Online-Tracking-Ads.aspx.

[2] Contrast, for example, the recent reports of 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. DEPARTMENT OF COMMERCE, INTERNET POLICY TASK FORCE, COMMERCIAL DATA PRIVACY AND INNOVATION IN THE INTERNET ECONOMY: A DYNAMIC POLICY FRAMEWORK (2010), FEDERAL TRADE COMMISSION, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (2010)[hereinafter FTC, PROTECTING PRIVACY].

[3] *See* Part I.B., *infra.*

[4] *Id.* For a discussion of the breach notification statutes, see DANIEL J. SOLOVE & PAUL M. SCHWARTZ, PRIVACY LAW FUNDAMENTALS 135-39 (2011). For an up-to-date listing of these statutes, see National Conference of State Legislatures, State Security Breach Notification Laws, at http://www.ncsl.org/Default.aspx?TabId=13489.

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.

At the same time, and surprisingly, information privacy law in the U.S. lacks a uniform definition of PII. Moreover, 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. Recent scholarship has also challenged PII as a fatally-flawed concept. In the view of Paul Ohm, privacy law must abandon its reliance on PII and find an entirely new regulatory paradigm.[5]

In contrast, this Article contends that information privacy law needs a concept of PII – it cannot jettison PII as one of its central dimensions. At the same time, PII must be reconceptualized if privacy law is to remain effective in the future. Therefore, we develop a conception of PII 2.0, and one which avoids the problems and pitfalls of the current approaches. The key to our model is to build two categories of PII, "identified" and "identifiable" data, and to treat them differently.[6] This approach permits tailored legal protections built around different levels of risks to individuals. It also represents a path forward, and one that avoids the reductionist view of PII of the U.S., and the expansionist one of the European Union (EU). In the reductionist view, the tendency is to consider PII as only that personal data which has been actually associated with a specific person. This model protects only identified data and 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 linked in the future. Thus, the EU treats identified and identifiable data as equivalent categories. In our view, the continuum of risk is different for these categories, and the necessary legal protections should also be different.

This Article proceeds in four steps. In Part One, we explore the central role of PII and the grounds for its current uneasy status. The concept of PII is one that only arose 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 for a set of first generation information privacy statutes. It was only in the Cable Communications Policy Act of 1984 that Congress settled on the classic model for these laws: once an entity collected or processed PII, it would be obligated to provide privacy safeguards.[7] Nonetheless, as Part One also demonstrates, there is no standard nomenclature for PII, and no standard definition of it. We explore the three basic approaches of U.S. lawmakers to defining PII and find the current formulations of PII to be deeply unsatisfactory.

In Part Two, we engage in a broader analysis of the weaknesses of PII as it is conceptualized today. First, many people believe in an "anonymity myth,"

---

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

[6] *See* Part IV.D, *infra.*

[7] *See* Part I.A., *infra.*

which is a belief is that people are anonymous unless they formally use 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 depends upon changing 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 Three, 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—and ones that exclude names but that contain a wealth of details about individuals. Online 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.[8] Due to the epidemic of obesity among minors in the U.S., the targeted marketing of unhealthy food products to youth is now a highly significant public health issue. The Children's Online Privacy Protection Act (COPPA) restricts websites from gathering and using information gathered from children, but it has also weaknesses that permit companies to argue that they are engaging in behavioral marketing without PII.

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, and a rule, its counterpart, is a harder-edged benchmark.[9] 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 by demonstrating the merits of this new approach in the context of behavioral marketing and food marketing to youth.

I. THE CENTRAL ROLE OF PII AND ITS UNEASY STATUS

In this Part, we examine how and why PII became a central concept in information privacy law. Due to computer processing of data, Congress was forced to confront the issue of the kinds of data that should matter for information privacy law. Despite legislative grappling with this issue over several decades, there is still no uniform definition today of PII in the U.S. We identity three current models of PII and demonstrate why each is a failure.

---

[8] *See* Part II.B., *infra.*

[9] *See* Part IV.B., *infra.*

A. THE RISE OF PII AND ITS SIGNIFICANCE

The concept of PII arose during the last fifty years. PII went from not being a consideration in privacy law to becoming one of its central concepts. The early jurisprudence of privacy law lacked a concept of PII. In their famous 1890 article, Samuel Warren and Louis Brandeis merely assumed that privacy would involve information identifiable to a person.[10] They conceived of privacy as a right of "personality."[11] 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.[12] The paradigmatic privacy invasion for Warren and Brandeis concerned the press invading the privacy of people by printing gossip about them.[13] Warren and Brandeis viewed such media reports as necessarily concerning information that would identify a person; otherwise, the gossip would have no sting. They thus did not consider PII as an issue warranting any 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.[14] 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.[15] 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.[16]

PII first became an issue in the 1960s with the rise of the computer. The computer permitted public bureaucracies and private companies to process

---

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

[11] *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 their conception of privacy as a right of personality, see Paul M. Schwartz & Karl-Nicholas Peifer, *Prosser's* Privacy *at Fifty*, 98 CALIF. L. REV. 1925, 1943-44 (2010).

[12] In their view, a privacy tort was needed to protect each person's "emotional integrity," as Robert Post later summarized their thought. Robert C. Post, *Rereading Warren and Brandeis: Privacy, Property, and Appropriation*, 41 CASE W. RES. L. REV. 647, 663 (1991). For many years after Warren and Brandeis' article, other authors on the subject of tort privacy rallied around the notion of the right of personality as the basis for such an interest. Schwartz & Peifer, *supra* note 11, at 1944-47.

[13] Warren & Brandeis, *supra* note 10, at 196.

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

[15] *Id.* at 389-99. 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 (1980).

[16] *See, e.g.*, Prosser, *supra* note 14, at 392-98 (discussion of "public disclosure of private facts").

personal data.[17] The computer did not merely increase the amount of information that entities collected; it changed how they could organize, access, and search it. 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."[18] Unlike manual systems, such as a telephone book, "computers [could] easily be programmed to sort or reorganize data on the basis of any particular index, attribute, or characteristic."[19] The key point, as the Commission noted, is 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.[20] This capacity of computers changed the way information could be linked to an individual. Previously, in order for information to be connected to people, it would have almost invariably had 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 applicability 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."[21] The Commission did not discuss the issue in terms of PII, but as "who and what is covered."[22] No longer was it possible to assume that 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. The development of computerized records, thus, required Congress to confront the issue of the kinds of information that should matter for information privacy law.

The initial focus of Congress was to view the types of records at stake as determinative in triggering a statute's protections. The Fair Credit Reporting Act (FCRA) of 1970, the Family Education and Records Privacy Act (FERPA) of 1974, and the Privacy Act of 1974 demonstrate this approach as well as the

---

17 Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1402 (2001). The classic early studies in American social sciences and law are ARTHUR MILLER, THE ASSAULT ON PRIVACY (1972) and ALAN WESTIN, PRIVACY AND FREEDOM (1967).

18 PRIVACY PROTECTION STUDY COMMISSION, TECHNOLOGY & PRIVACY 21-22 (1977)[hereinafter PRIVACY COMMISSION, TECHNOLOGY].

19 *Id.* at 21.

20 *Id.* at 21-22. In his prescient study, *The Assault on Privacy*, Miller also discusses the "retrieval capacity" of the computer. MILLER, *supra* note 17, at 54.

21 PRIVACY COMMISSION, TECHNOLOGY, *supra* note 18, at 45 (1977).

22 *Id.*

weaknesses associated with it.[23]

FCRA was the first federal sectoral privacy statute. It applies to any "consumer reporting agency" (CRA) that furnishes a "consumer report."[24] 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 a limited set of other purposes.[25] FRCA sets legal restrictions on the circumstances under which a CRA agency can furnish a consumer report to another party, as well as the use of these reports for purposes such as law enforcement and employment offers.[26] In sum, it focuses on the organization of data about a person (namely, whether it appears in a "consumer report"), and the party who collects and uses the information (the CRA).[27] Of the two categories, the concept of the consumer report is the most important.[28] Due to FCRA's definitional approach, moreover, there are notable gaps in its coverage.[29]

Enacted four years after FCRA, FERPA focuses on student privacy. It was also the first federal statute to refer to "personally identifiable information,"

---

[23] For FERPA, see 20 U.S.C. § 1232g (2006). For the Privacy Act, see 5 U.S.C. § 552a (2000).

[24] 15 U.S.C. § 1681b(d) (2006).

[25] *Id.*

[26] For a discussion, see SOLOVE & SCHWARTZ, FUNDAMENTALS, *supra* note 4, at 86-91.

[27] *Id.; Financial Privacy Law*, in Proskauer on Privacy, 2-7 to 2-14 (Kristen Matthews, ed. 2011).

[28] As *Proskauer on Privacy* observes, "Given that the definition of a CRA depends largely on the definition of a '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." *Financial Privacy Law*, in Proskauer on Privacy, *supra* note 27, at 2-10.

[29] 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). For case law reaching this conclusion, see *Mirfasihi v. Fleet Mortgage Corporation*, 551 F.3d 682, 686 (7th Cir. 2008)(Posner, C.J) and *Smith v. First National Bank of Atlanta*, 837 F.2d 1575, 1578 (11th Cir. 1988). Although such entities provide a CRA with information about consumers, the entities themselves are not in the business of supplying a consumer report to third parties. In addition, FCRA contains another problematic and explicit exception to its definition of consumer reports. The term does not extend to the sharing of information among affiliated entities so long as the consumer is given an opportunity to "opt-out" from such sharing. 15 U.S.C. § 1681a(d)(2)(A)(iii). In Congressional testimony in 2003, Joel Reidenberg already pointed to the consequences of this exemption: it "means that credit report information loses protection when shared with far-flung related companies." Testimony of Joel R. Reidenberg, Hearing: Affiliate Sharing Practices and Their Relationship to the Fair Credit Reporting Act (June 26, 2003), at http://banking.senate.gov/public/index.cfm?FuseAction=Hearings.Testimony&Hearing_ID=d545d5cd-9273-4ad3-a574-58f85d1e7af4&Witness_ID=bea3b438-e1a9-4bef-bf1a-152f1430af94.

or PII.[30] FERPA uses the term when prohibiting educational entities from releasing or providing access to "any personally identifiable information in education records."[31] Despite mentioning PII, however, the statute's key concept is "education records," which it defines as referring to "information directly related to a student" that an educational institution itself "maintains" in a file or other record.[32] The statute's coverage depends on whether or not a school has first organized and then stored data in education records.[33]

Due to FERPA's limitations, schools long profited by distributing "surveys" on behalf of marketers.[34] Since the collected information went from parents and children to the marketers without being "maintained" in "educational records" by schools, this practice fell outside of FERPA.[35] Congress finally responded in 2005, but in a limited fashion. It left FERPA unaltered and created a limited separate statutory interest that permits parents of elementary and secondary students the ability to opt out of the collection of student information for commercial purposes.[36] Congress neither revisited the reliance in FERPA on the concept of "educational records," nor created a more basic right to block release of student records for commercial purposes. As for universities, they remain able to sell essential student contact information to credit card companies.[37] Such data is considered "directory information," and, hence, not an "educational record."[38]

In a fashion similar to FCRA and FERPA, the Privacy Act's threshold turns on how record systems are organized rather than on whether the information can be linked to the individual. The key trigger of the Privacy Act

---

30 20 U.S.C. § 1232g(b)(2). 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.

31 20 U.S.C. § 1232g(b)(2).

32 20 U.S.C. § 1232g(a)(6).

33 JAMES RAPP, 5 EDUCATION LAW 13.04[7][a] (2010). The Supreme Court has also heard the siren call of protection based on the type of records. In 2002, in *Owasso Independent School District v. Falvo*, 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. Owasso Independent School District v. Falvo, 534 U.S. 426, 435-38 (2002).

34 Lynn M. Daggett, *FERPA in the Twenty-First Century: Failure to Effectively Regulate Privacy for All Students*, 58 CATH. U.L. REV. 59 (2008).

35 20 U.S.C. § 1232g(a)(6).

36 Daggett, *supra* note 34, at 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, § 507, 115 Stat. 272, 367-68 (codified as amended at 20 U.S.C. § 1232g(j)(1) (Supp. V 2005)).

37 Margaret O'Donnell, *FERPA: Only a Piece of the Privacy Puzzle,* 29 J. C. & U. L. 679, 684 (2003).

38 20 U.S.C. § 1232g(b)(2).

concerns how federal agencies retrieve information from a database; it applies only when information is retrieved from a "system of records."[39] Further, 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."[40] As a consequence, the Privacy Act only covers computer searches that identify an individual when retrieval of data was done through reference to a *specific personal identifier*, such as a name, or Social Security Number.[41]

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 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, and other personal attributes is *not* retrieving data from a system of records by use of an identifying particular assigned to a person.[42] Within three years of the statute's enactment, the Privacy Protection Study Commission had already drawn attention to and condemned this profound flaw.[43] 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. The statute not only refers to PII like FERPA, but also make this term the trigger for the applicability of the law.[44] 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 privacy law. These principles establish obligations for organizations that process personal information.[45] The Cable Act prohibits a

---

[39] The Privacy Act of 1974, 5 U.S.C. § 552a.

[40] 5 U.S.C. § 552a(a)(5). 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 number, symbol, or other identifying particular assigned to the individual, such as finger or voice print or a photograph." *Id.* at § 552a(a)(4).

[41] As the Department of Justice's guide to the Privacy Act summarizes, "The 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." Department of Justice, Overview of the Privacy Act of 1974 (2010), at Definitions, E System of Records, http://www.justice.gov/opcl/1974definitions.htm#system.

[42] PRIVACY PROTECTION STUDY COMMISSION, THE PRIVACY ACT OF 1974: AN ASSESSMENT 6-7 (1974); OMB Guidelines to Privacy Act, 40 Fed. Reg. 28,948, 28,952 (July 9, 1975).

[43] PRIVACY PROTECTION STUDY COMMISSION, PERSONAL PRIVACY IN AN INFORMATION SOCIETY 59-61 (1977) [hereinafter PRIVACY COMMISSION, PRIVACY REPORT].

[44] 47 U.S.C. § 551(a)(2). The legislative history of the Cable Act proves singularly unhelpful regarding the selection of this term. *See* H.R. Rep. No. 98-934, at 75-80 (1984).

[45] In the U.S., the Department of Health, Education and Welfare had first mentioned

cable operator from using a cable system from collecting PII "concerning any subscriber without the prior written or electronic consent of the subscriber concerned."[46] It provides for subscriber access to all PII "regarding that subscriber which is collected and maintained by a cable operator."[47] It requires notice to a subscriber about the nature of PII "collected or to be collected with respect to the subscriber of the nature of the use of such information."[48]

The contrast with FCRA, FERPA, and the Privacy Act is clear. 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, its statutory obligations fall on a cable operator as soon as this entity collects PII.

What inspired this important shift in the law between the early 1970s and 1984? First, a renewed focus on the topic of information privacy began during the latter part of the 1970s. Google Ngram provides a convincing demonstration of this development; this Google database permits statistical analysis of the use of words and phrases.[49] Appendix A to this Article contains the Ngram Viewer's chart for the term "information privacy" between 1950 and 2000. In particular, the chart 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.[50] Moreover, Congress debated and enacted the Cable Act in the shadow of George Orwell's signature year, 1984. This notable event heightened the concern about privacy in the U.S.[51]

FIPs in an influential report in 1973. U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS, COMPUTERS AND THE RIGHTS OF CITIZENS 29-30 (1973). On the policy history of FIPs, see PRISCILLA M. REGAN, LEGISLATING PRIVACY 75-85 (1995). For an introduction to FIPs, see Paul M. Schwarz, *Preemption and Privacy*, 118 YALE L.J. 907-908 (2009); DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 655-58 (3d ed., 2009)[hereinafter SOLOVE & SCHWARTZ, IPL].

46 47 U.S.C. § 551(b)(1).

47 *Id.* at § 551(d).

48 *Id.* at § 551(a)(1)(A).

49 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. *See* Patricia Cohen, *In 500 Billion Words, New Window on Culture*, N.Y. TIMES, Dec. 17, 2010, at A3.

50 The chart also shows how this attention only became more intense throughout the 1990s and the emergence of the Internet and other threats to privacy. Appendix A, infra.

51 As one law review article stated: "To prevent cable from turning the television set into an Orwellian nightmare, the Act creates a framework for the protection of subscriber privacy." Michael Meyerson, *The Cable Communications Policy Act of 1984: A Balancing on the Coaxial Wires*, 9 GA. L. REV. 543, 612 (1985); *see also* Mindy Elisa Wachtel, *Videotex: A Welcome New Explosion 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."); John

Most importantly, however, the collection by cable operators of personal information created the same kinds of issues that the Internet would later raise. In the 1980s, observers already 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.[52] The cable operator would collect these data, which permitted the construction of 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.[53] A "videotex explosion" would lead, in turn, to the conveying of detailed data about one's "interests, choice, and views to the central computer" of the system operator.[54] As a result of these concerns, the policy response in the Cable Act was to regulate around information rather than how the collector of the system organized data. This regulatory insight, once reached, established the model for information privacy regulation to come.

Subsequent to the 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.[55] 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.[56] Nevertheless, PII has become the preferred term of

---

Shattuck, *In the Shadow of 1984: National Identification Systems, Computer-Matching and Privacy in the United States*, 35 HASTINGS L.J. 991, 991 (1984) (observing how recent actions by the federal government had "brought the technology invasion from the realm of science fiction into the real world of public policy").

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

[52] 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 51, at 612.

[53] Wachtel, *supra* note 51, at 287.

[54] *Id.* at 289.

[55] For illustrative laws, see The Driver's Privacy Protection Act, 18 U.S.C. §§ 2721-2725 (2000); Children's Online Privacy Protection Act, 15 U.S.C. §§ 6501-6506 (2000); VPPA, 18 U.S.C. § 2710 (2000); California Breach Notification Statute, Cal. Civ. Code §§ 1798.29, 1798.82, 1798.84 (2008).

[56] For two examples, see William McGeveran, *Disclosure, Endorsement, and Identity in Social*

art since the mid-1990s.

Even more troublesome than the insistent 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 became 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 arose in privacy laws, but with little thought as to the selection of one rather than the other. As we will discuss in the next Section, moreover, all of these definitions are flawed.

The reason for these difficulties is that PII is a challenging conceptual issue at the heart of any system of regulating privacy in the Information Age. 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 non-identifiable and turn it into identifiable data. Moreover, industry is involved in practices that raise privacy concerns, but that do not fall within any of the current definitions of PII. Thus, despite being a ubiquitous and central concept in privacy law, PII lacks a consistent definition and its complexities have not been adequately explored. 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—ones that adequately protect privacy, are flexible and evolving, yet stable. But PII is a complicated and hard-to-pin-down concept.

While the edifice of privacy law is built on PII, only recently has some awareness emerged about the conceptual problem at the core of PII. 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."[57] The FTC pointed to the need to rethink PII, but did not make any headway beyond this call.[58] In scholarly literature, moreover, there has been surprisingly scant attention to the issue of PII. In 1997, Jerry Kang devoted several pages in a seminal early paper about Internet privacy to a discussion of when data became "personal information."[59] More recently, Paul Ohm published a major piece

---

Marketing, 2009 U. ILL. L. REV. 1105 (2009); Lior Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHICAGO L. REV. 919 (2005).

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

[58] In this report, the FTC stated that it would leave the question open as to 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. Its 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.*

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

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.[60] 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.

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 decision-making yardstick, and a rule, its counterpart, is a harder-edged decision-making tool.[61] 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).[62] It proves possible to organize the existing approaches to defining PII into the category of 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 defines PII as any information that identifies a person. It is an example of a standard. The Video Privacy Protection Act (VPPA) neatly demonstrates this model.[63] The VPPA, which safeguards the privacy of video sales and rentals, simply defines "personally identifiable information" as "information which identifies a person."[64] For its purposes, therefore, information which identifies a person is PII and falls under its

---

[60] Ohm, *supra* note 5, at 1742.

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

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

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

[64] 18 U.S.C. § 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. *Id.* at § 2710(b)(2)(B). It defines "videotape service providers" in a technological neutral fashion to permit the law to be extended to DVDs. § 2710(a)(4).

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. 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 toward defining PII is to focus on non-public information. Here, too, we see the use of a standard. The non-public 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 says that PII is all that is not aggregate. Its logic is that such information does not identify a person.

The Gramm-Leach Bliley Act (GLBA) epitomizes this approach by defining "personally identifiable financial information" as "nonpublic personal information."[65] The statute fails to define "nonpublic," but presumably it means information not found within the public domain.[66] In a related fashion, the Cable Act defines PII as something other than "aggregate data."[67] This statute, which protects the privacy of subscribers to cable services, views PII as excluding "any record of aggregate data which does not identify persons."[68] By aggregate data, the Cable Act presumably means purely statistical information that does not identify specific individuals.[69]

The problem with the non-public approach is that it does not map onto whether, in fact, the information is identifiable. It is a standard that is not likely to work well. The public or private status of data often does not match up to whether it can identify a person or not. A person's name and address might be considered public information; for example, 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 information. Yet, an approach that only

---

65 Gramm-Leach-Bliley Financial Modernization Act of 1999, 15 U.S.C. § 6809(4)(A) (1999).

66 During GLB 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 NC 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 government records, widely distributed media (including the Internet), or disclosures to the general public required to be made by federal, state or local law." *Id.*

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

68 *Id.*

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

protects non-public information as PII might not preclude such combinations.

3. THE SPECIFIC-TYPES APPROACH

The third approach is to list specific types of data that constitute PII. This technique is a classic approach to defining a rule. In the context of the specific-types approach, if the information falls into an enumerated group, it becomes a kind of statutory "per se" PII. To illustrate three different variations on this approach, we can examine the Massachusetts Breach Notification Statute of 2007, California's Song-Beverly Credit Card Act of 1971, and the federal Children's Online Privacy Protection Act (COPPA) of 1998.

The Massachusetts Breach Notification Statute requires notification of affected individuals in the case of a loss or leak of their personal information.[70] The Act defines PII as a person's first name and last name, or first initial and last name in combination with a limited amount of other elements: (1) Social Security Number; (2) a driver's license number; or (3) a financial account number, or credit or debit card number.[71]

The Song-Beverly Act prohibits merchants who accept credit cards from collecting a cardholder's "personal identification information" during business transactions with the cardholder.[72] 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."[73] 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."[74]

Finally, the federal Children's Online Privacy Protection Act (COPPA) regulates the collection and use of children's information by Internet websites or online services.[75] 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 individual collected online" that includes a number of elements beginning with "first and last name," and continuing through a physical address, Social Security Number, telephone number, and email address.[76] Its 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."[77] In 2000, the FTC made

---

[70] Massachusetts Breach Notification Statute, 201 Mass. Code Regs. § 17.00 *et seq.* (2010).
[71] 201 Mass. Code Regs. § 17.02 (2010).
[72] Song-Beverly Credit Card Act of 1971, Cal. Civ. Code § 1747.8 (2009). The Act uses the term "personal identification information." This language reinforces our earlier point that there is no standard nomenclature for PII. *See* Part I.A, *infra.*
[73] Cal. Civ. Code § 1747.8(a)(1).
[74] Cal. Civ. Code § 1747.8(b).
[75] Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501-6506 (1998).
[76] *Id.* at § 6501(8)(A)-(E).
[77] *Id.* at § 6501(8)(F).

use of this standard when it issued its COPPA Rule.[78] 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."[79]

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 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.[80] 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 a specific 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 wrongly assumes that the types of data that are identifiable to a person are static.[81] As we will argue later in this Article, however, this assumption is false. This variant of the specific-types approach is too rigid.

As for the version of the specific-types approach in the Song-Beverly Act, its text appears far less narrow than the Massachusetts statute.[82] Nonetheless, a recent series of decisions demonstrate how easy it is for PII to be interpreted only as information exclusive to one person. Two lower courts in California had interpreted this statute as providing extremely limited protection.[83] While the California Supreme Court in 2011 corrected their interpretation of the statute, the general flaw of the specific-types approach remains after this decision.[84]

In *Pineda v. Williams-Sonoma,* Jessica Pineda visited a 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 stated, "The cashier asked her for her zip code, but did not tell her the consequences if she declined to provide the information."[85] Pineda believed that she was obliged to provide this information to complete the transaction, and she supplied it to the cashier.[86] 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

---

78 16 C.F.R. § 312.2 (2011).

79 *Id.*

80 201 Mass Code Regs. § 17.02.

81 *Id.*

82 Cal. Civ. Code § 1747.8.

83 Trial Order, Pineda v. Williams-Sonoma Stores, Inc., 2008 WL 7414542 (Cal. App. Dep't Super. Ct. Oct. 29, 2008); Pineda v. Williams-Sonoma Stores, Inc., 100 Cal. Rptr. 3d 458 (Ct. App. 2009), *rev'd*, 51 Cal. 4th 524 (2011).

84 Pineda v. Williams-Sonoma Stores, Inc., 51 Cal. 4th 524 (2011).

85 *Id.* at 460.

86 *Id.*

number, the name on her credit card, and her zip code.[87]

As we have seen, the critical language in the Beverly-Song 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."[88] This language appears broad; nonetheless, 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.[89] 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."[90] For that court, "a zip code was not facially individualized information."[91]

The California Supreme Court corrected this verdict, but it did so on the narrowest possible grounds. It analyzed the statutory language and legislative history, and found that both supported a legislative intent to include a zip code as part of the "cardholder's address."[92] In other words, that statutory category included "not only a complete address, but its components."[93] Yet, the California Supreme Court had only tweaked a sub-category within the specific-types approach. It did not reach a 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 unique identification of a person. The zip code, although shared by as many as tens of thousands of people, was precisely the piece of information, when added to a person's name, which permitted linkage of the customer to a wealth of PII about her.[94] As the state Supreme Court itself observed, once the Williams-Sonoma store had the 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.[95] In fact, the store had created a database to market products to its customers as well as to have the possibility of selling "the information it has compiled to other businesses."[96]

As for COPPA, our third example of the specific-types approach, the federal

---

87 *Id.*

88 Cal. Civ. Code § 1747.8(b) (2009).

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

90 *Id.*

91 *Id.*

92 *Pineda*., 51 Cal. 4th 524, 529 (2011).

93 *Id.*

94 *Id.* at 534.

95 As the Court of Appeals itself conceded, the store "used customized computer software to perform reverse searches from databases that contain millions of names, e-mail addresses, residential telephone numbers and residential addresses." *Pineda*, 100 Cal. Rptr. 3d at 460.

96 *Pineda*, 51 Cal. 4th at 534.

statute has an aspect that the Massachusetts and California statutes lack. COPPA explicitly references FTC rulemaking as a way to expand and adapt its definitions of PII.[97] As we have seen, moreover, the FTC, in its COPPA rule, added one element to the statutory concept of PII, namely, the idea of "a persistent identifier," such as a cookie.[98] In the statutory definition, however, the FTC's ability to expand the list of identifiers is cabined by a requirement that the information be used to permit the "contacting of a specific individual."[99] As we will discuss in more detail later, there are also indications that this agency is unlikely to define "contacting" to include serving specific ads to a person.[100]

A final difficulty with COPPA, as typical for a rule, is that the statute requires that PII be defined in advance.[101] The COPPA twist is to permit the statutory listing to be expanded through agency rulemaking.[102] Nonetheless, the risk is that new technology will develop too quickly for this approach to be effective. For example, the 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.[103] In his study of co-regulatory privacy approaches, Ira Rubinstein traces a long cycle, one 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."[104]

* * *

Despite the importance of the concept of PII to privacy law and regulation, there remains a lack of consensus in the U.S. about how to define PII. All of the 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 on whether the data is 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 to distinguish the nature of the information that 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-

---

97 15 U.S.C. § 6501(8)(F).
98 16 C.F.R. § 312.2 (2011).
99 15 U.S.C. § 6501(8)(F).
100 *See* Part IV.C., *infra*.
101 15 U.S.C. § 6501(8).
102 15 U.S.C. § 6501(8)(F).
103 Ira Rubinstein, *Privacy and Regulatory Innovation: Moving Behind Voluntary Codes*, I/S (forthcoming 2011).
104 *Id.*

PII became less certain. Today, that line is not merely uncertain: Professor Ohm questions whether maintaining a distinction between PII and non-PII is even possible. 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 pre-determined *a priori*.

In this Section, we proceed through four steps to show defects in the existing distinction between PII versus non-PII. First, we discuss a widely shared misunderstanding about anonymity on the Internet. Many people believe that since they do not formally use their name in many settings in cyberspace that they are anonymous. Due to the growth of static IP addresses, however, once one crosses the threshold of cyberspace, there is a basic level of built-in identifiability. 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 forge a link to 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 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.

A. The Anonymity Myth and the IP Address

There is common myth about anonymity on the Internet. Many people 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 this incorrect assumption that as long as one does not explicitly do something under one's actual name 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. The assumption sometimes takes the form of a belief that so long as a person does not supply her name to a given website, then it is possible to surf it anonymously. An additional belief is 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 be anonymous online, this anonymity is only as protective as 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 and 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 wish to explore a threshold issue, one at the entry to cyberspace, which contributes significantly to traceability: the IP address. In later sections, we will discuss a number of other factors that contribute to such traceability on the Internet.

The IP address is a unique identifier that is assigned to every computer connected to the Internet.[105] Due to the shift from dial-up to static IP addresses, Internet Service Providers (ISP's) now have logs that link IP addresses with particular computers and, in many cases, eventually to specific users.[106] To understand why these links exist, it will be useful to trace the shift that has occurred from dial-up Internet service to broadband.

Like the Sony Walkman and cassette tapes, dial-up service is a cultural relict of fading significance. To take a trip down memory lane, we should recall that 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 Internet access.[107] 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 that she connects to the Internet.[108] As a consequence, many customers share a single IP address at different times over the course of a single day. Moreover, ISP's typically do not retain records about dynamic IP use for more than a few weeks.[109] 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 people on the Internet began to access it through high-speed services, such as cable or DSL.[110] The positive aspect of such broadband access is to permit a wide range of activities in cyberspace, including multi-media 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 that do not change. A long-standing DSL or cable account will have the same IP address for years.[111] 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.

---

105 PRESTON GRALLA, HOW THE INTERNET WORKS 14 (1999).

106 GARY BAHADUR ET AL., PRIVACY DEFENDED 194 (2002).

107 AARON SMITH, PEW RESEARCH CENTER'S INTERNET & AMERICAN LIFE PROJECT, HOME BROADBAND 6 (2010), at http://pewinternet.org/Reports/2010/Home-Broadband-2010.aspx.

108 BAHADUR ET AL., *supra* note 106, at 195.

109 *Id.*

110 JOHN B. HORRIGAN, PEW RESEARCH CENTER'S INTERNET & AMERICAN LIFE PROJECT: HOME BROADBAND 2 (2008).

111 BAHADUR ET AL., *supra* note 106, at 195.

The identification of a seemingly anonymous Internet user can easily follow from an IP address. Connection to a website requires a browser to share an IP address, and look-up tools available on the Internet permit certain information to be revealed about the IP address.[112] The details include the hostname, geographic location information, and a map.[113] With such access to the 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.

To be sure, IP addresses do not directly identify a particular person. Their function is to create a link to a specific computer to allow that computer access to the Internet. As a consequence, identification does not follow automatically through access to an IP address alone. For example, a computer in a house may be used by multiple members of the family. Not surprisingly then, some companies have argued that an IP address is non-PII.[114] Yet, this argument is misleading.[115] 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."[116] The incident gathered national attention when Seigenthaler wrote an editorial in *USA Today* condemning the defamation.[117]

As it turned out, Wikipedia had maintained the record of the IP address listed for the person who posted the contested information in the Seigenthaler biography.[118] A third party who read about the incident was able to obtain the IP

---

112 For a selection of these Websites, see http://ip-lookup.net/; http://network-tools.com/; http://whatismyipaddress.com/ip-lookup.

113 For a Website offering this information, see http://whatismyipaddress.com/ip-lookup.

114 Moreover, courts have agreed with this argument, see *Johnson v. Microsoft Corp.*, 2009 U.S. Dist. LEXIS 58174 (W.D. Wash. June 23, 2009); *Columbia Pictures Indus. v. Bunnell*, 2007 WL 2080419, at *3, n.10 (C.D. Cal. May 29, 2007); *Klimas v. Comcast Cable Comm., Inc.*, 465 F.3d 271, 276 n.2 (6th Cir. 2006).

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

116 John Seigenthaler, *A False Wikipedia "Biography,"* USA TODAY, Nov. 29, 2005, at http://www.usatoday.com/news/opinion/editorials/2005-11-29-wikipedia-edit_x.htm.

117 Katharine Q. Seelye, *A Little Sleuthing Unmasks Writer of Wikipedia Prank,* N.Y. TIMES, Dec. 11, 2005, at 15.

118 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.*

from Wikipedia and use IP lookup software to trace it to an address of a company in Nashville. The revealed hostname was for 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.[119]

Note, as well, that although Seigenthaler did not wish to file a lawsuit against the ISP to unmask the identity, so-called *John Doe* cases are now common.[120] Although caselaw is far from settled, ISP's generally require an entity seeking account information for an IP account to obtain a subpoena.[121] Courts tend grant these orders under a lenient standard; the party seeking the data must 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 who seek it. Of late, the Recording Industry Association of America (RIAA) has made active use of *John Doe* actions to unmask individuals who are engaged in a file-sharing of copyrighted works.[122] The revealing of IP account information has made legal actions possible against tens of thousands of patrons of sites such as bit-torrent.[123]

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."[124] Their paper provides the example of "an online consumer [who] visits websites, leaving the IP address of his computer logged at each site visited."[125] 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."[126] As the authors explain, "By 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."[127]

---

119 Seelye, *supra* note 117, at 15.

120 For a discussion of *John Doe* suits, see Patrick Fogarty, *Major Record Labels and the RIAA*, 9 HOUS. BUSINESS & TAX L.J. 140, 157-58 (2009); Julie E. Cohen, *Pervasively Distributed Copyright Enforcement*, 95 GEO. L.J. 1, 16-17 (2006).

121 Cohen, *supra* note 120, at 16.

122 Paul Roberts, *RIAA Sues 532 'John Does'*, PCWORLD, Jan. 21, 2004, at 15.

123 Casey J Dickinson, *Movie Industry Seeks Cornell Pirate*, 19 BUS. J. CENT. N.Y. issue 49, December 9, 2005. As Cohen notes: "Most defendants quickly settle for an amount reported to be in the $3,000-$6,000 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 120, at 17.

124 Bradley Malin, Latanya Sweeney & Elaine Newton, *Trail Re-identification,* Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory Technical Report, LIDAP-WP12. Pittsburgh: February 2003, at 1, at http://privacy.cs.cmu.edu/people/sweeney/trails1.pdf.

125 *Id.* at 2.

126 *Id.*

127 *Id.*

People may be surprised when linked to something they said anonymously, such as the person who wrote about Seigenthaler. They may be dismayed when a notice of a lawsuit arrives from the RIAA after they visited bit-torrent anonymously. They may be amazed that even scattered visits to Websites can lead to a linking of them to their IP address. These initial examples, all centered around IP addresses, demonstrate only some of the ways in which 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. AOL's release of search queries and research by Latanya Sweeney both demonstrate this point.

In 2006, America Online (AOL) released 20 million search queries for the benefit of researchers.[128] 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.[129] 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.[130]

AOL ultimately apologized for the disclosure. It recognized that it had violated the privacy of its users despite its attempts to anonymize the data.[131]

The AOL privacy debacle demonstrates a major problem with 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

---

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

129 *Id.*

130 *Id.*

131 Anick Jesdanun, *AOL: Breach of Privacy Was a Mistake,* WASH. POST, Aug. 7, 2006, at A1.

to identify 87% of individuals in the U.S.[132] 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 deidentified data since other personal characteristics, such as gender, date of birth, and ZIP code, often combine uniquely to identify individuals."[133] As a further example, during the 1970s, the U.S. government began to sell census data to marketers, and it supplied only addresses without names.[134] Marketing companies, however, were able to link names to addresses with data in telephone books and voter registration lists.[135]

A further problem with non-PII is that so much information about people is available. This phenomenon of data availability heightens the ability to trace non-PII to PII. This aspect of the re-personalization problem stems from a privacy problem that we will call "aggregation."[136] 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. A person who thinks she is anonymous at certain sites may provide explicitly identifying information as when completing a purchase.[137] Visit patterns can permit the use of an IP address to link de-identified data to names and addresses.[138] A further example involves a study of Netflix movie rentals by Arvind Narayanan and Vitaly Shmatikov, two computer scientists. The Narayanan-Shmatikov research demonstrated that at least some people in a supposedly anonymous dataset could be identified based on how they rated movies in a public available website.[139] This example is worth exploring in at least brief 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 these

---

132 Latanya Sweeney, *Uniqueness of Simple Demographics in the U.S. Population*, Laboratory for International Data Privacy Working Paper, LIDAP-WP4 (2000).

133 Latanya Sweeney, *Maintaining Patient Confidentiality When Sharing Medical Data Requires a Symbiotic Relationship Between Technology and Policy*. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, AIWP-WP344, May 1997, at 4-5, at http://privacy.cs.cmu.edu/dataprivacy/projects/law/aiwp.pdf.

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

135 *Id.*

136 DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 117-21 (2008).

137 *Id.* at 2.

138 *Id.*

139 Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets,* 2008 IEEE Symp. On Security and Privacy 111 (Feb. 5, 2008), *available at* http://arxiv.org/PS_cache/cs/pdf/0610/0610105v2.pdf.

individuals gave to films in the Internet Movie Database (IMDb), a popular site with information and ratings about movies.[140] 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.[141] As this study demonstrates, a single piece of non-PII does not exist alone. Rather, such data forms only part of a shifting landscape in which extensive information is available about almost every person. There are significant consequences of this rich tableau of available information.

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

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:

---

[140] http://www.imdb.com.

[141] Narayanan & Shmatikov, *Robust De-Anonymization, supra* note 139, 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." 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. LAW & TECH. – (forthcoming 2011).

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 and obtain his address.

In *Northwestern Memorial Hospital v. Ashcroft*, Judge Richard Posner aptly recognized that de-identified data can readily be re-identified.[142] The government had subpoenaed patient records of women who had partial birth abortions. The records were to be redacted so that the identities of the women would not be disclosed.[143] Despite the redaction, the court quashed the subpoena, concluding that de-identified patient records still violated the patients' right to privacy.[144] 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 and sex history, will put two and two together, "out" the 45 women, and thereby expose them to threats, humiliation, and obloquy.[145]

Through his concept of "skillful 'Googlers,'" Judge Posner has identified only one of the many powerful tools that now exist for retrieving de-identified information, analyzing it, and linking it to other information to re-personalize it.[146]

Posner's concern about unmasking information that is considered non-PII has been bolstered by research by computer scientists. For example, Sweeney notes that in many health care data sets, there will be unique data about people that can be used to de-identify them even when they are not identified in the data set. She argues, for example, 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 the patient."[147] In her view, medical data, stripped of identifying data such as names, addresses, phone numbers, and SSNs, is not really anonymized: "[T]he remaining data can be used to re-identify individuals by linking or matching the data to other databases or by looking at

---

142 362 F.3d 923 (7th Cir. 2004).
143 *Id.* at 932.
144 *Id.*
145 *Id.* at 929.
146 For an earlier court that recognized these same issues, see *Parkson v. Central DuPage Hospital,* 435 N.E.2d 140, 144 (Ill.App.1982).
147 Latanya Sweeney, *Maintaining Patient Confidentiality When Sharing Medical Data Requires a Symbiotic Relationship Between Technology and Policy*. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, AIWP-WP344, May 1997, at 5-6, *available at* http://privacy.cs.cmu.edu/dataprivacy/projects/law/aiwp.pdf.

unique characteristics found in the fields and records of the database itself."[148] In another study, Sweeney and co-author Bradley Malin demonstrate that "genomic data can often be re-identified in a distributed health environment."[149] Finally, as we have already noted, Ohm has brought the complexities of anonymization to the attention of the legal academy.[150]

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. Already in 1977, the Privacy Protection Study Commission observed generally of information technology, "A major problem created by the widespread adoption of computer and telecommunications technology to personal data recordkeeping is the inability to anticipate and control future use of information."[151] The Commission noted that systems were developed and then modified with an eye only to immediate needs, and not to long-range implications of the computerization of another area of record-keeping, which were, at any rate, difficult to predict.[152]

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. Specifically, new discoveries are constantly being made about combining data to reveal other data in ways that are surprising. For example, a recent study by Alessandro Acquisti and Ralph Gross demonstrates that people's Social Security Numbers (SSNs) can be predicted based on other pieces of data such as birth date and birth location.[153] Acquisti and Gross conclude: "We demonstrate that it is possible to predict, entirely from public data, narrow ranges of values wherein individual SSNs are likely to fall."[154] The implications of this study are dramatic; as Acquisti and Gross state, "Unless mitigating strategies are implemented, the predictability of SSNs exposes them to risks of identity theft on mass scales.[155]

In addition to new technological abilities that permit the re-identification of data, another important variable facilitates future re-identification – the

---

148 *Id.* at 1.

149 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. BIOMEDICAL INFORMATICS 179-192 (3/2004).

150 Ohm, *supra* note 5, at 1716-1731.

151 PRIVACY COMMISSION, TECHNOLOGY, *supra* note 18, at 26 (1977).

152 *Id.*

153 Alessandro Acquisti & Ralph Gross, *Predicting Social Security Numbers from Public Data,* 107 PNAS 975, 975 (2009).

154 *Id.*

155 *Id.*

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. To illustrate, we can consider the Facebook Beacon system and Google Buzz.

In 2007, Facebook introduced the Beacon online ad system, which tracked its users' online activities on third party websites. Without initial warning to Facebook users, this system shared collected information from the third party sites not only with Facebook, but with a user's Facebook friends.[156] Thus, activities such a purchase of 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 Facebook.[157] As a further example, Google introduced Buzz, its social networking platform, in 2010, and in a fashion that also led to a widespread proliferation of the personal information. Buzz permits users to share updates, comments, photographs, videos and other information through posts, or "buzzes."[158] As the FTC noted in its complaint against Google, however, "Without prior notice or the opportunity to consent, Gmail users were, in many instances, automatically set up with 'followers' (people following the user).[159] In additional, after enrolling in Buzz, Gmail users were automatically set up to 'follow' other users."[160]

In sum, whether re-identification of information can occur in the future depends on technology and corporate practices that permit the linking of de-identified data with already-identified data.[161] Moreover, as more pieces of already-identified data become available, it becomes easier to link it up to de-identified data since there will likely be more common data elements.

D. The Ability to Identify Depends on Context

In many cases, identification of PII as opposed to non-PII is complex because information does not readily fit into a single of these two categories. As noted above, identifiability is a complex concept because of the changing

---

156 Caroline McCarthy, Facebook's Zuckerburg: "We simply did a bad job" handling Beacon, Cnet (Dec. 5, 2007), at http://news.cnet.com/8301-13577_3-9829526-36.html?tag=mncol;txt

157 The social networking site also set up Beacon so these data were initially transmitted without a user being able to opt out from the program. In 2010, a federal judge approved a $9.5 million dollar settlement of a class action lawsuit concerning this matter, *Lane et al v. Facebook Inc*, No. 5:08-cv-03845-RS Settlement (N.D. Cal. Filed Aug. 12, 2008), at http://www.beaconclasssettlement.com/Files/SettlementAgreement.pdf.

158 Complaint at ¶ 7, In the Matter of Google Inc., FTC File. No. 102-3136 (FTC, March 30, 2011).

159 *Id.*

160 *Id.* at ¶8. The program automatically shared user information even shared if a Gmail user selected the "Nah, go to my inbox" choice from the initial Buzz screen. *Id.* Agreement Containing Consent Order at Section III, In the Matter of Google Inc., FTC File No. 102-3136, *available at* http://www.ftc.gov/os/caselist/1023136/110330googlebuzzagreeorder.pdf.

161 FTC, PROTECTING PRIVACY, *supra* note 2, at 37-38.

landscape of technology as well as social and corporate practices. As a further and related point, abstract determinations of whether a given piece information is PII are insufficient because the ability to identify information is context-driven.

Consider Internet search queries that are anonymized. A search query is the information that a person types into a search engine like Google when searching online.[162] In the abstract, if anonymized, search queries appear to be non-PII. In fact, AOL mistakenly believed such information was anonymous when it released its search query data. Yet, whether or not a search query is PII cannot be determined in the abstract. It depends upon the nature of the search in which the subject person had been engaged. If the only data is one search query for something general (such as a search for "poodles)" then identifying a user might be difficult. But if the user has engaged in a highly specific search, and more than one search, she becomes more identifiable. At some point, a search allows a person to be readily identifiable.

In *Gonzales v. Google,* the government sought to obtain from Google a sample of search queries users had made.[163] The court quashed the subpoena on privacy grounds. It 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.[164]

The district 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 de-masking 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 to fit into the binary system of much current privacy regulation is futile. Each instance depends upon the context and the specific things searched for as well as what other information is available about the person. The distinction between PII and non-PII is virtually impossible to make in the abstract, which is how it almost always is made when incorporated into privacy regulation. As Part IV demonstrates, this Article's concept of PII 2.0 responds to this situation by requiring context-based evaluations around a standard-based definition of PII.

---

[162] 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).

[163] 234 F.R.D. 674 (N.D. Cal. 2006).

[164] *Id.* at 689.

III. BEHAVIORAL MARKETING AND THE SURPRISING IRRELEVANCE OF PII AND PRIVACY LAW

The problems with the current approach to PII are most dramatically illustrated by looking at the burgeoning practice of behavioral marketing. This technique – sometimes referred to as targeted marketing – involves examining the behavior patterns of consumers to target advertisements to them. Public interest groups, scholars, and government regulatory agencies, such as the FTC, have examined this practice and raised objections to it on privacy grounds.[165] As we demonstrate in this Part, behavioral marketing occurs 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, targeting their audience by the demographics of those watching particular TV shows or reading particular periodicals. Today, companies direct offerings to specific consumers based on information collected about their specific characteristics, preferences, and behavior. As Don Peppers and Martha Rodgers defined the holy grail of modern advertising, it is "one-to-one marketing."[166] The result of such marketing is to create "advertising crafted to uniquely engage" each individual.[167] 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.

The key recent development has been, moreover, the ability of companies to engage in behavioral marketing without PII—-at least as this term is traditionally

---

[165] 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 & Kathryn Montgomery, *Alcohol Marketing in the Digital Age* (BMSG, May 2010); Jeffrey Chester & Kathryn Montgomery, *Interactive Food & Beverage Marketing* (BMSG, May 2007). The reports are posted online at http://www.digitalads.org/reports.php. For another NGO report, see Center for Digital Democracy, *Behavioral Targeting and the Online Assault on Personal Privacy* (March 2009), at www.democraticmedia.org/node/401.

The FTC reported on this topic in depth in 2009, FTC, SELF-REGULATORY PRINCIPLES FOR ONLINE BEHAVIORAL ADVERTISING (February 2009). As an example of international attention to the topic, see the EU's Article 29 Data Protection Working Party, *Opinion 2/2010 on online behavioural advertising* (June 22, 2010).

[166] DON PEPPERS ET AL., THE ONE TO ONE FIELDBOOK: THE COMPLETE TOOLKIT FOR IMPLEMENTING A 1-TO-1 MARKETING PROGRAM (1999); DON PEPPERS & MARTHA ROGERS, ENTERPRISE ONE TO ONE (1997).

[167] Jeff Chester & Kathryn Montgomery, *Interactive Food & Beverage Marketing: An Update* 2 (2008), *available at* http://www.digitalads.org/documents/NPLAN_digital_mktg_memo.pdf

understood. In this section, we trace the transformation from a past age of mass marketing to that of one-to-one marketing. We then explore modern information exchanges, and 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.[168] Such mass marketing was an international phenomenon: as part of his series of books about the Second Empire, Emile Zola devoted a brilliant novel, *Au Bonheur des Dames* (1883), to the events in a department store in Paris.[169] 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."[170] Advertisers and other "mass persuaders" during this period exploited broad patterns drawn from demographic data.[171] They sought to influence consumers within large demographical groups.

In contrast, contemporary behavioral marketing targets individuals by drawing on digital information about their past behavior as well as 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."[172] Miller was worried about the possibility of future "attempts at human manipulation" by organization that used computers to affect and shape their customers' behavior.[173] 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 Chester warned in 2007, "Advertisers are developing increasingly sophisticated technologies designed to track, analyze, and persuade us in the Internet era."[174] 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 favorite restaurant, recent purchases,

---

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

169 EMILE ZOLA, AU BONHEUR DES DAMES (Penguin Group 2001) (1883).

170 TUROW, *supra* note 168, at 180. In this sense, Turow also writes of a "democratization of shopping." *Id.* at 179.

171 For a popular early account of how advertisers enlisted social scientists in the 1950s, see VANCE PACKARD, THE HIDDEN PERSUADERS (1957).

172 MILLER, *supra* note 17, at 42.

173 *Id.* at 42, 58.

174 JEFF CHESTER, DIGITAL DESTINY 128 (2007).

favorite movies and TV shows.[175]

Individuals can now be tracked across different websites or digital media.[176] Moreover, online advertising networks follow people around the web.[177] In the 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.[178] In this tracking process, the advertising industry relies on diverse technology, such as basic cookies, "flash" cookies, and Web beacons.[179] Some technology, and in particular the beacon, which is also known as a "Web bug," permits real time observation of the user's activity on an Internet page, including where one's mouse moved and the information that one typed, such as a search query or a form that an individual filled out.[180] The cutting edge of this technology continues to advance, with some Internet Service Providers (ISPs) starting to engage in the controversial practices of deep-packet inspection.[181]

Marketers today engage in a pinpoint process that focuses on ever smaller groups of people.[182] Instead of companies selling ads for specific websites, advertisers now seek to buy access to people that fit a certain profile.[183] Behavioral marketing also depends on analytics to decide how to approach customers.[184] 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.[185] 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."[186] The idea is to take the information that entities have or to which they can gain access, and to convert it

---

175 Emily Steel & Julia Angwin, *On the Web's Cutting Edge, Anonymity in Name Only*, WALL ST. J., Aug. 4, 2010, at 1-2; Jessica E. Vascellaro, *Google Agonizes on Privacy as Ad World Vaults Ahead*, WALL ST. J., Aug. 10, 2010, at 3-4; Julia Angwin, *The Web's New Gold Mine: Your Secrets*, WALL ST. J., July 31, 2010, at 1.

176 Vascellaro, *Google Agonizes, supra* note 175*,* at 4-5.

177 *Id.*

178 Angwin, *Web's New Gold Mine, supra* note 175, at 1.

179 *Id.*

180 *Id.*

181 Electronic Privacy Information Center, *Deep Packet Inspection and Privacy*, available at http://epic.org/privacy/dpi/.

182 TUROW, *supra* note 168, at 1-3, 8.

183 *Id.* at 8.

184 THOMAS DAVENPORT & JEANNE G. HARRIS, COMPETING ON ANALYTICS 87-91 (2007).

185 *See* Thomas H. Davenport, *Competing on Analytics*, 84 HARV. BUS. REV. 98, 101, 104, 106-07 (v. 1, 2006).

186 DAVENPORT & HARRIS, *supra* note 184, at 7.

to actionable knowledge.[187] 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, "Analytics are now king."[188]

The information collected is packaged into profiles, which are sold on new kinds of "stock-market-like exchanges."[189] As an investigatory series in the *Wall Street Journal* observes, "Information 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."[190] Information about an individual's browsing habits sells for as little as a tenth of a cent online. All those slivers of a cent nonetheless add up; marketing online is a billion dollar industry and remains a growth field.[191]

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.[192] At the same time, and as a general matter when directed towards adults, advertising is an accepted and inescapable part of life. On occasion, Americans even look forward to it.

There are two core objections to behavioral advertising when directed towards adults. The first has to do with transparency, and the second with money. Regarding transparency, behavioral marketing takes place today in a multi-channel process in which individuals generally receive scant information about the data that organizations collect about them and how this 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."[193] The new kind of tracking largely takes place in the shadows, and Americans, not surprisingly, have responded with deep unease.[194] The instinct of many people is

---

[187] As Thomas Davenport and co-authors explain, "The analytic process makes knowledge from data." Thomas H. Davenport et al., *Data to Knowledge to Results*, 43 CAL. MGT. R. 117, 128 (2001).

[188] Michael Fertig, *Hire Great Guessers*, Harvard Business Review Website, at http://blogs.hbr.org/cs/2010/09/hire_great_guessers.html

[189] Angwin, *Web's New Gold Mine, supra* note 175, at 1.

[190] *Id.*

[191] According to one estimate, online advertising is a $23 billion a year industry. Interactive Advertising Bureau, *Internet Advertising Revenue Report* at 3 (2008), *available at* http://www.iab.net/media/file/IAB_PwC_2008_full_year.pdf.

[192] *See, e.g.* Julia Angwin & Emily Steel, *Web's Hot New Commodity: Privacy*, WALL ST. J., February 28, 2011, http://online.wsj.com/article/SB10001424052748703529004576160764037920274.html. ; Jeff Gelles, *When 'Behavioral Marketing' Turns Creepy*, THE PHILADELPHIA INQUIRER, Feb. 21, 2011, http://www.philly.com/philly/business/When_behavioral_marketing_turns_creepy.html.

[193] Angwin, *Web's New Gold Mine, supra* note 175, at 1.

[194] Joseph Turow et al., *Americans Reject Tailored Advertising and Three*

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.[195] Moreover, the very complexity of the marketing eco-system heightens the general ignorance of these corporate techniques, and reduces the value of the tools that some companies are making available to users.[196]

Regarding money, and as we have noted, marketing online is a billion dollar growth industry. Even more specifically, targeted advertisements command a considerable premium in the marketplace.[197] 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.[198] 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.[199] 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."[200] 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 it. This lack of knowledge places consumers at a profound disadvantage in negotiations, such as they may exist, with those who collect the information.[201] 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.

Activities that Enable It, SSRN ELIBRARY (Sept. 2009), at 1, http://ssrn.com/paper=1478214.; Gallup, *U.S. Internet Users Ready to Limit Online Tracking for Ads*, (2010), at 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.")

195 *See, e.g.*, Nicholas Carr, *Tracking is an Assault on Liberty, With Real Dangers*, WALL ST. J. (Aug. 6, 2010)("Personalization's evil twin in manipulation."); Angwin & Steel, *Web's Hot New Commodity*, *supra* note 192quoting former brand marketer at a credit company as to how "[p]eople feel targeted ads online are 'spooky'").

196 Available individual controls include an ability to opt out from some tracking, and to set preferences about the kinds of information that are collected. Google, Microsoft, and Yahoo are among the companies offering such privacy tools. Byron Acohido, *Google Chrome Will Join Other Browsers with Privacy Tools*, USA TODAY, Jan. 26, 2011, at A4.

197 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 165, at 15. 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.*

198 FTC, PROTECTING PRIVACY, *supra* note 2, at 37.

199 For a discussion, see Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2055, 2069-2075 (2004).

200 Pat Regnier, *The ID Theft Protection Racket*, MONEY MAGAZINE 112, 116 (Sept. 2005).

201 Schwartz, *Property, supra* note 199 at 2076-83.

2. Where's the PII (Adults)?

In behavioral marketing, companies generally do not track individuals through use of their names. Rather, they utilize software to build personal profiles that exclude this item, but that contain a wealth of details about the individual.[202] Typically, these firms associate these personal profiles with a single alphanumerical code placed on an individual's computer. In one reported case, for example, the file consisted of this string: 4c812db292272995e5416a323e79bd37.[203]

These codes are used to decide which advertisements people see as well as the kinds of products that are offered to them. Thus, Capital One Financial Corporation draws on [x+1], an ad network, to decide instantaneously the specific type of credit card to show first-time visitors to its website.[204] It uses the ad network's information about people to suggest products to individuals and to steer them to one card and not another.[205] As [x+1] argues, however, it does not gather the names of the individuals whose information it collects and analyzes.[206] Thus, behavioral marketing occurs without identifying, in the traditional sense, a specific individual.

While advertising networks may not know people's name, 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 seeming 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 out that they do not collect PII. The denial is ritualistic; companies are typically quoted the third-person and in this fashion: "They … say that they don't collect 'personally identifiable information.'"[207] One newspaper article even credited this view to marketers in general: "The 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 described in privacy policies."[208] Or, as the FTC quotes the conclusion of numerous parties from industry, "there is a reduced privacy interest in, and risk of harm from, non-PII."[209]

---

202 Angwin, *Web's New Gold Mine, supra* note 175, at 1.

203 *Id.*

204 Steel & Angwin, *Web's Cutting Edge, supra* note 175, at 2.

205 *Id.*

206 *Id.*

207 Steve Stecklow, *On the Web, Children Face Intensive Tracking*, WALL ST. J., Sept. 17, 2010, at 1.

208 Julia Angwin & Tom McGinty, *Sites Feed Personal Details to New Tracking Industry*, WALL ST. J., Aug. 10, 2010, at 1.

209 FTC, SELF-REGULATORY PRINCIPLES, *supra* note 165, at 31.

This defense points to a broader policy matter: industry may develop a strategy of compromise around a PII-based regulatory regime plus a narrow definition of PII. Currently, behavioral marketing is regulated only to a limited extent, but legal rules in this area may soon increase.[210] 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 Federal Trade Commission (FTC), which brings actions against companies that violate their own privacy policies.[211] In addition to the FTC's policing of the privacy promises of organizations to make sure that they are kept, it 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 as breaching its stated privacy policy and consequently as an "unfair or deceptive act" in the sense of the Federal Trade Commission Act of 1914.[212] Such FTC regulation is limited in scope: the agency merely ensures that companies live up to their promises, and companies need not promise much.[213] Moreover, studies have shown that few consumers read privacy policies, and those that do fail to understand them.[214] In fact, consumers commonly and falsely believe that a website with a posted "privacy policy" necessarily provides a positive level of substantive protection.[215]

In addition to the FTC's actions enforcing privacy promises, the agency has taken actions against companies that fail to provide adequate data security.[216] Such enforcement actions can occur even in the absence of a data breach, though

---

210 As the *N.Y. 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 Hansel, *The F.T.C. Talks Tough on Internet Privacy*, N.Y. TIMES, Feb. 12, 2009, at A17.

211 For an overview, see PRIVACY AND DATA SECURITY LAW DESKBOOK 16-01 (Lisa Sotto, ed., 2010)[hereinafter SOTTO, DESKBOOK]; SOLOVE & SCHWARTZ, IPL, *supra* note 45, at 776-86.

212 FTC Act, 15 U.S.C. §45.

213 For illustrative enforcement actions, see *In the Matter of Gateway Learning Corp.*, Docket No. C-4120 (FTC, Sept. 17, 2004); *In the Matter of Bonzi Software, Inc.*, Docket No. C-4126 (FTC, Oct. 7, 2004); *In the Matter of Vision I Props., LLC, d/b/a Cartmanager Int'l*, Docket No. C-4135 (FTC, Apr. 19, 2005).

214 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.

215 *Id.* at 47.

216 For illustrative FTC actions, see *In the Matter of Eli Lilly and Company*, Docket No. C-4047 (FTC, May 8, 2002); *In the Matter of ACRAnet, Inc.,* FTC File No. 092-3088 (FTC, Feb. 3, 2011); *In the Matter of Twitter, Inc.*, FTC File No. 092-3093 (FTC, June 24, 2010).

more typically the FTC acts only once a data spill occurs.[217] As part of these enforcement actions, as in its complaint against Eli Lilly in 2002, it will also seek to sanction a company for failing to train its employees about adequate data security practices.[218]

Finally, the FTC is also beginning to develop elements of a broader approach to privacy based on a "transparency" approach. Its policing of privacy notices and enforcement of adequate security already move in this direction. More broadly, however, the agency has begun to develop substantive notices of disclosure beyond its "broken promises" approach. Thus, in an enforcement action against Sears, which was settled in 2009, the FTC alleged that this company had engaged in an unfair practice by failing to adequately disclosure the extent of its tracking of customers who were paid to use a program that would record their Internet browsing.[219] The FTC acted even though Sears had provided users with a license agreement that, albeit with obscure language, had arguably informed users of the tracking.[220] The FTC charged that Sears's failure to provide adequate disclosure of the scope of the data collection was a deceptive act.[221] Its settlement order required Sears to provide clear and prominent disclosure of "the types of data the [software] will monitor, record, or transmit."[222]

The FTC's enforcement of transparency continued in 2010 with a settlement against EchoMetrix.[223] In that case, "parental controls" software led to the secret collection of data about children's computer activity and the feeding of the resulting database to marketers.[224] The FTC's theory of the case was that the disclosure of the tracking at stake in the case provided inadequate disclosure.[225] Finally, in its 2010 Report, *Protecting Consumer Privacy in an Era of Rapid Change*, the FTC explicitly emphasized the obligation of companies to increase the transparency of their data practices.[226]

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

---

217 For cases following upon a breach, see *United States v. American United Mortgage Co*., Docket No. 07C-7064 (N.D. Ill. 2007); *In the Matter of Superior Mortgage Corp*., Docket No. C-4153 (FTC, Dec. 16, 2005); *In the Matter of Sunbelt Lending Services, Inc*., Docket No. C-4129 (FTC, Jan. 7, 2005).

218 *Eli Lilly, supra* note 215, at ¶ 7.

219 Complaint at ¶¶ 13–14, In the Matter of Sears Holdings Management Corporation, File No. 082 3099 (FTC, Sept. 9, 2009).

220 *Id.* at ¶ 8.

221 *Id.* at ¶ 14.

222 Decision and Order at Section IA, In the Matter of Sears Holdings Management Corporation, File No. 082 3099.

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

224 Complaint for Permanent Injunction and Other Equitable Relief, Federal Trade Commission v. EchoMetrix, Inc., Civ. No. CV10-5516 at ¶¶ 10-14 (E.D.N.Y. 2010).

225 *Id.* at ¶¶ 16-18.

226 FTC, PROTECTING PRIVACY, *supra* note 2, at 69-78.

definition of PII. This strategy would present a kind of red herring to regulators. It would allow marketers to achieve their same goals with 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 still will be able to influence consumer behavior that falls outside the definition of PII-- and in ways that many would view as troublesome.

A single quotation, and one that we have already cited, concisely sums up this strategy; it occurred in a newspaper story about how U.S. websites are installing as many as one hundred tracking tools at a single time on the computers of people visiting their sites. The newspaper reported the response from the companies setting the tracking files: "The 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 described in privacy policies."[227] The strategy encapsulated in this quotation is two-prong; it proposes that: (1) as non-PII, the collection of information and the tracking carried out with it fall outside a PII-based regulatory regime; and (2) as long as 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 has demonstrated, marketing has changed greatly over the last century and is now conducted in highly sophisticated and potent ways. Marketing with personal information is focused not only on adults, however, but also on youth, a term that public health experts define to include children and adolescents. This group's large amount of disposable income and incompletely-developed tastes and interests make them appealing targets.

The issues of marketing to youth and the impact of legal definitions of PII raise distinct issues compared to marketing to adults and adult PII. American law generally views youth as deserving special protection. It is also highly concerned about food marketing to children and eager to act to assist parents in helping to make good choices. In short, policymakers consider youth, and especially children, as especially susceptible to advertising messages and such marketing as playing a role in influencing them to consume high-calorie and low-nutrient foods.[228] As a result, this Article will analyze marketing issues that concern them separately.

1. Digital Marketing and the "Net Generation"

In 1998, Donald Tapscott announced that "The Net Generation has

---

[227] Angwin & McGinty, *Sites Feed Personal Details, supra* note 208, at 1.

[228] For a summary of the available research, see INSTITUTE OF MEDICINE, FOOD MARKETING TO CHILDREN AND YOUTH: THREAT OR OPPORTUNITY? 226-318 (Washington, DC: National Academies Press, 2006) [hereinafter IOM, FOOD MARKETING].

arrived."[229] 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.[230] In addition, the commercialization of the Web and associated digital devices occurred all but simultaneously with their emergence. As Jeff Chester and Kathryn Montgomery observe, "The rapid growth of the Internet and the proliferation of digital media are fundamentally transforming how corporations do business with young people in the twenty-first century."[231] Corporations have actively sought to shape the experiences of minors with these new media.

Digital marketing may also be more intensively directed towards youth than adults. In September 2010, the *Wall Street Journal* reported that the fifty most popular websites aimed at children installed more tracking devices on personal computers than the top sites do for adults.[232] The fifty websites popular with minors placed 4,123 pieces of tracking technologies on the newspaper's test computers, which was more than thirty higher than the fifty most popular general audience U.S. websites.[233]

With enormous amounts of disposable income, young people will continue to be an attractive audience for marketers. Digital marketing now occurs around many kinds of products and services. The information collected about children is quite precise as to the data subject's characteristics, interests and pastimes, including categories such as race, pets, likelihood to post on the Internet, photography, "virtual worlds," concern about weight, and general location.[234] We will focus our discussion on the example of marketing activities that involve food products.

Over the past three decades, the extent of obesity among minors has risen dramatically throughout the U.S. In 2004, over one-third of children and adolescents in the U.S. were overweight or at risk of becoming overweight.[235] This number represents triple the rate in 1971 and double the rate in 1985.[236] A different study, one from 2005, raised the possibility that diet-related diseases will cause children in this country to be the first generation in the U.S. to have a shorter life span than their parents.[237] The stakes are high; as the Institute of Medicine has declared, "Prevention of obesity in children and youth should be a

---

229 DON TAPSCOTT, DIGITAL GENERATION 1 (1999).

230 *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, "and not just in video games." DON TAPSCOTT, GROWN UP DIGITAL 93 (2009).

231 Chester & Montgomery, *Interactive Marketing*, *supra* note 165, at 13.

232 Stecklow, *Children Face Tracking*, supra note 207, at 1.

233 *Id.*

234 *Id.;* Angwin, *Web's New Gold Mine, supra* note 175, at 1.

235 IOM, FOOD MARKETING, *supra* note 228; T.N. Robinson & J.R. Sirard, *Preventing Childhood Obesity*, 28 AM. J. PREVENTIVE MED. 194 (2005).

236 IOM, FOOD MARKETING, *supra* note 228, at 125.

237 S.J. Olshansky et al., *A Potential Decline in the Life Expectancy in the United States in the 21st Century*, 352 NEW ENG. J. MED. 1138-45 (2005).

national public health priority."[238]

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. In their report, the Institute of Medicine reached a concise conclusion: "marketing works."[239] In more detail, but to the same effect, a review in 2009 of the relevant psychological research into food marketing stated, "Youth 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."[240] Children and youth 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."[241] The net result? As a report by three psychologists summarizes: "Marketing practices that promote calorie-dense, nutrient poor food directly to children and adolescents present significant public health risks."[242]

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. Chester and Montgomery cite a corporate executive's explanation for his company's push away from traditional TV advertising into new forms of digital marketing: "the eyeballs have moved."[243] Food and beverage companies are now among the leaders of the new one-to-one digital marketing system.[244]

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.[245] The cutting edge of marketing now

---

238 IOM, FOOD MARKETING, *supra* note 228, at 5. 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, *Tobacco Funds Shrink as Obesity Fight Intensifies*, N.Y. TIMES, July 22, 2010, at B1.

239 IOM, FOOD MARKETING, *supra* note 228, at xiii; ELIZABETH MOORE, IT'S CHILD PLAY 1-2 (2006).

240 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 (no. 1 2009).

241 *Id.* at 224.

242 *Id.* at 211.

243 Chester & Montgomery, *Interactive Marketing*, *supra* note 165, at 13.

244 *Id.* at 61.

245 On viral marketing and other techniques, see Chester & Montgomery, *Interactive Marketing: An Update*, *supra* note 167, at 2. 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 (2011), at 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's or Oscar Mayer Lunchables, is placed within a digital entertainment property.

involves the use of "command centers" in which staff members of corporations interact with select consumers. As an example, Gatorade has developed a "mission control center" which tracks "sentiment analysis" in the social media ether.[246] In the center, Gatorade employees monitor social-media posts 24 hours a day.[247] Once someone mentions Gatorade in Twitter or other online media, the staff can weigh in and interact with consumer.[248]

As a final point, some empirical evidence suggests "an especially damaging potential role of targeted food marketing on at-risk minority youth."[249] 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."[250]

2. Where's the PII (Youth)?

The same basic issues concerning PII in the context of digital marketing arise for youth as well as for adults. Companies track young people through personal profiles that exclude names, but contain a wealth of details about the individual. As compared to such marketing to adults, the law responds differently when this practice is directed toward youth. While no specific federal statute regulates these practices for adults, the Children's Online Privacy Protection Act (COPPA) establishes certain rules for marketing to young children.[251] COPPA seeks to protect children under the age of 13, and it mandates that a covered website have a posted privacy policy and obtain parental consent when collecting, using, and disclosing children's information.[252] It also grants the FTC an enforcement role for its mandates.[253] The FTC has responded vigorously with sixteen enforcement actions, and over $6 million levied in fines collected pursuant to settlements.[254] In

---

Elizabeth S. Moore, Kaiser Family Foundation, *It's Child's Play: Advergaming and the Online Marketing of Food to Children* 4 (July 2006), *available at* http://www.kff.org/entmedia/upload/7536.pdf.

246 Valerie Bauerlein, *Gatorade's 'Mission': Sell to Teens*, WALL ST. J., Sept. 13, 2010 at 12.

247 *Id.*

248 *Id.*

249 Harris et al., *supra* note 240, at 245.

250 *Id.*; *see* INSTITUTE OF MEDICINE, PREVENTING CHILDHOOD OBESITY 58-61 (2005) (examining relevant information about socioeconomic and ethnic make-up of high risk groups for childhood obesity).

251 Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501-06 (1998).

252 15 U.S.C. § 6502(b)(1)(A)(ii).

253 15 U.S.C. § 6501(8)(F).

254 United States v. Playdom, FTC File No. 1023036 (May 12, 2011); United States v. Iconix Brand Group, Inc., FTC File No. 0923032 (Oct. 20, 2009); United States v. Sony BMG Music Entertainment, FTC File. No. 082-3071 (S.D.N.Y. Dec. 11, 2008); United States v. Industrious Kid, Inc,, Docket No. CV-08-0639 (N.D. Cal. Jan. 30, 2008); United States v. Xanga.com, Inc., Docket No. 06-CIV-6853 (S.D.N.Y. Sept. 7 2006); United States v. UMG Recordings, Inc., Docket No. CV-04-1050 (C.D. Cal. Feb. 18, 2004); In

May 2011, its most recent enforcement action, led to a settlement that included a $3 million fine against an operator of an online "virtual worlds."[255]

COPPA has several notable weaknesses. First, it only applies to children under 13.[256] Advertisers and marketers can therefore ply their trade with teenagers unaffected by the statute. Yet, teenagers may be even more vulnerable to targeted marketing than younger children.[257] Second, COPPA extends only to a "website or online service" and, thus, it does not regulate new digital platforms that are independent of the Internet, such as cell phones.[258] Third, COPPA regulates only a website or online service when it is "directed to children," or where the operator of the website "has actual knowledge that it is collecting personal information from a child."[259] It is relatively easy for website operators to avoid acquiring actual knowledge they are collecting information from a child.[260]

Even if these problems with COPPA are addressed, COPPA suffers from a

---

the Matter of Bonzi Software, Inc., FTC File No. 042-3016 (FTC, Feb. 18 2004); United States v. Hershey Foods Corp., Docket No. 4:CV-03-350 (M.D. Pa. Feb. 27, 2003); United States v. Mrs. Fields Famous Brands, Inc., Docket No. 2:03-CV205-JTG (D. Utah Feb. 27, 2003); United States v. The Ohio Art Co., FTC File No. 022-3028 (N.D. Ohio Feb. 22, 2002); United States v. American Popcorn Co., Docket No. C02-4008DEO (D.C.N.D. Iowa Feb. 14, 2002); United States v. Lisa Frank, Inc., Docket No. 01-1516-A (E.D. Va. Oct. 2, 2001); United States v. Looksmart, Ltd., Docket No. 01-606-A (E.D. Va. Apr. 19, 2001); United States v. Monarch Services, Inc., Docket No. AMD-01-CV-1165 (D. Md. Apr. 19, 2001); United States v. Bigmailbox.com, Inc., Docket No. 01-605-A (E.D. Va. Apr. 19, 2001); Federal Trade Comm'n v. Toysmart.com, LLC, Docket No. 00-11341-RGS (D. Mass. July 21, 2000). For a concise discussion of COPPA enforcement actions, see SOLOVE & SCHWARTZ, FUNDAMENTALS, *supra* note 4, at 110-11.

255 United States v. Playdom, FTC File No. 1023036 (May 12, 2011), at http://www.ftc.gov/opa/2011/05/playdom.shtm.

256 15 U.S.C. § 6501(1).

257 Harris et al., *supra* note 240, at 236. As the authors state, social science has shown that the impact of "[m]edia, including marketing messages" is especially strong for "older children and adolescents … as they focus on the world beyond their families and actively develop their independent identities." *Id.*

258 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 (February 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 27.

259 15 U.S.C. § 6502(b)(1)(A). In instances where a website has a special section for children, it would be subject to COPPA since this statute explicitly applies to "that portion of a commercial website or online service that is targeted to children." 15 U.S.C. § 6501(10)(A).

260 The result of the requirement, however, is that many websites that might otherwise fall under COPPA have a simple way to avoid its reach: the use of "drop-down" age menus. This result follows because it is not especially difficult for children to determine the appropriate birthday that will allow them to access a website.

fundamental flaw – its concept of PII. As this Article discussed in Part I.B.3, COPPA defines PII through the specific-types paradigm. But it employs this approach with a twist. In addition to the traditional list of types of PII ("first and last name," Social Security Number, and email address, and other elements), it provides an authorization for the FTC to add additional factors.[261] Although the FTC made limited use of this power in its COPPA Rule in 2000, adding a "a persistent identifier" used to track a person to the list,[262] the statute's key concept remains whether or not the "identifier" will permit "the physical or online contacting of a specific individual."[263] The meaning of the concept of "a 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 offered by the FTC's definition of "online contact information," which it views as involving "an e-mail address or any other substantially similar identifier that permits direct contact with a person online."[264]

IV. PII 2.0

The existing approaches to defining PII have proven problematic. Nonetheless, we reject the idea that privacy law should abandon the concept of PII. If the law did so, it would be left without a means for establishing coherent boundaries on necessary regulation. Next, we re-conceptualize the PII standard, compare its model of PII 2.0 to existing approaches in the U.S. and EU, and defend the new approach against possible objections. Finally, we apply its 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 reidentification 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 will never stop growing until it includes everything."[265] 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."[266]

In fairness to Ohm, his primary focus is not on abandoning PII, but on alerting the legal academy and policymakers to the problem of new means for re-

---

[261] 15 U.S.C. § 6501(2)(A)-(F).
[262] 16 C.F.R. § 312.2 (2011).
[263] 15 U.S.C. § 6501(2)(F).
[264] 16 C.F.R. § 312.2 (2011).
[265] Ohm, *supra* note 5, at 1742.
[266] *Id.*

identification of data.[267] This effort is a valuable and meritorious one. Nevertheless, he pushes his argument further when he 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."[268] Ohm calls for a clamping down on the flow of information through society. He would replace the current reliance on PII as a gatekeeper for privacy law with a cost-benefit analysis for *all* data processing and data collection of any kind.[269] Ohm proposes that privacy regulation "should weigh the benefits of unfettered information flow against the cost of privacy harms."[270]

Abandoning PII is problematic, however, because the concept serves a crucial function: it establishes the boundaries of privacy regulation. In a world overflowing with information, the law cannot possibly regulate all of it. 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. There would be no limits on the scope of privacy law. Moreover, Ohm's proposal to assess the costs and benefits of every collection and release of data will be tremendously difficult because all the costs and benefits are often not known in advance. Ohm suggests that when in doubt, the law should limit the release or even creation of large data sets.[271] Such data sets, however, play an important role in research, health care, data security, and in 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.[272] 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 an "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."[273] 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."[274]

---

267 *Id.* at 1704.
268 *Id.* at 1706.
269 *Id.* at 1765-68.
270 *Id.* at 1759.
271 *Id.* at 1766-68.
272 INSTITUTE OF MEDICINE, BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 112 (Sharyl J. Nass et al., eds., 2009).
273 *Id.*
274 *Id.*

These benefits have included the development of Herceptin, the use of an "open source" approach for Alzheimer's research, and medical database research involving children. As an initial example, through analysis of the records of a cohort of 9,000 breast cancer patients, scientists were able to identify the HER-2 oncogene.[275] Scientists then developed a targeted therapy, Herceptin, that is effective with women with HER-2 breast cancer. In another major research effort, one that started in 2003, universities, the drug and medical-imaging industries, and non-profit groups joined in a collaborative effort to find biological markers that reveal the progression of Alzheimer's disease in the human brain.[276] As the *New York Times* summarized, "The 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."[277] There have already been more than 3,200 downloads of the entire data set, and almost a million downloads of the database that contains images from brain scans.[278] As a final example, medical database research has improved children's health. The results include 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.[279]

Analytics also play an important role in data security. For example, a multi-institutional response is necessary to combat data security breaches.[280] 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.[281] 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.[282]

Analytics are used in creating new products and services for direct use by

---

[275] *Id.* at 114.

[276] The data itself is posted at ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE, http://adni.loni.ucla.edu/ (last visited March 4, 2011).

[277] Gina Koalta, *Sharing of Data Leads to Progress on Alzheimers*, N.Y. TIMES, Aug. 13, 2010, at A1.

[278] *Id.*

[279] In a more recent study that drew on database analysis, Flaura Winston and other researchers drew on "child-focused crash surveillance information" reported to the State Farm Insurance Companies in 15 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 (2000).

[280] *See* Paul M. Schwartz & Edward J. Janger, *Notification of Data Security Breaches*, 105 MICH. L. REV. 913 (2007).

[281] *Id.* at 962.

[282] 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 information. ID Analytics, Technology Overview, *ID Network*, at http://www.idanalytics.com/technology/index.php#id-network.

individuals. For example, Google Flu Trends is a free service that furthers early detection of influenza epidemics throughout the world.[283] 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 respiratory illnesses. [284] There is also growing concern about the possibility of a future pandemic with millions of possible fatalities worldwide if a new strain of influenza virus emerges.[285] 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 web search queries that millions of individuals submit to the Google search engine each day.[286]

Although analytics have great benefits, they can also implicate information privacy concerns. Yet, an approach where the first step is to restrict the flow of information is a move in the wrong direction. 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 his play *Wallensteins Lager* (1798): "*Was nicht verboten ist, ist erlaubt*" ("What is not forbidden is allowed.").[287] Information collection and processing is thus permitted unless a law forbids it. This approach wisely encourages the flow of information and the benefits it brings while setting up 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, the cost-benefit analysis would be 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 that may prove to be in vain.

Privacy rights should attach when data pertains 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. To be sure, certain types of aggregate data can be used in ways to harm people. 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 by insurance companies affects coverage and rate decisions. These decisions can cause harm to people, and these harms do involve information. Nonetheless, this category of harm is far broader than the category of information privacy harms. As a policy matter, these issues raise

---

283 Jeremy Ginsberg et al., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012 (Feb. 19, 2009).

284 FACT SHEET NO. 211, WORLD HEALTH ORGANIZATION, INFLUENZA (Seasonal) (Apr. 2009), *available at* http://www.who.int/mediacentre/factsheets/fs211/en/.

285 Ginsberg, *Detecting Influenza Epidemics*, *supra* note 283, at 1012.

286 *Id.* at 1014.

287 FRIEDRICH SCHILLER, WALLENSTEINS LAGER, Act 1, Scene 6 (1798).

questions that predominately sound in civil rights, discrimination, and insurance law.[288] At least as far as the analysis of the aggregate data is concerned, the critical issues are not those of information privacy law.

When data is disclosed or used, a privacy harm is created because the data pertains to specific people. Disclosing that ten thousand 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 privacy.[289] 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.[290] In our view, privacy law cannot abandon PII because doing so would make it impossible to establish coherent boundaries on the scope of necessary regulation. Instead, as we argue below, the concept of PII must be reconceptualized rather than jettisoned.

B. A STANDARD FOR PII

In devising an approach to conceptualizing 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.[291] The legal discussion of PII to this point has not yet considered the issue of whether PII ought to be a rule or a standard, but focusing on 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, lists the kinds of data that are PII. To categorize these three models within the rules-standard framework, the first two are standards, and the last one, a rule. The concept of rules and standards also provides a window into the grounds for understanding the current failure of all three approaches.

Due to their open-ended nature, the definitions of PII that we have characterized as the tautological and non-public approaches are standards.[292] They allow the decision-maker to take into account relevant factors and permit

---

288 On the role of antidiscrimination law, see TIMOTHY P. GLYNN, EMPLOYMENT LAW 515-576 (2007).

289 Whether or not the U.S. provides enough protections in privacy law for such information is, of course, another matter. For a discussion, see SOLOVE & SCHWARTZ, IPL, *supra* note 45, at 565-72.

290 Thus, the disclosure of a person's membership in an organization can affect freedom of association and the groups needed to affect social change. *NAACP v. State of Alabama*, 347 U.S. 449, 461-23 (1958).

291 *See* Part I.B., *supra*.

292 Isaac Ehrlich & Richard A. Posner, *An Economic Analysis of Legal Rulemaking*, 3 J. LEG. STUDIES 257, 258 (1974). ("A standard indicates the kinds of circumstances that are relevant to a decision on legality and is thus open-ended,").

broad discretion.[293] 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."[294] 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.[295] 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. The main problem in the face of any standard's inevitable generality has been the tendency in information privacy law to interpret a specific definition of PII as applying only to information that taken in isolation, or at that single moment, 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 reading 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 EU 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, or outdated, as in the COPPA Rule.[296] As Kathleen Sullivan already pointed out in 1991, one problem with rules is that they "tend toward obsolescence."[297] Indeed, while COPPA permits the FTC to add to the definition of PII, this authorization has languished unused since 2001.[298] Here, we can draw on an insight of Louis Kaplow, who noted that rules require the legal system to expend more work ex ante, and standards, to engage in more work ex post.[299] As Kaplow observed, "When the government promulgates a rule, it gathers information before individuals act and announces its findings."[300] The difficulty for rules is that the government entity designated to revise it might be unable to or unwilling to expend the necessary resources to do so.[301] As an illustration, the FTC has been

---

293 Sullivan, *supra* note 61, at 58-59 ("Standards allow the decision-maker to take into account all relevant factors or the totality of the circumstances.").

294 Video Privacy Protection Act of 1988, 18 U.S.C. § 2710(a)(3); Cable Communications Policy Act of 1984,, 47 USC § 551(a)(2)(A) (1994).

295 *See* Sullivan, *supra* note 61, at 58 ("A legal directive is 'standard' – like when it tends to collapse decision-making back into the direct application of the background principle or policy to a fact situation.").

296 Massachusetts Security Breach Law, 201 Mass. Code Regs. 17.00 *et. seq.*,; Children's Online Privacy Protection Act of 1998, 15 USC § § 6501-6506 (2006).

297 Sullivan, *supra* note 61, at 33.

298 COPPA Rule, 16 C.F.R., § 312.

299 Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 58-90 (1992).

300 *Id.* at 585.

301 *Id.* at 623.

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 re-conceptualization 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. Rules can become obsolete.[302] Indeed, rules function best when an area of social and technological development has reached a fairly settled status. Sullivan cogently observed that a rule reflects an area of "epistemological maturity."[303] The myriad routes that can lead to creation of PII do not fit into a set of neat categories. 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 concerns the heterogeneous nature of the behavior to be regulated. As this Article's Part III has demonstrated, the tracking of individuals and the behavior that can re-identify information are quite diverse. Numerous scholars, including Isaac Ehrlich and Richard Posner, have demonstrated that rules are quite poor at handling situations involving many different types of behavior that should be treated distinctly.[304] 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.

This pattern suggests that the best starting point for information privacy law, at least under present conditions, is through a standard for PII. The question then becomes the nature of this standard. In the next section, we consider two existing models: (1) the U.S. reductionist approach to PII, and (2) the E.U. expansionist approach.

C. REDUCTIONISM, EXPANSIONISM, AND PII 2.0

Information privacy law is now divided between reductionist and expansionist regulation of PII. The U.S. offers examples of the former, and the EU of the latter. Both approaches are flawed. In this section, we develop a different concept, which we term, "PII 2.0."

1. Reductionism in the U.S.

In the US, as we have seen, the law often engages in a reductionist reading of PII. The tendency manifests itself when statutes, judges, or policymakers consider PII to be only information that refers to a currently-identified person. Although computer scientists and data security experts in the U.S. recognize the category of identifiable information, the law has, by and large, failed to understand this concept. Identified information already refers to a specific

---

302 Sullivan, *supra* note 61, at 33.

303 *See id.* at 62.

304 Ehrlich & Posner, *supra* note 292, at 258.

person, and the concept of identifiability means that such a connection has not yet occurred, but is possible. To be sure, the second "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 this agency views the applicable statutory language and its own COPPA rule as regulating this technology only when there is information about an "identified" person.[305] 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 email.[306] When a company engages in the same gathering of information only to send the same person a targeted ad based on cookies placed on her computer, however, it is likely to fall outside the COPPA rule.

Another example of the reductionist tendency in the U.S. involves the Privacy Act's definition of a "system of records," which turns on whether federal agency records involve an "identified" person.[307] The Privacy Act does *not* apply to data processing if a person is "identifiable" within a federal agency's database, but not located through a unique identifier.

2. Expansionism in the EU

In comparison, the EU has an expansionist approach to PII. For example, the EU Data Protection Directive defines "personal data" as "information relating to an identified or identifiable natural person."[308] This accord sets out common rules for data protection in EU Member States and requires these countries to enact legislation that follows its standards.[309] Through this supranational agreement, a definition of PII as relating to "identifiable" individuals has been fixed deep in the DNA of EU information privacy law. The EU Data Protection 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."[310] 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

---

305 *See* Part I.B.3, *supra*.

306 FTC, COPPA REPORT, *supra* note 258, at 25.

307 Privacy Act of 1974, 5 U.S.C. § 552a (1976);*see* Part I.A., *supra*.

308 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, O.J. (L281) 31, art. 2(a) (1995) [hereinafter EU 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).

309 Schwartz, *European Data Protection Law*, *supra* note 308, at 484.

310 EU Data Protection Directive, *supra* note 308, at art. 2(a).

person."[311]

In the EU, moreover, information that refers to an "identifiable" person is treated in the same fashion as that which refers to an "identified." The treatment in privacy of "identified" and "identifiable" as equivalents is a German innovation. The Federal German Data Protection Act (*Bundesdatenschutzgesetz*, or BDSG) of 1977 defines "personal-specific" information as data relating to both "identified" and "identifiable" individuals.[312] Whether the data is "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 specific data" if "the reference person is identifiable."[313] He adds, "It is irrelevant for the BDSG's application whether the person is identified or identifiable."[314]

To be sure, the concern in EU law about the risks of "identifiable" data has proven prescient. Already in 1978, Dammann zeroed in on a threat that he called "re-individualization" (*Re-Individualisierung*) of data.[315] He observed, "Where the 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."[316] According to Dammann, the critical question concerning the nature of PII turns on the availability of "additional knowledge" (*Zusatzwissen*) about the concerned individual.[317]

The EU expansionist approach to PII is more in tune with technology than the U.S. reductionist approach. It also has been of significant international influence. Thus, 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.[318] The OECD is a group of leading industrial countries, including the U.S., and the OECD Guidelines provide a non-binding framework for its member nations.[319] It defines personal data as "any information relating to an identified or identifiable individual (data subject)."[320] The OECD Guidelines applies its eight privacy principles to all PII, and, in doing

---

311 *Id.* at Recital 26.

312 Gesetz zum Schutz vor Mißbrauch personenbezogener Daten bei der Datenverarbeitung (Bundesdatenschutzgesetz) [Federal Data Protection Act], Jan. 27, 1977, BGBl. I at 201, *last amended by* Gesetz, Aug. 22, 2006, BGBl. I at 1970.

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

314 *Id.* at § 3, marginal no. 23.

315 DAMMANN, KOMMENTAR ZUM BUNDESDATENSCHUTZGESETZ § 2, marginal no. 25 (Spiros Simitis, ed., 1st ed. 1978).

316 *Id.*

317 *Id.* at marginal no. 26.

318 Organization for Economic Cooperation and Development (OECD), Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, Doc. C(80)58 Final (Sep. 23, 1980).

319 For a discussion of the OECD Guidelines, see SOLOVE & SCHWARTZ, IPL,*supra* note 45, at 997-998.

320 OECD Guidelines, *supra* note 318, at 1(B).

so, demonstrates the EU expansionist approach.[321] Nonetheless, its influence on this question in the U.S. has been negligible as the US has followed its own reductionist path. In a fashion similar to the OECD Guidelines, the Privacy Framework of the Asian-Pacific Economic Cooperation of 2004 defines PII as "any information about an identifiable or identifiable individual."[322]

Finally, Canada reflects the influence of the EU approach, but goes even further in its approach to PII by dropping the concept of "identified." Its federal privacy law, the Personal Information Protection and Electronic Documents Act (PIPEDA), regulates the collection, use, and transfer of personal information by private organizations.[323] 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."[324] As a leading treatise on Canadian privacy law summarizes the result, "In essence, almost any information in any form that can be attributed to an identified individual is caught by this expansive definition."[325] The federal Privacy Commissioner plays a key role in deciding whether information is identifiable. The general tendency has been expansionist. As the Privacy Commissioner stated in his annual report to Parliament, 2001-2002, "The 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 about an identifiable individual."[326]

Notwithstanding its widespread adoption by other international documents, the EU expansionist approach is flawed because it treats data about an identifiable and identified person 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

---

321 *Id.* at §§7-14.

322 Asian-Pacific Economic Cooperation (APEC) Privacy Framework § 9 (2004).

323 Personal Information Protection and Electronic Documents Act, 2000 S.C., ch. 5 (Can.) at Pt. 1 § 26(b) [hereinafter PIPEDA]. PIPEDA also regulates the use of personal information by federal organizations and data flows between Canadian provinces. *Id.*

324 *Id.* at Pt. 1 § 2(1).

325 BARBARA MCISAAK ET AL., THE LAW OF PRIVACY IN CANADA 4-7 (2011); *See* 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 information.").

326 Office of the Privacy Commissioner of Canada, Annual Report to Canada 2001-2002, Part Two, Report on the Personal Information Protection and Electronic Documents Act, The Definition of Personal Information, at http://www.priv.gc.ca/information/ar/02_04_10_02_e.cfm.

For important caselaw interpreting this term in PIPEDA, see *Gordon v. Canada (Helath)*, 2008 FC 258 (CanLII); *Wyndowe v. Rousseau*, 2008 FCA 39 (CanLII).

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 relates 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 FIPs: (1) notice, access, and correction rights for the individual; and (2) security for personal data.[327] For information that merely relates to an *identifiable* person, the law should *not* generally require that the entity that processes information provide a full panoply of notice, access, and correction rights. Indeed, to do so might be counterproductive in many circumstances by requiring that the data first to be associated with an identified person.

3. The Benefits of PII 2.0

A benefit of having two categories of PII, identified and identifiable data, is to open the path for an assessment of the optimal nature of legal protections. Rather than a hard "on-off" switch, this approach allows legal safeguards for both identified and identifiable information, ones that permit tailored FIPs built around the different levels of risk to individuals. In our model of PII 2.0, information refers to (1) an identified, (2) identifiable, or (3) non-identifiable person. The continuum runs from actually being identified to no risk, and our three categories divide up this spectrum and provide three different regimes of regulation. Because these categories do not have hard boundaries and are fluid, 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 U.S., 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.[328] 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."[329] In German data protection law, as Dammann explains, "The person is identified when it is clear that the information refers to this person and not to

---

327 On these two FIPs, see Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609, 1614 (1999); FTC, PRIVACY ONLINE 7-12 (1998).

328 NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, GUIDE TO PROTECTING THE CONFIDENTIALITY OF PERSONALLY IDENTIFIABLE INFORMATION (PII) 2-1 (2010); GENERAL ACCOUNTING OFFICE, PRIVACY: ALTERNATIVES EXIST FOR ENHANCING PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION (May 2008); Office of Management & Budget, Memorandum 07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information* (2007).

329 Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concept of Personal Data* 12 (June, 20, 2007).

another."[330]

In the middle of the risk continuum, information refers to an *identifiable* individual when a specific identification, while possible, is not a significantly probable event. In other words, an individual is *identifiable* when there is some possibility of future identification, but it is not a remote one. The risk level is moderate to low. This information should be treated differently than an important sub-category of nominally identifiable information, where a linkage to a specific person has not yet been made, but where such a connection is more likely. As we shall explain shortly, 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 population of the U.S., China, and Japan, and their relative access to telecommunications.[331]

Practical tools also exist for assessing the risk of identification. In fact, computer scientists have developed metrics for assessing the risk of "identifiability." For example, Khaled El Eman has identified benchmarks for assessing the likelihood that de-identified information can be linked to a specific person, that is, made identifiable.[332] The critical axes in El Eman's work concern the "mitigating controls" that the party with the information places on it, and the likely motives and capacity of the outsiders who might to seek to tie the data to a person.[333] The decades spent by computer scientists in developing more secure software also offer useful lessons. The achievement of a "secure development lifecycle" requires computer scientists to assess on an ongoing basis: (1) the nature of internal and external threats to a data asset, and (2) the effectiveness of possible countermeasures.[334]

There are certain instances where identifiable information should be

---

330 DAMMANN, KOMMENTAR, 6TH ED., *supra* note 313, at § 3 marginal no. 21.

331 The CIA's World Factbook provides online access to such information, see CIA: World Facbook, at https://www.cia.gov/library/publications/the-world-factbook/.

332 Khaled El Eman, *Risk-Based De-Identification of Health Data*, IEEE SECURITY & PRIVACY 64 (May/June 2010); Khaled El Eman, *Heuristics for De-Identifying Data*, IEEE SECURITY & PRIVACY 58 (July/Aug. 2008).

333 El Eman, *Risk-Based, supra* note 304, at 65-66. For an important essay that summarizes current research on new methods for randomizing data sets with personal information, see Cynthia Dwork, *A Firm Foundation for Privacy Data Analysis*, 54 COMMUNICATIONS OF THE ACM 86 (2011).

334 MICHAEL HOWARD & STEVE LIPNER, THE SECURITY DEVELOPMENT LIFECYCLE (2006). Moreover, Adam Shostack and Andrew Stewart have proposed that data security analyze objective information about data breaches, draw on other fields, such as economics and psychology, and use the scientific method in testing hypothesizes. ADAM SHOSTACK & ANDREW STEWART, THE NEW SCHOOL OF INFORMATION SECURITY 145-50 (2008).

treated akin to 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 carry out the necessary linkage or linkages. This essential sub-category 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 on 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 future likely development of relevant technology, and the likely incentives of parties to link identifiable data to a specific person.[335]

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 of PII, the key is to think about identification in terms of risk. Our model of PII 2.0 conceives of identifiability as a continuum of risk rather than as an either-or. 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: (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.[336]

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 be different 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.[337] 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. Yet, 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?

---

335 Article 29 Working Group, *Opinion on personal data, supra* note 347, at 15.

336 Schwartz, *Preemption*, *supra* note 45, at 907.

337 On the development of privacy legislation in the U.S., the classic study remains REGAN, *supra* note 40, at 174-211.

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, the law's creation of such interests would decrease rather than increase privacy by requiring that all such data be associated with a specific person. This connection would be necessary to allow an individual to exercise her rights of notice, access, and correction. In this fashion, the law would promote a vicious circle of identifiable data being made identified. 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 cripple socially productive uses of analytics that did not raise significant risks of harms to individuals. [338]

At the same time, some FIPs should apply to identifiable data. The key obligations 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."[339] 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 for 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, an openness about information use allows for improved policies and law. As Louis Brandeis famously stated, "Sunlight is said to be the best disinfectants; electric light the most efficient policeman."[340] Brandeis was also concerned about privacy, of course, and this interest was reflected first in his famous 1890 article with Samuel Warren, and then in his opinions as Supreme Court Justice.[341] Yet, Brandeis' attention to privacy for individuals was accompanied by his interest in open flows of information about "social and industrial diseases." Characteristic is his argument about the need for "publicity as a remedy" in reducing abusive practices among financial institutions and bankers in the early twentieth century.[342] In an

---

338 At the Article 29 Working Party of the EU, there has been recent openness to a concept of "proportionality" in the use of information privacy law. Article 29 Data Protection Working Party, *Opinion 3/2010 on the Principle of Accountability* 3 (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.

339 SOTTO DESKBOOK, *supra* note 211, at 14-3.

340 LOUIS BRANDEIS, OTHER PEOPLE'S MONEY 92 (1914).

341 Warren & Brandeis, *supra* note 10, at 193. The most famous of his opinions about privacy as a Supreme Court Justice is his dissent in *Olmstead v. U.S.*, 277 U.S. 438, 478 (1928) (Brandeis, J., dissenting).

342 BRANDEIS, *supra* note 340, at 90-140.

analogous fashion, behavioral marketing and food marketing to children are controversial today, and there is a need for transparency about these emerging practices.[343]

Second, identifiable information can have great value. As we have discussed, "stock-market-like exchanges" now exist around information that is collected online.[344] 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. The requirement is one that depends, moreover, 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 things matter more than others, however, and the stakes are low in 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 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, 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.[345] When one clicks on Google Flu Trends, there is only high level information that is safely aggregated.

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 kind of 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. Access to such data should be accompanied by obligations that travel with the information. Companies that

---

343 *Id. at* 106. For more on Brandeis as a progressive advocate and his belief in public advocacy and in shaping opinion, see Neil M. Richards, *The Puzzle of Brandeis: Privacy and Speech*, 63 VAND. L. REV. 1295 (2010).

344 Angwin, *Web's Gold Mine, supra* note 175, at 1.

345 Ginsberg et al., *Detecting Influenza Epidemics, supra* note 283, at 1012-14.

handle identifiable information can structure these obligations by associating metadata, or information about information, with data sets.[346]

Thus, one benefit of PII 2.0 is to tailor 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 on keeping data in the most de-identifiable state that is practicable. As an illustration of such a disincentive in action, the EU's Article 29 Group, an independent advisory body on privacy, has articulated "an absolute certainty test" for ISP's and search engine operators.[347] Under it, 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."[348] 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. The payoff here is that the company by making information identifiable or non-identifiable will benefit from FIPs that become easier for it to meet as it moves 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 because the risk of identification can never be eliminated. From the EU's perspective, the problem is that treating "identifiable" 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." Potential PII is everywhere, and attempts to predict where it will appear, or in his metaphor, "pop right up," are pointless.[349] 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

---

346 Regarding metadata, see *Schwartz, Property*, *supra* note 199, at 2077.

347 Article 29 Working Party, *Opinion 1/2008 on data protection issues relating to search engines* 8 (April 4, 2008); Article 29 Working Party, *Opinion 4/2007 on the concept of personal data* 17 (June 20, 2007).

348 Article 29 Working Party, *Opinion 1/2008 on data protection issues relating to search engines* 8 (April 4, 2008).

349 Ohm, *supra* note 5, at 1742.

notification law.

In tort law, the concept of "reasonable" functions as a way for juries to sift through an otherwise unordered universe of "facts" that are of possible relevance each time an accident occurs.[350] 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.[351] To shift from the common law to a modern statutory regulation of a high tech issue, we can consider data breach notification laws, which now have been enacted in forty-four states.[352] These statutes typically require notification of an individual when evidence exists for a reasonable belief that an outside party has gained access to or acquired personal data.[353] These laws do not require a showing that a third party actually acquired the information, that is, gained control of it.[354] This standard has led to the development of contextual benchmarks regarding relevant indices of "access or acquisition" of information.[355] No more is required for PII 2.0; here, too, there is a need for developing 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."[356] 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.[357] Nonetheless, we do think that the sheer rate of technological change in this area counsels introduction of a "track and audit" approach, as set out above.

---

350 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).

351 For an introduction to the variable factors in concepts of reasonable and unreasonable behavior in negligence determinations, see RICHARD EPSTEIN, TORTS 169-284 (9th ed. 2008).

352 SOLOVE & SCHWARTZ, FUNDAMENTALS, *supra* note 4, at 136-38. For sample laws, see Cal. Civ. Code § 1798.81.5 (2006) and 201 Mass. Code Regs. § 17.00 *et seq* (2010).

353 SOLOVE & SCHWARTZ, FUNDAMENTALS, *supra* note 4, at 136.

354 *Id.*

355 As an example of such benchmarks, see California Office of Privacy Protection, *Recommended Practices on Notice of Security Breaches Involving Personal Information* 12-13 (2009).

356 Ohm, *supra* note 5, at 1711-1712.

357 For examples of the different attempts to develop effect, strong statistically-based methods of de-identification, see Dwork, *supra* note 333, at 86-91; Arvind Narayanan & Vitaly Shmatikov, *Privacy and Security: Myths and Fallacies of "Personally Identifiable Information,"* 53 COMM. OF THE 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 "the opportunities and motivations for the hypothetical adversary, see Yakowitz, Data Commons, *supra* note 141, at 22, 35-37.

The EU 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 of the EU, the goal must be to avoid "unduly restricting the interpretation of the concept of personal data."[358] The fear is that any other definition would narrow the jurisdictional sweep of the law. Nonetheless, the Article 29 Working Party has also conceded, "The scope of the data protection rules should not be overstretched."[359] Nonetheless, its first step is to claim as much information as possible to be "personal data" as demonstrated, for example, in its "absolute certainty test." Only then does it concede 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."[360] The evidence is at best mixed regarding whether such flexibility has, in fact, been forthcoming.[361]

In PII 2.0, in contrast, flexibility follows through 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 to avoid regulation and responsibility. PII 2.0 addresses the EU concern that regulation might be skirted by drawing the boundaries too narrowly, because we propose no hard line between identified and identifiable data and because our 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, which are behavioral marketing to adults and food marketing to youth. Regarding the former, PII 2.0 leads to a contextual analysis of the data used in behavioral marketing. In many instances, the information now

---

358 Article 29 Party, *Opinion on Personal Data*, *supra* note 347, at 5.

359 *Id.*

360 *Id.*

361 On the Article 29 Working Party's sweeping definition of PII in the use of Radio Frequency ID tags and highly detailed follow up requirements for Privacy Impact Assessments for all uses of RFID, see *Opinion 9/2011 on the revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications*, WP 180 (Feb. 11, 2011); *Opinion 5/2010 on the Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications*, WP 175 (July 13, 2010); *Working Document on the Protection of Individuals with Regard to the Processing of Personal Data, Working Document on data protection issues related to RFID technology*, WP 105 (Jan. 19, 2005). There have been complaints in the EU 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,* 2 INT'L J. L. & INF. TECH. - (forthcoming 2011).

being gathered is in fact identified and 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 far less likely a possibility. The PII 2.0 model also will promote heightened disclosure of commercial practices and demonstrate limits in the FTC's current approach and in 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 matter for 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.[362] These efforts involve the use of tracking files being placed on a user's computer and, in some instances, include the sale of 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 a name, address, or SSN.

The information at the heart of targeted marketing is not *non-identified* data. Indeed, the promise of these new forms of marketing is to go beyond advertising's past reliance on crude demographical categories and be able 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 individual. 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, correlation with static

362 *See* Part III, *supra.*

IP addresses, or links with explicitly identifying information at other websites.[363] 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 to reduce the risk of identification of personal data. The goal would be to structure data operations so that identification of specific individuals becomes truly remote, which will then lower the risk of data collection and processing.

Beyond the benefits of PII 2.0 leading to contextual determination regarding identified and identifiable information, we think that this approach suggests four insights about the current privacy law landscape. First, when behavioral marketing carries a significant risk of identification of specific individuals, there is a need for the same kinds of heightened disclosure of a company's practices as in other circumstances involving the collection of personal data. Since 2009, the FTC has been developing a jurisprudence of "transparency" that finds deceptive, and hence legally actionable, a company's failure to adequately disclosure its processing practices.[364] Emerging 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).[365] The absence of such transparency should be viewed as falling under the FTC's enforcement of unfair and deceptive practices.[366]

Second, PII 2.0 demonstrates certain limits in the FTC's approach and weaknesses of the current legal regime. PII 2.0 will likely lead to a classification of at least some behavioral marketing as involving identified information. Moreover, traditional privacy FIPs extend far beyond providing only transparency to consumers.[367] 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."[368] The lack of a general online privacy statute, or a specific behavioral marketing statute leaves questionable practices today free of effective regulation. The new classifications of PII 2.0 thereby provide support for the idea of additional sectoral privacy laws, a number of which have been introduced in Congress.[369]

---

363 Angwin, *Web's New Gold Mine*, *supra* note 175, at 1; Part II, *supra*.

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

365 In the Matter of Sears Holdings Management Corp., FTC File No. 082 3099 (Federal Trade Commission Sept. 9, 2009); Federal Trade Comm'n v. EchoMetrix, Inc., Docket No. CV10-5516 (E.D.N.Y. 2010); FTC, PROTECTING PRIVACY, *supra* note 2, at 41.

366 FTC, PROTECTING PRIVACY, *supra* note 2, at 12-13.

367 Schwartz, *Preemption, supra* note 45, at 907-908.

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

369 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 co-sponsoring. Julia Angwin, *Proposed Bill Would Put Curbs on Data Gathering*, WALL ST. J., March 10, 2011, at B1.

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

As for the possibility of a general privacy statute in the U.S., the leading candidate at present is "the Commercial Bill of Rights Act of 2011," which Senators John Kerry and John McCain have co-sponsored.[370] The Bill employs the specific-types approach, but its list of PII is extremely broad, including a catch-all category of "[a]ny other information concerning an individual that may be reasonably be used by the party using, collecting, or storing that information to identify that individual."[371] This Bill begins to resemble the EU expansionist approach.

Fourth, as regards *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.[372] 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 policy proposal**,** the Cabinet Office in the United Kingdom is leading a consumer empowerment effort that includes its "mydata" initiative.[373] The goal is to enable consumers greater knowledge of how organization's 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."[374] As we have already argued, such increased transparency will go far to correcting the asymmetry of knowledge between consumers and the companies that track their online behavior.[375]

---

370 The Commercial Privacy Bill of Rights Act of 2011, S. -, 112th Cong. (2011).

371 *Id.* at §3(5)(A).

372 *See* Part III.A.2, *supra.*

373 Cabinet Office, Department for Business Innovation & Skills, Better Choices: Better Deals (2011).

374 *Id.* at 17.

375 Schwartz, *Property, supra* note 199, at 2076-80. The treatment of both identified and identifiable information alike will heighten consumer awareness of behavioral marketing at a critical moment. An introduction of a concept of PII 2.0 occurs at time when consumers know little about behavioral marketing, but are also predisposed to be skeptical towards it.

Regarding the skepticism that Americans have towards industry practices, a 2010 survey by Joseph Turow and associates revealed that a majority of Americans do *not* want marketers to tailor advertisements to their interests. Turow et. al, *supra* note 1, at 1-4. The results of the Turow study suggest that greater transparency in this area will promote

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 24 hours a day.[376] 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 the 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 towards children under age 13, COPPA should fully apply involving identified information, including those where there is a significant risk of identification. In enacting COPPA in 1988, Congress was concerned with providing a mechanism for parental consent before the collection of personal information on the Internet.[377] Consistent with Congress' intention, PII 2.0 would update this policy concern and apply it to one 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, and now will merely summarize the statute's weaknesses. First, the Act only applies to children under 13.[378] 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.[379] These restrictions, and the limited vision of technology that they imply, means that COPPA-- like Grunge music and Beanie Babies-- remains entrenched in the 1990s.

---

greater options for consumers and more informed choice.

376 Valerie Bauerlein, *Gatorade's 'Mission': Sell to Teens*, WALL ST. J., Sept. 13, 2010 at 12.

377 *Children's Online Privacy Protection Act of 1998: Hearing on S. 2326 Before the Subcomm. on Communications of the S. Comm. on Commerce, Sci., and Transp.*, 105th Cong. 2 (1998) (opening statement of Hon. Conrad Burns, Senator from Montana).

378 15 U.S.C. § 6501(1).

379 *Id.* at §§ 6501(2)(A), 6502(a)(1).

Due to these limitations on COPPA, 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 found a high level of concern about this topic among young people.[380] Large majorities of young people also believe that a person should have legal rights to know the information that websites have about her and to require them to delete all such stored information.[381]

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 controls" software to collect data about children's computer activity and to feed it to marketers.[382] The FTC did not bring an enforcement action under COPPA; its theory of the case was one of "inadequate disclosure" by the company— a theory that it advanced although the company supplied language to its customers that arguably covered the underlying activity.[383] 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.

Thus, one part of a response to food marketing to children should rely on a transparency approach. The need is for greater information 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, multi-pronged public policy effort is now being directed toward this public policy issue. The highest profile participant in the debate is First Lady Michelle Obama, who is directing the

---

380 Chris Hoofnagle et al., *How Different are Young Adults from Older Adults When It Comes to Information Privacy Attitudes & Policies* (April 14, 2010) at http://www.ftc.gov/os/comments/privacyroundtable/544506-00125.pdf.

381 *Id.* at 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.

382 Complaint at ¶ 8-14, Federal Trade Comm'n v. EchoMetrix, Inc., Docket No. CV10-5516, (E.D. N.Y. 2010).

383 *Id.* at ¶ 12. The critical information was both buried in a Terms of Service notice and obscure in its phrasing. *Id.*

nation's attention to the many dimensions of this public health crisis.[384] At the federal interagency level, a working group is developing nutrition principles to guide industry when it marketes foods to children ages 12-17 years old.[385] The FTC, Food and Drug Administration, Centers for Disease Control and Prevention, and United States Department of Agriculture (USDA) are the agencies involved in this effort "to improve the nutritional profile of foods marketed to children."[386]

CONCLUSION

Personally identifiable information (PII) is one of the central concepts in information privacy regulation. The basic assumption behind the relevant statutes is that their applicability will turn on whether PII is present. 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.

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, but that associates different legal interests with each category. This flexible approach also provides the safeguard of treating identifiable information with a substantial risk of being identified as a form of identified data. Such an approach has the merit as well of preventing tactical attempts to use readily identifiable data in lieu of identified data to avoid regulation and responsibility.

PII 2.0 represents a way beyond the reductionist reading of PII in the U.S., and the expansionist reading in the EU. Its use would represent a significant step forward in responding to the privacy implications of behavioral marketing and the marketing of unhealthy food products to youth. In this Article, we have argued that PII cannot be abandoned, and that this concept is essential as a way to define 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 the response to the negative implications of the food industry's marketing techniques.

---

384 Sheryl Gay Stolberg, *Restaurant Nutrition Draws Focus of First Lady*, N.Y. TIMES, Feb. 6, 2011, at A7.

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

386 *Id.* at 1. As the *N.Y. Times* summarized the new guidelines, "Regulators 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, April 28, 2011, at B1. Public health advocates have also developed a sound methodology of possible regulatory approaches. In one of the most useful methodologies, developed by the Berkeley Media Studies Group, policy strategies are targeted along the concepts of the "four P's," namely, products, places, promotions, and price. BERKELEY MEDIA STUDIES GROUP, FIGHTING JUNK FOOD MARKETING FOR KIDS 18 (2006).

Appendix A: *Google Books Ngram Viewer*, References to "Information Privacy" 1950 – 2000