GOODBYE PII: CONTEXTUAL REGULATIONS FOR ONLINE BEHAVIORAL TARGETING

Yuen Yi Chung*
I. INTRODUCTION

Internet privacy has become a highly contentious issue.\(^1\) One after another, prominent news media have exposed numerous tricks in tapping consumer data to create personalized advertisements.\(^2\) Shortly after the 2012 elections, consumers grew increasingly concerned when it was revealed that even politicians used behavioral advertising for their campaigns.\(^3\) Despite the Federal Trade Commission’s

---


\(^{3}\) See Emily Steel, *A Web Pioneer Profiles Users by Name*, WALL STREET JOURNAL (Oct. 25, 2010), archived at www.perma.cc/T2FD-HJPW (disclosing RedLeaf’s ability to identify and target individuals with political advertisements via its massive database).
(“FTC”) best efforts in balancing the potential benefits of behavioral advertising and privacy concerns by setting self-regulatory principles, there is currently no law in the United States that expressly addresses behavioral targeting.4

Current privacy regulations center on Personally Identifiable Information (“PII”).5 PII defines the scope and boundaries of many federal and state privacy laws.6 PII serves as a jurisdictional trigger for these statutes and regulations; without PII, there is no privacy harm.7 Research has, however, demonstrated that non-PII may potentially turn into PII when additional information is made public or when data is aggregated.8

This note suggests that the Legislature should no longer measure privacy risks based on the distinction between PII and non-PII. Privacy laws should abandon the concept of PII and regulate behavioral targeting based on a contextual continuum of reasonable expectations.9 Part II describes the evolution of the role of PII in privacy laws.10 Part III provides a background of online behavioral targeting in relation to PII and illustrates the flawed concept of anonymity with

---

4 See Andrea Stein Fuelleman, Right of Publicity: Is Behavioral Targeting Violating the Right to Control Your Identity Online?, 10 J. MARSHALL REV. PROP. L. 811, 815 (2011) (indicating the lack of privacy law in the area of online behavioral targeting notwithstanding FTC’s effort in regulating behavioral targeting).

5 See Paul M. Schwartz & Daniel J. Solove, PII 2.0: Privacy and a New Approach to Personal Information, BUREAU OF NATIONAL AFFAIRS (Nov. 23, 2012), archived at www.perma.cc/EB5T-PM3Y (stating PII is “one of the most central concepts in privacy regulation”).

6 See id. (illuminating the role of PII in privacy laws).

7 See id. (stressing the particular importance of PII in bringing a privacy violation claim).

8 See Arvind Narayanan & Vitaly Shmatikov, Robust De-anonymization of Large Sparse Datasets, THE UNIVERSITY OF TEXAS AT AUSTIN (Feb. 5, 2008), archived at www.perma.cc/PJ5J-7GLP (identifying some people in an alleged anonymous data sample conducted by Netflix); see also Latanya Sweeney, Simple Demographics Often Identify People Uniquely, CARNEGIE MELLON UNIVERSITY LABORATORY FOR INT’L DATA PRIVACY (2000), archived at www.perma.cc/ZNW3-2A99 (discussing likelihood of uniquely identifying individuals from basic information).

9 See infra at Part V (concluding that the privacy laws should abandon the concept of PII’s and begin the use of reasonable expectations).

10 See infra Part II (providing history of privacy laws as related to PII).
three distinct cases. Part III also introduces European privacy law on behavioral advertising as well as several proposals for regulatory reforms in the U.S. Part IV proposes a new approach to contextual regulations for behavioral targeting and possible challenges.

II. THE ROLE OF PII IN U.S PRIVACY REGULATIONS

In the last century, PII has evolved from an irrelevant issue, to a recognized privacy tort, and then to one of the most fundamental aspects of the current privacy statutory schemes in the United States. Given its importance, there is surprisingly no uniform definition of PII.

A. The Rise of Privacy Law

In their much celebrated law review article, Samuel Warren and Louis Brandeis advocated a right of privacy in 1890. Alarmed by tabloid journalism, Warren and Brandeis conceived the right of privacy as a “right of personality” and described privacy deprivation as a form of mental suffering. After seventy years of privacy common law development, William Prosser categorized privacy law into four privacy torts commonly recognized in common law: (1) intrusion upon the plaintiff’s seclusion or solitude, or into his private affairs, (2) public disclosure of embarrassing private facts about the plaintiff, (3) publicity that places the plaintiff in a false light in the public eye, and (4) appropriation. Prosser did not explore the idea of PII because

11 See infra Part III (illustrating the failures of the current state of the law).
12 See infra at Part III (introducing European law on privacy issues).
13 See infra at Part VI (suggesting a workable alternative).
14 See infra at Part II (discussing PII).
15 See infra at Part II.C (discussing the evolving definition of PII).
16 See Samuel D. Warren & Louis D. Brandeis, The Right to Privacy, 4 HARV. L. REV. 193, 193-195 (1890) (proposing a right of privacy as a type of tort by pointing out the conflicts between technology and private life).
17 See id. at 195 (highlighting the harm from publicity given to sensitive information and personal life by comparing the right of privacy to the law of defamation).
his four distinct types of violations require actual injury of an identified person - as do all torts.\textsuperscript{19}

\textbf{B. Harm Prevention and Significance of PII}

PII first became an issue when the advent of the mainframe computer changed how information could be collected and processed.\textsuperscript{20} In the 1960’s, public bureaucracies began to computerize citizen records.\textsuperscript{21} The public became concerned of such actions because the compilations of data lead to easily accessible, massive databases that offered little protection of sensitive information.\textsuperscript{22} In response to the growing privacy concerns, the Secretary of Health, Education and Welfare introduced the Fair Information Principles (FIPS) in 1973.\textsuperscript{23} FIPS is a data protection framework that requires, among other principles, notice and consent, access, data integrity, enforcement and remedies.\textsuperscript{24} More significantly, FIPS recognizes the “creation of risk that a person might be harmed in the future”.\textsuperscript{25} Not only does this include permanent harm such as identity theft, but it also covers potential embarrassment and damaged reputation from misuse of information.\textsuperscript{26} FIPS has inspired legislation to shift from merely redressing past harm to avoiding privacy problems.\textsuperscript{27}

\textsuperscript{19} See id. at 392-98 (providing examples of privacy violations that all required injury to identified persons).
\textsuperscript{21} See PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY 72-73 (1995) (describing the U.S governments as one of the biggest users of computerized systems for personal data records).
\textsuperscript{22} See id. at 82 (providing a background on the legislation in privacy law).
\textsuperscript{24} See id. (asserting that FIPS is not only remedial in nature).
\textsuperscript{25} See id. (listing different forms of damages from leaked PII); see also Daniel J. Solove, A Taxonomy of Privacy, 154 U. PA. L. REV. 477, 487-88 (2006) (establishing a general proposition regarding the potential harm the lack of privacy may place on a person).
\textsuperscript{26} See Fighting Back Against Identity Theft, FEDERAL TRADE COMMISSION (2012), archived at www.perma.cc/BSF4-Y8XD (cautioning that identity theft is serious because some consumers not only suffer from damaged reputations, loss of educa-
After 1970, Congress began to enact privacy laws that were preventive in nature. This process required the legislature to first identify a problem, and then categorize the types of information that might contribute to that risk. This data-centric assessment of whether or not a particular data category constitutes “sufficient” harm to be regulated marked the beginning of the PII-era. To this day, Congress continues to develop various privacy laws around the concept of PII.

C. Three Approaches of Defining PII

Despite its significant role in privacy law, there is surprisingly no consistent definition of PII. While some laws and regulations view PII as a rule, others favor PII as a standard. Paul Schwartz and Da-
vid Solove have synthesized three approaches to defining PII in various privacy laws and regulations: (1) the tautological approach, (2) the non-public approach, and (3) the specific-types approach.34

1. The Tautological Approach

The tautological approach defines PII as any information that identifies a person.35 Any information that identifies a person is PII and triggers protection of the right of privacy.36 While this approach allows flexibility and evolvement, like all standards, it fails to define PII because it merely reiterates PII as PII.37

2. The Non-Public Approach

The non-public approach is a variant standard of the tautological approach.38 Instead of defining what PII is, privacy standards under this approach outline what is not PII—information that is either publicly accessible or purely statistical.39 The Gramm-Leach-Bliley Act,40 for example, simply defines personally identifiable financial information as “nonpublic personal information.”41 This approach is problematic because it fails to take into account whether such information is identifiable and overlooks the possibility that other non-
public information may readily be matched to this type of public information.\(^{42}\)

### 3. The Specific-Types Approach

The specific-types approach exemplifies the qualities of a classic rule – if information falls into an enumerated category, it automatically triggers the privacy law or regulation.\(^{43}\) The Children’s Online Privacy Protection Act of 1998\(^ {44}\) is an example of the specific-type approach in defining PII.\(^ {45}\) The federal statute defines PII as “individually identifiable information about an individual collected online”, such as first and last names, address, social security number, telephone number and email address, and “any other identifier that the [FTC] determines permits the physical or online contacting of a specific individual.”\(^ {46}\) Though clearer than other approaches, the specific-type approach is very restrictive in its definition of PII as it always carries the possibility of being under inclusive.\(^ {47}\) In addition, the list of identifiable information is not static because technology continues to grow and non-PII always has the potential to become personally identifiable.\(^ {48}\)

Despite its fundamental role in privacy regulations, there appears to be no uniform definition and application of PII.\(^ {49}\) All three of the current approaches are flawed and offer no concrete guidance as to what type of information belongs to the list of PII.\(^ {50}\) As a result, the

---

\(^{42}\) See Schwartz & Solove, supra note 32, at 1830 (highlighting the shortcomings of the non-public approach).

\(^{43}\) See Schwartz & Solove, supra note 32, at 1831 (reiterating qualities of a rule that requires a triggering event).


\(^{45}\) See id. (listing the factors that would trigger the Act).


\(^{47}\) See Schwartz & Solove, supra note 32, at 1832 (criticizing the specific-approach as too limited).

\(^{48}\) See Ohm, supra 29, at 1742 (questioning the concept of PII in light of modern re-identification technology and naming this problem “whack-a-mole”).

\(^{49}\) See Schwartz & Solove, supra note 32, at 1828-29 (summarizing various inconsistent approaches in defining and applying PII).

\(^{50}\) See Schwartz & Solove, supra note 32, at 1835 (acknowledging inconsistent definitions of PII as the main drawback of current privacy regulations).
fine line between PII and non-PII continues to fluctuate based on contextual and ever-changing technology.  

D. Using Anonymization in Balancing Internet Privacy

In an effort to protect the privacy of individuals, data administrators or collectors anonymize data when storing or disclosing person-specific information. Anonymization is the “process of removing or modifying the identifying variables in the microdata dataset.” Typical anonymization techniques include data reduction and data perturbation. Data reduction hides unique or rare recognizable data by increasing the number of individuals in the sample sharing similar identifying characteristics or selectively revealing such data. Data reduction methods include removing variables, removing records, global recording, top and bottom coding and local suppression. Data perturbation, on the other hand, modifies values of the identifying

51 See Schwartz & Solove, supra note 32, at 1836 (discussing the existing defects in the current distinctions between PII and non-PII); see also Helen Nissenbaum, Privacy as Contextual Integrity, 79 WASH. L. REV. 119, 136 (2004) (arguing that three different types of principles against privacy intrusion provide significant force to the reasonableness of privacy claims covered by the principles, but all fail to protect anything outside of them).

52 See Bin Zhou, Jian Pei & Wo-Shun Luk, A Brief Survey on Anonymization Techniques for Privacy Preserving Publishing of Social Network Data (Jan. 30, 2014), archived at www.perma.cc/YWY3-UVVE (stating the goal of anonymization of data is to minimize the amount of personal information in the course of running pattern-based searches); Mary DeRosa, Data Mining and Data Analysis for Counterterrorism, CENTER FOR STRATEGIC AND INTERNATIONAL STUDIES REPORT (March 2004), archived at www.perma.cc/Y66C-YKER (stating that anonymization techniques mask identifying information so data may be shared without exposing individual identities).


54 See Anonymization Techniques, International Household Survey Network (2009), archived at www.perma.cc/Q5VQ-3QXV (listing the two common anonymization techniques in masking sensitive data).

55 See Ira S. Rubinstein, Ronald D. Lee & Paul M. Schwartz, Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches, 75 U. CHI. L. REV. 261, 268 (2008) (explaining how data reduction works to anonymize personal data); see also International Household Survey Network, supra note 54 (detailing the actual process of data reduction in anonymizing data).

56 See International Household Survey Network, supra note 54 (supporting the proposition that data reduction successfully removes sensitive information from data).
attributes using a randomized process.\textsuperscript{57} Data perturbation techniques include micro-aggregation, data swapping, post randomization, adding noise and resampling.\textsuperscript{58} Latanya Sweeney, a well-known computer scientist in data privacy\textsuperscript{59}, suggests a privacy model called k-anonymity that ensures that no data disclosure will allow a person to be distinguished from fewer than “k-1” other individuals, leaving the value of “k” up to policy makers.\textsuperscript{60} A common challenge with all anonymization techniques is to identify the PII that allows an inference of identity and then controlling this inference, a problem that will be addressed in Part III of this note.\textsuperscript{61}

Anonymization has encouraged the Legislature to outweigh what seemed to be a minimal risk of sharing de-identified data against important values such as security, innovation and free flow of information.\textsuperscript{62} Even if information falls within the scope of PII, Congress permits a more flexible regulatory system as long as such data is anonymized.\textsuperscript{63} Therefore, sensitive information may be traded publicly as long as the data administrator makes the PII unidentifiable.\textsuperscript{64}


\textsuperscript{58} See International Household Survey Network, supra note 54 (listing several widely used data perturbation techniques).

\textsuperscript{59} See THE CARNEGIE MELLON UNIVERSITY: QUALITY OF LIFE TECHNOLOGY CENTER, infra note 75 (offering Latanya Sweeney’s biography).

\textsuperscript{60} See Latanya Sweeney, Achieving k-Anonymity Privacy Protection Using Generalization and Suppression, 10 INT’L J. ON UNCERTAINTY, FUZZINESS AND KNOWLEDGE-BASED SYS. 571, 572 (2002), archived at www.perma.cc/434N-QD46 (explaining the concept and application of k-anonymity in protecting person-specific information).

\textsuperscript{61} See DeRosa, supra note 52, at 18 (finding a challenge with anonymization research is to find the PII that allows an inference of identity and to control that inference); see also infra Part III (describing the problem with inference and control of identity).

\textsuperscript{62} See Ohm, supra note 29, at 1735-1736 (hinting at how the Legislature overlooks the risks of privacy harms and focuses on the advantages of information flow); but see Jane Yakowitz, Tragedy of the Data Commons, 25 HARV. J.L. & TECH. 1, 4 (2011) (suggesting the social utility of data sharing is significantly undervalued by most privacy scholars such as Paul Ohm).

\textsuperscript{63} See e.g., 42 U.S.C § 1320d-2 (2013) (allowing health professionals to trade PII as long as the data has been made unidentifiable).

\textsuperscript{64} See id. (exemplifying the reliance of congress on de-identifying information to provide security and the free flow of information).
With the help of anonymization, Congress has developed law around the concept of PII to avoid weighing the cost and benefits of privacy regulations. At first glance, Congress seems to have developed an approach that evaluates the inherent privacy risk of data categories by assessing with mathematical precision whether or not a data category causes sufficient harm to be regulated. In reality, Congress has just used anonymization in order to avoid making a real decision in balancing privacy interests.

III. THE PII PROBLEM AND ONLINE BEHAVIORAL TARGETING

Although anonymization has served well as online privacy protection’s first line of defense in the past, current technology demonstrates that anonymization has outlived its usefulness. “Anonymity myth,” a term coined by Schwartz and Solove, describes the common mistaken assumption that most internet activities are not identifiable so long as real names are not used. Schwartz and Solove underscore the conflation between momentary anonymity with actual untraceability by arguing that traceability exists whenever one is online. For example, Internet service providers have database entries that link IP addresses with particular computers or devices and,

65 See Ohm, supra note 29, at 1736-1738 (proposing that Congress avoided the difficult task of balancing the costs and benefits of privacy regulations by relying on the concept of PII); see also Cynthia Dwork, Differential Privacy, in AUTOMATA, LANGUAGES AND PROGRAMMING, 33rd INT’L COLLOQUIUM 4 (Springer, 2006) (showing that regulations based on the concept of PII does not require an utility versus privacy harm analysis).
66 See Ohm, supra note 29, at 1736-1738 (developing the idea that the assessment of whether a data category warrants regulation is ineffective and superficial).
67 See Ohm, supra note 29, at 1736-1738 (criticizing Congress’s unwillingness to address the problem of conflicting privacy interests).
68 See Ohm, supra note 29, at 1716 (contending that there is more than enough evidence for researchers to reject anonymization as a privacy-providing panacea).
69 See Schwartz & Solove, supra note 33, at 1836-1837 (proffering the term “anonymity myth”).
70 See Schwartz & Solove, supra note 33, at 1837 (clarifying the common mistaken belief that anonymity exists most of the time on the Internet).
under many circumstances, to specific individuals. Some courts have held that the non-PII nature of an IP address does not trigger a privacy claim because a computer may be shared by multiple individuals. Technology development is, however, slowly yet steadily breaking down the distinction between PII and non-PII.

A. De-Anonymization

Research shows that it is highly possible to reconstruct anonymized data and to re-identify individuals from non-PII and have well documented such a trend in recent years. The concept of de-anonymization is best illustrated by three famous case studies, described below.

1. Simple Demographics Can Uniquely Identify Us

Latanya Sweeney, professor of computer science at Carnegie Mellon University, conducted an experiment using the 1990 census and discovered that combinations of just a few unique characteristics can be used to identify some individuals. Her study revealed that 87.1% of people in the United States can be identified by the combi-
nation of their ZIP code, birth date and gender. More significantly, her study showed that even less-specific data combinations such as city, birth date and sex can identify 53% of United States citizens. A subsequent study replicated Sweeney’s analysis of the 2000 census and found that only 63% of United States citizens can be identified by the combination of their ZIP code, birth date and gender.

Reidentification can easily be done by lay people without access to such databases. In the mid-1990’s, Group Insurance Commission (GIC), a government agency in Massachusetts, decided to release summary records of state employee’s hospital visits at no cost to anybody who requested them. GIC anonymized the record by removing identifiers such as names, addresses and social security numbers before releasing the data. In response, then-graduate student Sweeney purchased the voter rolls from the City of Cambridge, where then Governor Bill Weld lived. The voter rolls contained,

77 See Sweeney, supra note 8, at 16 (warning that a large percentage of people can be identified by the combination of three easily obtainable pieces of information); but see Philippe Golle, Revisiting the Uniqueness of Simple Demographics in the US Population, 5TH ACM WORKSHOP ON PRIVACY IN THE ELECT. SOC’Y (2006), archived at www.perma.cc/UUB8-TWCP (revisiting Sweeney’s study and calculating that only 61% of population could be identified by the combination of ZIP code, birth date, and gender).

78 See Sweeney, supra note 8, at 2 (showing the possibility of identifying half of the U.S population through a combination of generic data).

79 See Golle, supra note 78 (reinforcing the theory that anonymization has no effect before the use of aggregated data). Golle noted in his paper that he could not explain the discrepancy because he lacked detailed information about the data collection and analysis techniques of Sweeney’s original study. See Golle, supra note 78 (explaining that his conclusion was hindered due to the lack of detailed information that was available to Sweeney).


82 See id. (alleging the process GIC used in anonymizing voters’ record before releasing such data).

83 See Greely, supra note 80 (describing how Sweeney gained access to voter information for her research).
among other things, the name, address, ZIP code, birth date and sex of every voter in Cambridge. By combining the GIC data with the voter rolls, Sweeney easily identified Governor Weld’s medical record because only six people in Cambridge shared his birth date, only three were men and of the three, he was the only one who lived in that ZIP code.

2. The Netflix Prize Surprise

Netflix, an online movie rental service, launched a prize with a result that took America by surprise in 2006. Intending to improve its movie rating system, Netflix proposed that the first team to use its expansive database of user movie ratings to improve Netflix’s recommendation algorithm would win one million dollars. Two weeks after one hundred million records were released, Arvind Narayanan and Vitaly Shmatikov demonstrated the power of combining databases when they identified two Internet Movie Database (IMDb) users out of a pool of fifty, in the Netflix database. By cross-referencing a user’s public IMDb ratings and user information, along with the private database released by Netflix, one may deduce a user’s real identity, along with sensitive information such as political views, religion and sexual orientation. In 2009, Netflix announced

---

84 See Greely, supra note 80 (illustrating the type of information Sweeney used in identifying individuals).
85 See Sweeney, supra note 8, at 2 (rejecting anonymization as an effective method of protecting privacy on the basis that individuals can be identified under aggregation of data).
86 See Netflix Prize Rules, NETFLIX, archived at www.perma.cc/3FYM-ZAXH (dating the Netflix prize contest commencement as October 2, 2006).
87 See id. (outlining the goal and prize of the Netflix contest).
88 See Narayanan & Shmatikov, supra note 8, at 13 (demonstrating the power of combining two or more anonymous databases to identity specific subjects). Due to a fear of violation of IMDb’s terms of service, the authors only sampled around fifty IMDb users, thus the results do not imply anything about the percentage of IMDb users who can be identified in the Netflix Prize dataset; this should only be viewed as a proof of concept. See Narayanan & Shmatikov, supra note 8, at 13 (demonstrating the issues with the experiment).
89 See Narayanan & Shmatikov, supra note 8, at 13 (suggesting individuals may be linked to the specific types of movies they have watched or rated). Unlike Netflix, IMDb posts user movie ratings publicly on its website and these users frequently rate movies under their real names. See, e.g., Reviews & Ratings for The Godfather, IMDb (2012), archived at www.perma.cc/C4DU-QBMY (revealing IMDb users’ willingness to use their real names when posting movie ratings online).
that it settled a class action lawsuit where its customers alleged violations of various state and federal privacy laws. That plans for a second contest were abandoned.

3. AOL Research Shows You Are What You Search

In 2006, America Online (AOL) released twenty million search queries for over 650,000 individuals that used its search engine over a period of three months under a new initiative named “AOL Research”. This data were assumed to be anonymized after AOL removed identifying information such as the AOL username and IP address. Instead, AOL replaced such identifying information with unique identification numbers so researchers could not link different searches to individual users. While some condemned AOL, others argued that there was no violation of privacy because there was no linkage between the anonymized queries and actual individuals.

The New York Times shattered AOL’s shield of anonymity when it revealed the true identity of user “4417749” as Thelma Arnold. After cross-referencing this user’s queries, such as “landscapers in

---

91 See Ohm, supra note 29, at 1722 (discussing that movies such as “Fahrenheit 9/11”, “Jesus of Nazareth” and “Queer as Folk” may disclose movie reviewer’s sexual orientation and religious or political views).
92 See Michael Barbaro & Tom Zeller, Jr., A Face is Exposed for AOL Searcher No. 4417749, NEW YORK TIMES, (2006), archived at www.perma.cc/WA6Y-VMHM (articulating how the reporters tracked down the real identity of AOL user 4417749 by cross-referencing anonymized search queries).
93 See id. (alleging the process AOL used in anonymizing users’ identifying information was flawed).
94 See id. (explaining the anonymizing method AOL used before making a public release of such data).
96 See, e.g., Greg Linden, A Chance to Play with Big Data, GEEKING WITH GREG (2006), archived at www.perma.cc/63PT-3U3P (classifying the privacy concerns of AOL data release as merely a theoretical possibility because no one had yet been identified).
97 See Barbaro & Zeller, supra note 92 (announcing the failure of “AOL Research” when Thelma Arnold is identified as user 4417749 in the so-called anonymized data).
Lilburn, Ga”, the last name “Arnold” and “homes sold in shadow lake subdivision Gwinnett county Georgia”, the New York Times reporter quickly tracked down Thelma Arnold, an elderly widow from Lilburn, Georgia.98 A shocked Arnold confirmed that she had authorized embarrassing searches such as “numb fingers”, “60 single men”, and “dog that urinates on everything”.99 Consequently, AOL admitted its failure in anonymizing the data and apologized for the privacy violation.100

These three cases illustrate a common problem: a combination of non-PII can be, and has been, used to produce PII.101 The FTC has identified three possible ways to re-construct anonymized data.102 First, it may be possible to merge non-PII with PII.103 Sweeney’s study demonstrated how combinations of non-PII and PII help narrow the number of individuals and can be linked to identifying information to reveal individual identities.104 Second, it will likely become easier to identify individuals based on information traditionally considered to be non-PII.105 For example, although most IP addresses have traditionally been considered as non-PII, the development of

98 See Barbaro & Zeller, supra note 92 (condemning AOL for releasing user data that lead an adversary to discover the real identity of its user).
99 See Barbaro & Zeller, supra note 92 (reporting the shock that Thelma Arnold experienced when she found out her identity and search inquiries were publicly released).
101 See supra notes 75-100 and accompanying text (detailing three examples of PII flaws).
103 See id. at 22 (explaining the possibility of merging non-PII and PII). “For example, a website might collect anonymous tracking data and then link that data with PII (i.e name, address) that the consumer provided when registering at the site.” Id. at 22.
104 See Sweeney, supra note 8, at 34 (revealing the ease of reidentifying individuals as long as there is more than one available data set).
105 See Self-Regulatory Principles, supra note 102, at 22 (noting that the scope of PII continues to grow). See e.g., Narayanan & Shmatikov, supra note 8 (articulating that data that would traditionally be categorized as non-PII are now becoming PII).
technology has shown the possibility to link IP addresses to not only electronic devices, but also individuals.\textsuperscript{106} Similarly, Barbaro and Zeller found Thelma Arnold by combining multiple sets of non-PII data.\textsuperscript{107} Third, anonymized data may become identifiable when combined and linked by a common identifier.\textsuperscript{108} Narayanan and Shmatikov were able to turn abstract identification results into concrete names by comparing Netflix ratings to IMDb data.\textsuperscript{109} These studies suggest that the uniqueness of non-PII can turn into PII in the hands of those who have access to more than one set of data.\textsuperscript{110}

Auxiliary information, more commonly known as outside information, is the piece of puzzle that completes the entire picture.\textsuperscript{111} In theory, anonymization is the ideal protection for data if there were no external sources to cross-reference.\textsuperscript{112} In reality, a wide range of information about people is available through many easily accessible means.\textsuperscript{113} The Legislature has created PII to avoid balancing the importance of privacy versus free flow of information,\textsuperscript{114} yet it is the free flow of information that allows aggregation of data and heightens the ability to create PII from non-PII.\textsuperscript{115}

\textsuperscript{106} See Schwartz & Solove, supra note 32, at 1837 (warning that technology will continuously expand the scope of PII until it includes virtually everything); see also Self-Regulatory Principles, supra note 102, at 22 (commenting on the effect of technology on current privacy regulations).
\textsuperscript{107} See Barbaro & Zeller, supra note 92 (strengthening the theory that aggregation of data increases the likelihood of reidentification).
\textsuperscript{108} See Self-Regulatory Principles, supra note 102, at 22 (explaining the role common-identifier plays in re-identification).
\textsuperscript{109} See Narayanan & Shmatikov, supra note 8, at 13 (demonstrating how Narayanan and Shmatikov reidentified Netflix users by comparing Netflix’s data to IMDb’s database).
\textsuperscript{110} See Ohm, supra note 29, at 1723 (highlighting the privacy risks of aggregating large databases).
\textsuperscript{111} See Ohm, supra note 29, at 1724 (describing the role and effect of ancillary information in reconstructing anonymized identities).
\textsuperscript{112} See Ohm, supra note 29, at 1724 (reinforcing the danger of comparing aggregated data from multiple databases).
\textsuperscript{113} See Ohm, supra note 29, at 1724 (stressing the impossibility of eliminating the cross referencing of data).
\textsuperscript{114} See Ohm, supra note 29, at 1736-1738 (arguing that Congress used the concept of PII to avoid weighing privacy interests).
\textsuperscript{115} See Schwartz & Solove, supra note 33, at 1821 (blaming reidentification on aggregation of data).
B. Uncovering the Anonymity Myth

By definition, anonymity does not equal privacy.\textsuperscript{116} The value of privacy rests in its protection of individual autonomy as well as the power of decision-making.\textsuperscript{117} Privacy is also associated with fairness because it guarantees a level playing field for information flow.\textsuperscript{118} Therefore, the underlying value and alleged effect of anonymization do not directly follow the value of privacy.\textsuperscript{119} Re-identification techniques can easily reconstruct PII and undermine Internet users’ autonomy.\textsuperscript{120}

The breakdown of the PII dichotomy further exposes Internet users to actual harm.\textsuperscript{121} Each piece of information, PII or non-PII, develops its greatest value and power when pieced together.\textsuperscript{122} The common myth that Internet activities are anonymous unless specific PII are given has been proven wrong.\textsuperscript{123} Once the distinction between PII and non-PII starts to blur, data collectors and businesses will find themselves competing “to see who can convert customer secrets into the most pennies.”\textsuperscript{124} Not only are these groups building

\textsuperscript{116} See Catherine Dwyer, Behavioral Targeting: A Case Study of Consumer Tracking on Levis.com, 4 (2009), archived at www.perma.cc/75SF-NCMN (stating “[t]he problem with the equating anonymity with privacy becomes apparent by considering the instrumental value of privacy.”).
\textsuperscript{117} See id. (correcting the common view that privacy regulations only concern protecting information).
\textsuperscript{118} See id. (applying the concept of fairness to privacy regulations).
\textsuperscript{119} See id. (criticizing the concept of anonymization with respect to fairness issues).
\textsuperscript{120} See id. (condemning anonymization techniques as depriving individual’s decision making power).
\textsuperscript{121} See Ohm, infra 124 (illustrating that the blurring of PII and non-PII ultimately exposes Internet users to privacy harms).
\textsuperscript{122} See William J. Fenrich, Common Law Protection of Individual Rights in Personal Information, 65 FORDHAM L. REV. 951, 952 (1996) (stating “the information develops its greatest value, and greatest power, when the individual pieces are gathered and layered on top of one another, creating a detailed profile of who you are and what you do”); Peter R. Orszag, Memorandum for the Heads of Executive Departments and Agencies (June 25, 2010), archived at www.perma.cc/93Z8X9UW (discussing the potential of non-PII turning into PII when additional information is made public or when other pieces of information are strung together).
\textsuperscript{123} See Schwartz & Solove, supra note 33, at 1836-1837 (stressing the concept that PII no longer protects sensitive information or personal identities).
\textsuperscript{124} See Paul Ohm, Don’t Build a Database of Ruin, HARVARD BUSINESS REVIEW BLOG NETWORK (Aug 23, 2012), archived at www.perma.cc/Q8NQ-GVCE (ex-
massive databases to house an indefinite amount of information about every person who has ever used the Internet, but they are also combining these databases.\textsuperscript{125}

Professor Paul Ohm has adopted the name “Database of Ruin” for this single, powerful database.\textsuperscript{126} The Database of Ruin is the worldwide collection of all information held by third parties that may be used to probe into private lives.\textsuperscript{127} Once a connection between a piece of data and a person’s real identity is established, it can be used to unlock other anonymized databases.\textsuperscript{128} Even the most seemingly non-sensitive personal information has the potential to cause privacy harm because it increases the ability to link data.\textsuperscript{129} As a result, re-identification techniques defeat the purpose of almost every privacy law and regulation in the U.S.\textsuperscript{130}

The abstract distinction between PII and non-PII has slowly faded away over the years.\textsuperscript{131} Further, the rise of online behavioral advertising now challenges the traditional definition of consumer PII and highlights an urgent need to revisit the current privacy regulation on online behavioral targeting.

\textbf{C. Online Behavioral Targeting}

Online behavioral targeting, sometimes known as behavioral advertising or behavioral marketing, involves tracking individual’s
online activities for the purpose of delivering tailored advertising to potential customers. Companies generally use “cookies”, amongst other methods, to track consumer activities by associating those activities with a particular computer or device. “Cookies” are small data text files that a website stores on a consumer’s electronic device, which transmits the device’s browsing activities back to its server. This information allows companies to compile and classify expansive consumer profiles based on demographics and personality traits. Advertisers then use advanced algorithms to analyze this information and predict a user’s likely purchasing inclinations.

Online behavioral tracking is extremely valuable because not only does it allows businesses to align their advertisements with what the potential consumer is likely to click on, browse and eventually purchase, it also enables useful features to Internet users such as saving customized personal preferences and settings on the web. These benefits have generated a new form of service provider: network advertisers, companies that select and deliver appropriate advertisements to participating websites across their network. These vast networks may include “hundreds or thousands of different, unrelated websites”, allowing network advertisers to generate rich profiles about the activities of a specific device user. In general, the data that behavioral advertisers collect does not fall under PII because it

---

133 See Self-Regulatory Principles, supra note 102, at 2 (detailing how companies generally track consumer online activities).
134 See Self-Regulatory Principles, supra note 102, at 2 n.3 (defining “cookies” in the contextual of behavioral targeting).
135 See Fuelleman, supra note 4, at 813-814 (summarizing the organizational process of behavioral advertising after companies collect consumer information).
136 See Joanna Penn, Behavioral Advertising: The Cryptic Hunter and Gatherer of the Internet, 64 FED. COMM. L. J. 599, 601 (explaining the role of algorithm in predicting a consumer’s preferences).
137 See id. at 601 (showing that behavioral targeting offers positive consequences).
139 See Self-Regulatory Principles, supra note 102 at 3 n.5 (introducing the role of network advertisers).
140 See Self-Regulatory Principles, supra note 102, at 3 n.5 (hinting at the danger of aggregation of such data).
does not include the user’s real name or other personally identifying information that can tie the user to his or her real identity.\textsuperscript{141}

As discussed earlier, however, current technology allows data collectors to reconstruct anonymized data through a process that involves multiple parties and databases.\textsuperscript{142} Website owners, often service providers, use third-party analytical tools to monitor and assess their site traffic, of which the end-user is unaware.\textsuperscript{143} This trend of merging data promotes exchange of information in online advertisement networks and encourages the aggregation of non-PII.\textsuperscript{144}

Behavioral targeting, like the three breaches of privacy discussed earlier in this Note, implicates not only concrete personal information, but also a consumer’s “inner identity.”\textsuperscript{145} A complete consumer data profile is, in a way, an attempt to replicate the consumer’s personality in order to create customized advertisements.\textsuperscript{146} Such exploration into one’s personality is more private and valuable than aspects of external identity and corresponds with the right of privacy proposed by Warren and Brandeis in 1890.\textsuperscript{147} For instance, the de-

\begin{itemize}
\item \textsuperscript{141} In re DoubleClick, Inc. Privacy Litig., 154 F. Supp. 2d 497, 506 (S.D.N.Y 2001) (alleging the collection and use of data by network advertisers are typically what is traditionally considered as non-PII).
\item \textsuperscript{142} See Omer Tene & Jules Polonetsky, To Track or “Do not Track”: Advancing Transparency and Individual Control in Online Behavioral Advertising, 13 Minn. J.L. Sci. & Tech. 281, 302-303 (demonstrating the use of non-PII in tracking and analyzing consumer preferences).
\item \textsuperscript{143} See Sarah Lacy, Web Numbers: What’s Real?, BLOOMBERG BUSINESSWEEK MAGAZINE (Oct. 22, 2006), archived at www.perma.cc/4Y96-9P2X (detailing the methods and confusion over measuring online traffic).
\item \textsuperscript{144} See Self-Regulatory Principles, supra note 102, at 3 n.5 (describing the trend in moving towards creating massive databases that virtually combines information about everybody).
\item \textsuperscript{145} See Andrew J. McClurg, A Thousand Words are Worth A Picture: A Privacy Tort Response to Consumer Data Profiling, 98 NW. U. L. REV. 63, 124 (2003) (arguing that behavioral advertisement violates privacy by examining consumer’s inner identity).
\item \textsuperscript{146} See id. at 124-125 (reasoning that consumer profiling is gathering intelligence and attempting to replicate an individual’s personality); see also Noam Cohen, As Data Collecting Grows, Privacy Erodes, NEW YORK TIMES (2009), archived at www.perma.cc/VUP5-6XDJ (describing behavioral targeting as “the surveillance business model”).
\item \textsuperscript{147} See McClurg, supra note 145, at 108 (suggesting consumer data profiles are, in some aspect, more personal and private than some external personal data); see also
\end{itemize}
livery of behavioral advertisement may reveal private information to other users of the same electronic device because cookies track the online activity of a device, not any particular person. In other words, behavioral targeting may lead to not only identify thefts, but also embarrassment, inconvenience and unfairness.

The FTC has long recognized that the traditional definitions of PII and non-PII are becoming less and less meaningful, and the dichotomy should no longer be used to determine the protections for consumer data. Such concern is echoed in In Re DoubleClick, where DoubleClick, a subsidiary of Google that provides Internet advertising services, was accused of creating a super database that matches users’ online activities with their real identities by merging with Abacus Direct, a data collection company. The FTC launched an investigation and concluded that DoubleClick had not engaged in unfair trade practices after DoubleClick announced that it planned to forgo the combination of databases. However, the court specifically underscored FTC’s concern by warning that “…DoubleClick’s practices and consumers’ privacy concerns… are not unknown to Congress. Indeed, Congress is currently considering legislation that specifically recognizes and regulates the online harvesting of user information.”

Warren & Brandeis, supra note 16, at 195 (articulating the significance of right of privacy because otherwise our inner personalities and identities would be violated).

148 See Self-Regulatory Principles, supra note 102 at 2 (illustrating the problem of tracking with “cookies”).
149 See Memorandum from Clay Johnson III on Safeguarding Personally Identifiable Information to the Heads of Departments and Agencies, The Executive Office of the President (2006), archived at www.perma.cc/654C-HWJ4 (stating “…the loss of personally identifiable information can result in substantial harm, embarrassment, and inconvenience to individuals and may lead to identity theft or other fraudulent use of the information.”).
150 See Self-Regulatory Principles, supra note 102 at 21 (acknowledging the blurring of PII and non-PII).
152 See id. at 505 (suggesting the possible harmful results of combining databases that links individuals to their real identities).
153 See id. at 505-506 (reporting FTC’s investigations and results).
154 Id. at 526.
Similarly, consumers have conveyed a genuine fear regarding consumer data collection on the Internet.\textsuperscript{155} The FTC has suggested in its 2009 report that “consumers are concerned about the collection of their data, regardless of whether the information is characterized as PII or non-PII.”\textsuperscript{156} In an independent phone survey, 55% of a sample group composed of 18 to 24 years old objected to behavioral targeting.\textsuperscript{157} Critics have also warned that behavioral targeting contradicts the traditional concept of privacy, limits user options and violates user expectations.\textsuperscript{158}

### D. Attempts to Regulate Behavioral Targeting

While there is no privacy law in the U.S regarding behavioral targeting, there are some regulations.\textsuperscript{159} The current consumer data privacy framework centers on the Consumer Privacy Bill of Rights.\textsuperscript{160} The FTC, under the 1914 Federal Trade Commission Act,\textsuperscript{161} has broad authority to regulate unfair and deceptive business practices.\textsuperscript{162}

\footnotesize{\textsuperscript{155} See Stephanie Clifford, \textit{Two-Thirds of Americans Object to Online Tracking}, \textit{NEW YORK TIMES} (2009), archived at www.perma.cc/PN7Z-V79F (reporting that approximately two-thirds of Americans object to online tracking and behavioral advertisement).  
\textsuperscript{156} See \textit{Self-Regulatory Principles}, supra note 102 at 23 (voicing consumer concerns regarding behavioral targeting).  
\textsuperscript{157} See Clifford, supra note 155 (reporting the amount of individuals who oppose the concept of behavioral targeting).  
\textsuperscript{158} See Tracy A. Steindel, \textit{A Path Toward User Control of Online Profiling}, 17 MICH. TELECOMM. & TECH. L. REV. 459, 468-469 (2011) (condemning online profiling as a harmful practice because it violates the traditional concept of privacy and takes away users’ control over the flow of personal information); \textit{see also} Dustin D. Berger, \textit{Balancing Consumer Privacy with Behavioral Targeting}, 27 SANTA CLARA COMPUTER & HIGH TECH. L. J. 3, 18-19 (2011) (recognizing that behavioral targeting results in compilation of “sensitive consumer data that exists outside of her ability to protect, control or monitor” and can paint an embarrassing “picture” of the consumer).  
\textsuperscript{159} See Penn, supra note 136, at 611 (outlining different self-regulatory principles in behavioral targeting).  
\textsuperscript{160} See \textit{The White House, Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy} 7 (Feb. 2012), archived at www.perma.cc/79D6-F2L5 (outlining the goal of Consumer Bill of Rights as to advance the following objectives for consumers: individual control, transparency, respect for context, security, access and accuracy, focused collection, accountability).  
\textsuperscript{162} See id. (outlining the scope of FTC’s authority).}
Although the FTC had first reported the potential benefits and harms of online profiling to Congress in 2000,\textsuperscript{163} it has taken no direct measures to control behavioral targeting and has simply relied on the industry’s self-regulation.\textsuperscript{164} The Online Privacy Alliance (OPA) and Network Advertising Initiative (NAI), both comprised of leading industry companies, have taken the initiative to implement guidelines and principles for their members.\textsuperscript{165} Both groups ultimately failed because of a lack of enforcement, inadequate participation from the industry and their standards offered little privacy protection.\textsuperscript{166}

Nine years later, the FTC issued a set of self-regulatory principles to guide behavioral advertisers and other companies.\textsuperscript{167} The proposed principles focus on four concepts: (1) transparency and consumer control,\textsuperscript{168} (2) reasonable security and limited data retention for consumer data,\textsuperscript{169} (3) affirmative express consent for material changes to existing privacy promises,\textsuperscript{170} and (4) affirmative express consent to using sensitive data for behavioral targeting.\textsuperscript{171} Some pri-


\textsuperscript{164} See Fuelleman, supra note 4, at 815 (commenting on the FTC’s reliance on the industry’s questionable self-regulation).

\textsuperscript{165} See Online Privacy Alliance, Guideline for Online Privacy Policies (2010), archived at www.perma.cc/DY3A-YVAS (focusing on the adoption and implementation of a privacy, notice and disclosure, choice/Consent, data security, data quality and access); NAI Staff, 2008 NAI principles: The NAI’s Self-Regulatory Code of Conduct 3, NAI 2008, archived at www.perma.cc/ZT9Y-RJ94 (emphasizing NAI’s commitment to self-regulate “with respect to notice, choice, use limitation, access, reliability, and security”).


\textsuperscript{167} See Self-Regulatory Principles, supra note 102 (documenting self-regulatory principles issued by the FTC).

\textsuperscript{168} See Self-Regulatory Principles, supra note 102, at 30-37 (summarizing the principles of transparency and consumer control)

\textsuperscript{169} See Self-Regulatory Principles, supra note 102, at 37-38 (summarizing the principle of reasonable security and limited data retention for consumer data).

\textsuperscript{170} See Self-Regulatory Principles, supra note 102, at 39-41 (summarizing the principle of affirmative express consent for material changes to existing promises).

\textsuperscript{171} See Self-Regulatory Principles, supra note 102, 42-44 (summarizing the principle of affirmative consent to using sensitive data for behavioral promises). It is
vacy advocates specifically criticize that the FTC principles should be more rigid and explicit to produce meaningful results. Others continue to argue that the system of self-regulation in behavioral targeting is fundamentally flawed.

In Europe, the European Data Protection Directive and the European e-Privacy Directive largely make up a legal framework that applies to behavioral targeting. The European Data Protection Drive regulates the collection, processing, storage and transfer of personal data. It sets forth principles that are wider in scope than those proposed by the FTC in 2009, including notice, consent, proportionality, purpose limitation and retention periods.

In addition, the e-Privacy Directive protects, among other things, communications, traffic and location data and precisely controls the use of cookies. It provides useful insight into a co-regulatory approach. The Directive requires each member of the European Union (EU) to legislate the collection, use and disclosure of personal in-

worth noting that the FTC principles are largely consistent with the OPA guidelines and NAI principles. See Self-Regulatory Principles, supra note 102, 42-44. See Penn, supra note 136, at 607-10 (criticizing the FTC self-regulatory principles for excluding several forms of behavioral targeting from their scope).


See Tene & Polonetsky, supra note 142, at 310 (introducing the legal framework of privacy protection in Europe).

See Tene & Polonetsky, supra note 142, at 310 (explaining the functions of European Data Protection Drive).

See Tene & Polonetsky, supra note 142, at 310 (detailing the scope of the European Data Protection Drive).

See Tene & Polonetsky, supra note 142, at 310 (outlining the scope of the e-Privacy Directive).

See Hirsch, supra note 166, at 469 (contrasting the e-Directive co-regulatory approach to self-regulatory principles).
Article 27 of the Directive not only mandates all member nations to establish national data protection authorities and identify the conditions in which it is inappropriate to collect personal information, but also prohibits the collection of information such as race, religion or sexual orientation. Article 27 also requires all Internet firms and any other business that processes data to obtain informed consent from the data protection authority, as well as individuals, before commencing any data collection and processing.

Despite the Directive’s burdensome obligations, the EU lawmakers also avoided the question of utility versus privacy by relying on anonymization. The EU lawmakers tried to strike a balance between freedom of expression and privacy by not including all types of data under the scope of the Directive. The Directive excludes anonymized data, including data that is not “directly or indirectly identifiable”, from regulation. Much like their U.S counterparts, this classification triggers numerous debates as on the definition of effective anonymization in tracking consumers’ online activities. For instance, Google argues that it has provided adequate protection of its users by throwing away only part of every IP address it records, while Yahoo! and Microsoft have chosen to throw away the entire IP address. This debate, much like the PII and non-PII de-

---

180 See id. at 47 (specifying that Article 27 requires the creation of national data protection agencies and legislation on collection of online data).
181 See id. (imposing the restrict requirement of obtaining informed consent before starting any data collection or processing).
182 See Ohm, supra note 29, at 1736 (stating that the EU Legislature, like its U.S counterpart, did not have to weigh privacy interests in passing the e-Directive).
183 See Ohm, supra note 29, at 1736 (explaining EU Legislature’s failed attempt to balance privacy interests).
184 See EU Data Protection Directive, supra note 179, at 33.
185 See Ohm, supra note 29, at 1739 (comparing the debate on anonymization to the debate on PII).
187 See Katitza Rodriguez, European Privacy Officials: Google, Yahoo and Microsoft are Still Breaking European Privacy Law, ELECTRONIC FRONTIER
bate in the U.S, rests on the distinction between data that is “directly or indirectly identifiable” and those that are not. This choice to either anonymize data or comply with the Directive may now appear meaningless in light of re-identification technology.

At the national level, several proposed bills address the issue of online data privacy but none would resolve the PII problem. This problem is further worsened by the speed of ever-changing technology. As Schwartz and Solove put it, “today’s non-PII might be tomorrow’s PII.” It is also important to recognize that the ability to identify is driven by context, such as what specifically the search query includes and what ancillary information is available about the particular user. In conclusion, the PII problem continues to exist because PII cannot be defined in the abstract and the fine line between PII and non-PII is constantly shifting.

See Ohm, supra note 29, at 1764 (suggesting that easy re-identification may force every data collector to comply with the Directive because any database containing facts relating to people, no matter how indirectly, possibly falls under the scope of the Directive).

See Ohm, supra note 29, at 1763 (reiterating the idea that anonymization no longer offers any meaningful protection).

See, e.g., Best Practices Act, H.R. 611, 112th Cong. (2011) (stating the goal of the Bill is to “foster transparency about the commercial use of personal information, provide consumers with meaningful choice about the collection, use, and disclosure of such information, and for other purposes”); see also Do-Not-Track Online Act of 2011, H.R. 654, 112th Cong. (2011) (requiring the FTC to regulate online tracking).

See Schwartz & Solove, supra note 32, at 1845 (pointing out the impact advancing reidentification technology has on the concept of PII in the U.S).

See Schwartz & Solove, supra note 32, at 1846.

See Schwartz & Solove, supra note 32, at 1846 (explaining reidentification depends on many factors such as the specificity of search inquiries and other ancillary information); See also Nissenbaum, supra note 51, at 145 (proposing contextual integrity as the benchmark of privacy).
IV. GOODBYE PII: CONTEXTUAL REGULATIONS

There is an urgent need for the Legislature to regulate online behavioral targeting because of the high stakes involved in information leakages and the disappointing results of self-regulations. As categories of PII constantly expand, policy makers can no longer avoid a cost-benefit analysis of privacy regulations.

A. Abandoning the Concept of PII

The Legislature must re-evaluate laws and regulations that draw on the distinctions of PII and non-PII because such a sectorial approach no longer guarantees any meaningful privacy protection. Abandoning the concept of PII may at first seem uncomfortable and problematic. For decades, PII served as the center of privacy regulation. One may argue that without the concept of PII, there would be virtually no limits on the scope of privacy regulation. In reality, specific regulations based on contextual expectations complemented by a comprehensive regulation scheme that sets a general floor of privacy regulations can solve this problem.

B. A Proposal: Contextual Integrity

This section proposes a scheme that not only contextually regulates specific sectors with a focus on online behavioral targeting, but

194 See Penn, supra note 136, at 611 (criticizing the FTC self-regulatory principles for behavioral targeting); see also Campbell, supra note 173, at 771 (elaborating on the ineffectiveness of self-regulation in protecting privacy); Cody, supra note 173, at 1217 (noting the difficulties in monitoring compliance under a self-regulatory regime); Culnan, supra note 173, at 25 (doubting the implementation and effectiveness of self-regulation in protecting consumer privacy); Hirsch, supra note 166, at 469 (describing the result of self-regulation as “discouraging”).
195 See Ohm, supra note 29, at 1736-1738 (calling for a review of current privacy regulations).
196 See Ohm, supra note 29, at 1763 (drawing the inference that meaningful protection is not a product of anonymization).
197 See Ohm, supra note 29, at 1745 (recognizing any approach that is not PII-centric may be disruptive but necessary); see e.g., Schwartz & Solove, supra note 32, at 1866 (rejecting the idea of abandoning PII under the current privacy regulations model).
198 See supra Part II (outlining the role of PII in U.S privacy regulations).
199 See Schwartz & Solove, supra note 32, at 1866 (discarding the idea of abandoning PII because PII currently establishes the boundaries of privacy regulations).
also sets a general requirement for privacy regulations. Several scholars have suggested this approach to privacy regulations in the past. Solove and Ohm have both highlighted the significance of simultaneously looking to the specific context as well as a general framework to identify privacy harms and to understand the problems behind them. Helen Nissenbaum, on the other hand, offers the concept of “contextual integrity.” Contextual integrity builds on the idea that almost everything – events, transactions and human behaviors – happen in a context. Privacy norms, therefore, are rooted in the details of these societal, cultural and political expectations.

The proposed scheme begins with specific regulations around how one would reasonably expect data to be collected, stored, distributed and used for under the context the information was collected. If the collection, storage, distribution or use of data falls outside of this reasonable expectation, it automatically triggers the universal privacy protection; parties that violate this reasonable expectation are required to give clear and explicit disclosure and obtain informed consent.

Regulators must weigh different factors that serve as indicators of risk and instruments for reducing risk when making specific regulations under given contexts. What is necessary to safeguard health

---

200 See Ohm, supra note 29, at 1762 (proposing that the PII problem can be solved on a case by case basis by basis looking at specific context and using a general broad framework to identify privacy harms) (citing DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 46-49 (2008)).
201 See Ohm, supra note 29, at 1762 (offering the strength of both general and contextual regulation).
202 See Nissenbaum, supra note 51, at 136-37 (introducing the concept of contextual integrity).
203 See Nissenbaum, supra note 51, at 136-37 (explaining the rationale of contextual regulation).
204 See Nissenbaum, supra note 51, at 136-37 (describing contextual norms as a fluctuating concept based on varying expectations).
205 See Nissenbaum, supra note 51, at 136-138 (stressing the importance of reasonable expectations under different circumstances).
206 See, e.g., EU Data Protection Directive, supra note 179 (demanding informed consent prior to collecting or processing any data).
207 See Ohm, supra note 29, at 1764 (offering a basis for determining respective contextual expectations).
records may not be necessary for online search inquiries. These factors may include, but are not limited to, motive and purpose, data-handling techniques, private versus public release, sensitivity and quantity. Although all risks and benefits are rarely known in advance, these factors provide a rough sense of the risk of re-identifying a particular type of data under specific contexts. If the risk is very high, regulators should feel obligated to create more specific and restrictive regulations around the collection and use of information under specific circumstances.

Regulating information collection and usage on a contextual expectation requires an examination of contextual norms. As discussed above, contexts are largely made up of norms, which define essential elements such as expectations, behaviors, boundaries and much more. Nissenbaum introduces two types of informational norms: norms of appropriateness and norms of flow, also known as norms of distribution.

These two types of informational norms form the basis of contextual integrity in information privacy. Norms of appropriateness circumscribe the type and nature of information about a person that is allowed and even expected to reveal under specific circumstances. For example, a patient is expected to share her medical history and condition with her physician under a medical context, but not with

---

208 See Ohm, supra note 29, at 1768 (reinforcing the significance of catering to the reasonable expectations under various contexts).
209 See Ohm, supra note 29, at 1765-68 (outlining factors that may help in creating context-specific regulations); see also The European watchdog, the Article 29 Working Group, 2007 Working Party opinion. (offering factors that may be used to determine risks involved).
210 See Ohm, supra note 29, at 1764 (explaining the benefits of understanding the proposed factors in order to assess possible risks under specific contexts).
211 See Ohm, supra note 29, at 1768 (urging Legislature to take a more proactive role in regulating circumstances in which consumers are particularly vulnerable).
212 See Nissenbaum, supra note 51, at 138 (reinforcing the idea that contextual expectations are based on contextual norms).
213 See Nissenbaum, supra note 51, at 138 (defining contextual norms).
214 See Nissenbaum, supra note 51, at 138 (offering two types of informational norms).
215 See Nissenbaum, supra note 51, at 138 (introducing the concept of contextual integrity).
216 See Nissenbaum, supra note 51, at 138 (defining norms of appropriateness).
her employer under a different context. In other words, norms of appropriateness protect the varying degrees of knowledge concerning different relationships with different people. Norms of flow, on the other hand, govern the transfer of information from one party to another. Information distribution equality can only be upheld with the information provider’s freedom of choice and discretion. Contextual integrity in information privacy is violated when either of these two norms is violated.

Online behavioral targeting regulations, therefore, should focus on reasonable norms of appropriateness and norms of flow in the collection and use of information. Regulators must envision the reasonable, appropriate information that one may expect to produce to a first party under a specific context, and whether the information distribution to a third party respects a reasonable standard of information flow. For instance, when a person orders a dozen Fuji Apples on a grocery website, she may well expect to see Fuji Apples as the first option in the drop-down menu for her shopping convenience, but not to receive Fuji Apple phone advertisements or see pictures of Fuji apples on every webpage she visits. Similarly, a person can book a gay pride retreat to San Francisco on a travel website but remain private about her sexual preference at her home and workplace. Thus

217 See Nissenbaum, supra note 51, at 138 (providing context for an example of appropriate release of personal information in certain circumstances).
218 See Nissenbaum, supra note 51, at 138 (illustrating that norms of appropriateness allows individuals to be selective in sharing their information with different people).
219 See Nissenbaum, supra note 51, at 138 (defining norms of information flow).
220 See Nissenbaum, supra note 51, at 148-49 (emphasizing the significance of autonomy and free-choice in the information flow).
221 See Nissenbaum, supra note 51, at 138 (providing that contextual integrity can only be achieved by maintaining the norm of appropriateness and the norm of flow).
222 See Ohm, supra note 29, at 1763-64 (approving Nissenbaum’s concept of contextual integrity because specific sectors demand a different standard).
223 See Nissenbaum, supra note 51, at 141-42 (suggesting respecting contextual norms in the distribution or flow of data is as significant as collecting only appropriate information under a given context).
224 See Nissenbaum, supra note 51, at 121 (providing context for an example of suggestion of appropriate advertisements upon collection of personal data).
225 See Nissenbaum, supra note 51, at 140 (citing Ferdinand Schoeman, GOSSIP AND PRIVACY, IN GOOD GOSSIP 72, 73 (Robert F. Goodman & Aaron Ben-Ze’ev
cookies used to save preferences on the use of the website or user location would fall under the reasonable expectation of collecting and using such information while revealing the age and gender of the traveler and the nature as well as destination of the trip to a third party for marketing purposes would not.\textsuperscript{226} Such privacy regulations allow people the power to share information selectively, determined by the trust and nature of their relationships.\textsuperscript{227}

The parties whose collection, storage, distribution or use of data would violate the contextual integrity in information privacy are obliged to obtain informed consent before collecting data.\textsuperscript{228} This includes not only first parties or first party domains that collect information and third parties that use this information, but also those who store information in massive databases.\textsuperscript{229} This echoes Ohm, who urges privacy laws to include these entities that he calls “large entropy reducers.”\textsuperscript{230}

Although the requirement to obtain informed consent largely mirrors Article 27 of EU’s e-Privacy Directive, this second step of the proposed scheme slightly differs.\textsuperscript{231} On one hand, it is broader than Article 27 because it does not rely on the concept of PII and regulates all data, including those anonymized.\textsuperscript{232} On the other hand, it is less burdensome than Article 27 because it is triggered only if the collection, storage, distribution or use of the information violates a specific contextual regulation.\textsuperscript{233} Those who collect, store, distribute and use

\textsuperscript{226} See Nissenbaum, \textit{supra} note 51, at 140 (differentiating what would fall under the reasonable expectations under the suggested contexts).

\textsuperscript{227} See Nissenbaum, \textit{supra} note 51, at 139 (reiterating that the norm of appropriateness enables people the power to share information discriminately).


\textsuperscript{229} See Ohm, \textit{supra} note 29, at 1760-1761 (proposing regulations on those database owners who possess the key to re-identification).

\textsuperscript{230} See Ohm, \textit{supra} note 29, at 1760-1761 (defining “large entropy reducers” as those who store information in massive databases).

\textsuperscript{231} See \textit{supra} Part III.D (describing Article 27 of EU’s e-Privacy Directive, particularly its effort in regulating behavioral targeting).

\textsuperscript{232} See \textit{supra} Part III.D (questioning Article 27’s exemption of anonymized data).

\textsuperscript{233} See \textit{supra} Part III.D (focusing on Article 27’s burdensome requirements).
information within the reasonable expectation under the particular circumstances have no duty to comply with this general privacy regulation.\footnote{See EU Data Protection Directive, supra note 179, at 26 (outlining exemption from privacy regulation when data is anonymized).}

The proposed scheme brings back the original privacy tort liability standard proposed by Warren and Brandeis in the 1890.\footnote{See Warren & Brandeis, supra note 16, at 193-195 (proposing a right of privacy as a type of tort by pointing out the conflicts between technology and private life).} Online behavioral targeting regulations should rely heavily on judicial interpretation of the reasonable person standard used in administrative and constitutional law.\footnote{See Mayo Moran, The Reasonable Person: A Conceptual Biography in Comparative Perspective, 14 LEWIS & CLARK L. REV. 1233, 1234 (2010) (introducing the culpability-determining role of the reasonable person standard in torts and criminal law, as well as its judgment-related role in administrative and constitutional law).} Since behavioral targeting is so interconnected with ordinary everyday lives, it would only be logical and fair to require those who hold the key to unlock private data to act similarly to how a reasonable person under the circumstances would.\footnote{See id. at 1282 (stating the reasonable person was initially created for equality-sensitive areas of public law and can serve to correct structural deficiencies).}

\section*{C. Benefits and Challenges}

Now is the best time to initiate a new regulatory regime in the United States. Since the sectorial approach in the United States is flawed and the EU e-Privacy Directive proves to be too burdensome,\footnote{See Ohm, supra note 29 at 1762-63 (disapproving both the U.S’s exclusively sectorial approach and the European approach to privacy regulations).} the proposed scheme aims to takes a middle ground with a sector-specific privacy regulation, supplemented by a universal data privacy regulation.\footnote{See Ohm, supra note 29, at 1763 (stating the proposed privacy regulatory scheme was inspired by both Nissenbaum’s concept of contextual integrity and EU’s e-Privacy Directive).} Like any regulatory approach, the proposed scheme offers several benefits and faces possible challenges.\footnote{See Ohm, supra note 29, at 1763 (explaining that Nissenbaum’s approach has potential drawbacks according to other privacy scholars).}

The combination of comprehensive data regulation and enhanced obligations for specific sectors target the specific needs of the post-
anonymization United States.\textsuperscript{241} In comparison to a sweeping, society-wide approach, the proposed two-tiered scheme offers several benefits such as eliminating unenforceable guiding principles and unequal treatment of data.\textsuperscript{242} A reasonable person standard will create a uniform, foreseeable and neutral objective tool to determine liability in online behavioral targeting.\textsuperscript{243} It will be predictably more enforceable because such a standard offers more concrete measurements than existing guiding principles and allows judicial interference even when there is no injury.\textsuperscript{244} In addition, it breaks down the PII-oriented data hierarchy and treats all information equally.\textsuperscript{245}

The proposed scheme has a particularly high prescriptive value because it abandons the technology-centric approach to privacy regulations.\textsuperscript{246} Instead of trying to keep up with constantly changing technology\textsuperscript{247}, the proposed scheme is preventive in nature and focuses on the intent of the parties collecting, storing and using the data prior to data collection.\textsuperscript{248} It does not focus strictly on punishing those who harm or providing remedies for those harmed.\textsuperscript{249} As a result, contextual privacy regulations can always stay at least one step ahead of the realm of re-identification technology.\textsuperscript{250}

\textsuperscript{241} See Ohm, \textit{supra} note 29, at 1764 (explaining a new method of data privacy by using data regulation and an enhanced obligation structure for specific sectors).
\textsuperscript{242} See \textit{supra} Part III.D (pointing out the flaws of current self-regulation, particularly its unenforceability and unequal treatment of PII and non-PII).
\textsuperscript{243} See \textit{supra} Part III.D (emphasizing the need for an uniform and objective standard).
\textsuperscript{244} See \textit{supra} Part III.D (condemning the current guidelines as weak and unenforceable).
\textsuperscript{245} See \textit{supra} Part III.D (elaborating on the breakdown of the PII-centric approach to privacy regulations). \textit{See also supra} Part IV.A (emphasizing the need to abandon the distinction of PII and non-PII).
\textsuperscript{246} See Part III.D (urging a review of the current PII-centric approach to privacy regulations); \textit{supra} Part IV. (reiterating the need to abandon the concept of PII).
\textsuperscript{247} See Part III.C-D (describing how new technologies threatens to expand PII categories infinitively).
\textsuperscript{248} See Ohm, \textit{supra} note 29, at 1742 (noting the importance of preventive regulations).
\textsuperscript{249} See Ohm, \textit{supra} note 29, at 1742 (suggesting HIPAA’s approach is similar to a carnival whack-a-mole game).
\textsuperscript{250} See Ohm, \textit{supra} note 29, at 1742 (warning that the list of potential PII will never stop growing regardless of how regulators update their re-identification research).
Most significantly, the proposed scheme intends to strike a balance between utility and privacy. Noted as two concepts at war in privacy regulation, scholars are split on how to weigh the benefits of unconstrained information flow and possible privacy harm. Not only does the proposed scheme allow easy access to data that will be used in a way that is reasonable and expected, but it also ultimately guarantees an unrestricted information flow as long as the target Internet user gives informed consent. The Legislature may also take this concept further and create variations of the reasonable person standard as courts have created, such as a reasonable marketer standard and a reasonable researcher standard. As a result, Internet users will have more power to control whether their data will pass the second test and become public information.

The proposed scheme will also increase privacy awareness among Internet users and open up privacy discussions under the context of behavioral targeting. Like the EU e-Privacy Directive, the proposed scheme encourages disclosure and consent. It will force those who wish to collect, store, distribute and use data to consider their purposes and methods and give a concrete explanation in their legal disclaimers if such use is not reasonably expected under the circumstances. For example, search engines that store data in massive databases will be regulated because they will no longer be able

---

251 See Ohm, supra note 29, at 1752 (explaining the clash between utility and privacy in resolving anonymized data).
252 See Ohm, supra note 29, at 1752-53 (highlighting the conflicting concept of utility and privacy in privacy regulations); see also Dwork, supra note 65, at 4 (recognizing the inability to find a balance between utility and privacy would prevent lawmakers from creating a suitable privacy regulatory scheme).
253 See Ohm supra note 29, at 1759 (explaining another expected benefit of the proposed theory).
254 See e.g., J.D.B. v. North Carolina, 131 S. Ct. 2394 (2011) (holding the test for determining whether a juvenile was in custody in such that he should have received Miranda warnings must be evaluated through a reasonable juvenile standard).
255 See id. at 2407 (implying that the reasonable person test empowers the public with regard to the personal data).
256 See EU Data Protection Directive, supra note 179 (highlighting that privacy is enhanced by deletion of search data no later than six months post collection).
257 See EU Data Protection Directive, supra note 179 (detailing the restrict requirement to obtain informed consent).
258 See EU Data Protection Directive, supra note 179 (explaining a benefit of the proposed method in forcing data collectors to provide concrete reasons for their data collection).
to sell such data to marketing firms without naming the parties that will be involved and intended use for the data.\textsuperscript{259} The scheme will also promote privacy awareness because it requires Internet users to make the effort to actively agree to the collection and use of data before they are able to proceed to use the website.\textsuperscript{260} Although users may not meaningfully read the privacy disclaimer before agreeing to it, this gives users who are not familiar with the practice of behavioral targeting more incentive to learn about the practice and voice their concerns.

There are challenges that merit mention in the course of regulating online behavioral targeting. Like all new regulations, the proposed scheme requires extensive legislation and a high level of judicial interpretation.\textsuperscript{261} Since there is no clear measurement of breach of privacy and resulting harm, contextual regulations of behavioral targeting entail the Legislature to consider endless scenarios and the judiciary to interpret the meaning of reasonable expectations under various circumstances.\textsuperscript{262} Although the scheme aims to strike a balance between utility and harm, the difficult task of weighing indicators of risk and instruments for reducing risk under different contextual norms is inevitable.\textsuperscript{263} Nevertheless, there are already established legislation and interpretation of reasonable person standards in torts, administrative, criminal and constitutional law on both state and federal levels that may be used as examples.\textsuperscript{264}

One may also argue that the nature of behavioral targeting essentially warrants almost all first party domains and third parties that conduct online business, research and advertising activities to comply

\textsuperscript{259} See \textit{supra} Part III (detailing privacy problems caused by the use massuse massive aggregated databases).

\textsuperscript{260} See \textit{Ohm, supra} note 29, at 1763 (explaining the benefit of having users actively agree to collection methods).

\textsuperscript{261} See \textit{Moran, supra} note 236, at 1233 (alleging that the reasonable person standard requires extensive judicial interpretation).

\textsuperscript{262} See \textit{Petersen v. Magna Corp.}, 773 N.W.2d 564, 567 (2009) (arguing that it is the role of the court to construe ambiguous statutory terms).

\textsuperscript{263} See \textit{Ohm, supra} note 29, at 1764-68 (exemplifying the tests for assessing the risk of re-identification).

\textsuperscript{264} See \textit{Moran, supra} note 236, at 1233 (guaranteeing a concrete foundation in judiciary’s ability to interpret the reasonable person standard in different areas of law).
with the universal privacy protection.\footnote{265 See Self-Regulatory Principles, supra note 102, at 2 (describing the concept of behavioral targeting).} Although the practice of behavioral targeting is designed to predict Internet user behaviors, it serves two distinct purposes: to allow businesses to deliver tailored advertisement to potential customers and to promote convenient use of the Internet by saving customized personal data.\footnote{266 See Penn, supra note 136, at 602-603 (explaining the role of algorithm in predicting a consumer’s preferences).} Much of the privacy debate surrounding behavioral targeting is on the first purpose.\footnote{267 See supra Part III.C-D (revealing the risks and harms of online behavioral targeting).} Contrary to the second purpose, allowing third parties to access data for the purpose of delivering tailored advertisement often violates users’ reasonable expectations and trusts, and involves high risks of harm in return for minimum utility.\footnote{268 See supra Part III.C-D (exposing the cost of free flow of information in behavioral targeting).} The proposed scheme seeks to regulate behavioral targeting that generates solely profit-making activities on a stricter standard than activities that users reasonably expect and trust to occur under the circumstances.

V. CONCLUSION

Re-identification technology has not only changed our understanding of data privacy, but also called for an update on current privacy regulations. The sectorial concept of PII must be abandoned because it no longer guarantees any meaningful protection to online privacy. This Note suggests a scheme that contextually regulates specific sectors with a focus on online behavioral targeting, supplemented by a universal privacy protection. If the collection, storage, distribution or use of data falls outside of how one would reasonably expect data to be collected, stored, distributed or used, all parties that violate the reasonable expectation must give clear and explicit disclosure and obtain informed consent.

The preventive nature of the proposed scheme allows privacy regulations to always stay ahead of re-identification technology by regulating all types of data that violate the reasonable norms under
which it was collected, stored or used. It offers a new way of balancing utility as well as privacy of data. Its implementation would represent a giant step in revolutionizing privacy regulations as well as behavioral targeting regulations in the United States.