Vanderbilt Journal of Transnational Law

Volume 56
Issue 5 *Issue 5 - November 2023*

Article 4

11-2023

The Identifiability Problem in Transnational Privacy Regulation

Xiaowei Yu
University of Illinois College of Law

The Identifiability Problem in Transnational Privacy Regulation

Xiaowei Yu*

ABSTRACT

Commercial surveillance pervasively compromises data privacy by tracking consumers without meaningful consent or knowledge, yet there is no consensus on when data privacy laws should intervene. The crux lies in the standard of identifiability, which functions as the threshold trigger for when regulation is permissible. Ascertaining the identifiability of information is therefore critical to consumers, companies, and regulators, who must understand, comply with, and implement data privacy laws. As this Article shows, the world's key privacy jurisdictions—the European Union, United States, and China—continue to struggle in similar ways with inadequately defining and inconsistently applying the concept of identifiability.

This Article generalizes from the transnational convergence and refers to it as "the identifiability problem." Recognizing the identifiability problem reveals an overlooked phenomenon across jurisdictions. Besides, it lays a factual foundation for synthesizing international efforts into solving the threshold issue of data privacy law. Moreover, it calls forth a normative inquiry. Whether we can justify the use of identifiability as the threshold for regulation is the first and foremost task prior to revising identifiability or abandoning it.

[*] Post-Doctoral Fellow at the Law & Technology Center, Faculty of Law, the University of Hong Kong. J.S.D., the University of Illinois College of Law. Thank you to Anupam Chander, Heidi Hurd, Robin Kar, Jedidiah Kroncke, Colleen Murphy, Arden Rowell, Chiying Wang, and participants in the HKU Law Tech Talk for helpful comments and suggestions.

I. INTRODUCTION

The Internet knows you. Amazon recommends the exact book that you need for your research project. [1] Instagram sends you advertisements for baby clothes just a few hours after you confirm your pregnancy.[2] Facebook suggests that you add the person you just met at a conference or a party as a friend.[3] These are all possible because of commercial surveillance, which is "the business of collecting, analyzing, and profiting from information about people."[4]

Commercial surveillance has become the core of the digital economy. According to Google's 2020 annual report, over 80 percent of its revenue came from online ads, which are operated by tracking

[1] *See* Kate O'Flaherty, *The Data Game: What Amazon Knows About You and How to Stop It*, GUARDIAN (Feb. 27, 2022, 6:00 PM), https://www.theguardian.com/technology/2022/feb/27/the-data-game-what-amazon-knows-about-you-and-how-to-stop-it [https://perma.cc/ZCX3-6WQX] (archived Aug. 23, 2023).

[2] *See* Gillian Brockell, *Dear Tech Companies, I Don't Want to See Pregnancy Ads After My Child Was Stillborn*, WASH. POST (Dec. 12, 2018, 3:10 PM) https://www.washingtonpost.com/lifestyle/2018/12/12/dear-tech-companies-i-dont-want-see-pregnancy-ads-after-my-child-was-stillborn/ [https://perma.cc/38GN-F8QE] (archived Aug. 23, 2023).

[3] *See* Curtis Silver, *How Facebook's 'People You May Know' Section Just Got Creepier*, FORBES (June 28, 2018, 12:10 PM), https://www.forbes.com/sites/curtissilver/2016/06/28/how-facebooks-people-you-may-know-section-just-got-creepier/?sh=614705505f5a [https://perma.cc/4FBP-FGQ2] (archived Aug. 23, 2023).

[4] *See FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices: Agency Seeks Public Comment on Harms from Business of Collecting, Analyzing, and Monetizing Information About People*, F.T.C. (Aug. 11, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-explores-rules-cracking-down-commercial-surveillance-lax-data-security-practices [https://perma.cc/78TP-8LGC] (archived Aug. 23, 2023).

consumers and targeting their needs.[5] While playing a significant role in the global economy, commercial surveillance puts data privacy in peril. Data privacy is threatened and even violated when companies digitally track consumers without their meaningful consent or knowledge. Unbeknownst to most users of commercial surveillance, companies sell nameless but detailed profiles about consumers to undisclosed third parties. Users have no say in the generation and transformation of their information, nor a voice in who is eligible to access information about their daily lives.

In modern U.S. informational privacy litigation, whether data are identifiable is a crucial issue. If a person wants to stop being tracked by, say, Facebook, and seeks to sue the company for a privacy violation, this person must prove that his or her data are identifiable. Put differently, unless the person establishes that these data exclusively identify him or her, there is no violation of privacy under the law. Identifiability denotes the legal standard that distinguishes identifiable and non-identifiable data.[6] How to define and apply identifiability is a threshold issue for triggering privacy regulation.

Identifiability is notoriously arduous to apply and leads to a serious problem of judicial predictability. For example, U.S. courts have held unique anonymized IDs identifiable in one situation while non-identifiable in another.[7] The legal status of the same piece of data may vary across contexts and lacks certainty. Consequently, the implementation of data privacy is opaque to everyone. Regulators must make extensive efforts to ascertain the legal status of data on a case-by-case basis. Companies must allocate considerable money and human resources to fulfill their obligations. Consumers are confused by the fact that a piece of data is protected in one case but is not protected in the other case; they don't know when they can legally claim their privacy rights and what is required to seek legal protection.

Despite this, more and more jurisdictions adopt identifiability as the gatekeeper of data privacy regulation.[8] The gatekeeping role of

[5] Megan Graham & Jennifer Elias, *How Google's $150 Billion Advertising Business Works*, CNBC (Oct. 13, 2021, 12:52 PM), https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html [https://perma.cc/6ZVB-8ZL5] (archived Aug. 23, 2023).

[6] *See infra* Part VI.B.

[7] *Compare In re* Nickelodeon Consumer Priv. Litig., 2014 WL 3012873, at *19 (D.N.J. July 2, 2014) (holding that an IP address was non-identifiable under the VPPA because IP addresses alone cannot identify a specific person) *with* Yershov v. Gannett Satellite Info. Network, Inc., 820 F.3d 482, 486 (1st Cir. 2016) (holding that an Android ID was identifiable under the VPPA because it was reasonably and foreseeably likely to reveal the plaintiff's identity).

[8] Jurisdictions generally adopt identifiability as a critical condition to trigger privacy regulation. *See* LEE BYGRAVE, DATA PRIVACY LAW, AN INTERNATIONAL PERSPECTIVE 129 (2014) ("Data privacy law generally applies solely to 'personal' data or information."). When more than 140 countries start to legislate the EU-style data privacy standards and laws, the global convergence on defining and applying identifiability is

identifiability refers to it functioning as the threshold. Meeting the standard of identifiability is a sufficient condition for triggering data privacy regulation. Most identifiability studies aim to clarify how a state defines and applies identifiability within one legal context and rarely compare and contrast identifiability in multiple jurisdictions.[9] Such a paradigm overlooks the dual role of identifiability: it is not only a domestic but also a multinational, if not global, gatekeeper to data privacy laws. Traditional discussions are inadequate to grasp a complete understanding of identifiability both as a domestic and multinational gatekeeper. Understanding identifiability from a comparative perspective is much needed to fill the scholarly gap.

This Article pioneers a comparative study on legal discourses over identifiability in the European Union (EU), the United States, and China. These three critical jurisdictions are distinct in many aspects but demonstrate remarkable homogeneity in defining and applying identifiability. All three adopt identifiability as the threshold trigger for their own data privacy regulation, put little effort into defining identifiability, and rely on a presumed relationship between the meaning of identifiability and the identification of a person.

Their applications of identifiability are all enormously uncertain and share at least three challenges that blur the boundary between identifiable and non-identifiable data. First, the same piece of data is non-identifiable to one company but is identifiable to another company ("audience challenge").[10] Second, the same piece of data can be treated as anonymous in the beginning but turn identifiable in a snap second ("re-identification challenge").[11] Third, the same piece of data is non-identifiable in isolation but becomes identifiable in combination with other data ("aggregation challenge").[12]

This Article will dive into this transnational confluence among three jurisdictions. More specifically, this Article will first describe the three legal discourses over identifiability by examining primary legal documents and representative rulings and summarizing their features. It will then compare and contrast discourses. In so doing, this Article will present a comparative picture of how the EU, United States, and China define and apply identifiability. After, this Article will generalize from the comparative depiction to "the identifiability problem," which refers to common difficulties of implementing the data privacy regulation caused by the uncertainty of whether data is identifiable. Such a problem exists everywhere as long as identifiability is the vital threshold of data privacy protections.

becoming inevitable. *See* Graham Greenleaf & Bertil Cottier, *2020 Ends a Decade of 62 New Data Privacy Laws*, 163 PRIV. L. & BUS. INT'L REP. 24, 24–25 (2020).

[9] *See infra* Part II.

[10] *See infra* Part VI.

[11] *Id.*

[12] *Id.*

Articulating the identifiability problem is significant in at least three aspects. First, it reveals a common but overlooked phenomenon in transnational privacy regulation. This does not mean that no one else has ever noticed the problem. People have had the unchecked impression that identifiability is commonly used. But privacy literature provides little insight into how various states define and apply identifiability in a doctrinal sense.[13] This Article fills the gap and takes legal discourses over identifiability seriously from a comparative perspective.

Second, the existence of identifiability problem lays a factual foundation to synthesize global doctrinal efforts to encounter the threshold issue of data privacy. Traditionally, issues of domestic law can only be solved within a certain jurisdiction. The transnational emergence of the identifiability problem breaks this territorial limit to some degree. States can learn from each other about how to solve the problem doctrinally because they share huge doctrinal convergence on identifiability.

Third, recognizing the identifiability problem in transnational privacy regulation points us in a new direction. Before making a choice either in revising identifiability[14] or abandoning it[15], we need to answer a normative question: is it possible to justify the until-now insufficiently justified criterion of identifiability? Put differently, should identifiability be the criterion that triggers privacy protections? An increasing number of jurisdictions take identifiability as the threshold of data privacy laws but no single state expressly provides reasons for its gatekeeping role. To avoid this pitfall, this Article asserts that the first and foremost task is to clarify the relationship between identifiability and data privacy. Such a clarification is the necessary first step towards justifying or changing the role of identifiability. Without establishing the foundation of using identifiability as a primary threshold of data privacy regulation, neither revising nor abandoning identifiability is plausible.

This Article proceeds as follows: Part II will introduce the current landscape of commercial surveillance. Part III will explain the necessity of conducting comparative studies on identifiability. Part IV will describe how the EU, United States, and China define and apply identifiability. Part V will present a sketch of comparison and generalize the identifiability problem in transnational privacy regulation. Part VI will investigate possible solutions to challenges of

[13] *See infra* Part III.

[14] Those who believe identifiability is useful seek to revise identifiability to make it more feasible. *E.g.*, Paul Schwartz and Daniel Solove's PII 2.0 and EU scholars' group privacy theory. *See infra* Part VI.A.

[15] Those who think identifiability is useless call for abandoning it and propose alternative mechanisms. *E.g.*, Paul Ohm's utility approach and Helen Nissenbaum's contextual integrity. *See* Part *infra* VI. B.

applying identifiability and highlight the importance of a normative inquiry into the relationship between identifiability and data privacy.

II. COMMERCIAL SURVEILLANCE

Commercial surveillance is an integral part of the Internet. Companies employ various tracking technologies to collect and analyze consumer data to gain profits. Despite its pervasiveness, commercial surveillance remains mysterious to most consumers. Considering the privacy implications of commercial surveillance, it is essential to begin with an understanding of the significant forms of tracking technologies. This Part lays a foundation for the ensuing discussion by focusing on the three most prominent examples of tracking technologies—cookies as online tracking, smart devices as offline tracking, and algorithms as the next level of tracking.

A. *Cookies as Online Tracking*

Cookies are small text files that websites send to users' browsers and store on users' computers temporarily or permanently.[16] Cookies are used to keep information about users' visits to websites. Among the most well-known forms of information kept are usernames, passwords, frequency of visiting the same websites, and commodities in online shopping carts.[17] In addition to cookies, websites track users' online activities with other similar methods, including web beacons, SDKs, JavaScript, and device identifiers.[18] Through these online trackers, websites are capable of collecting a variety of data, such as IP address, location, operation system, browser, browser language, URLs of visited pages, device identifiers, advertising identifiers, and other usage information.[19]

Websites claim that cookies and other online trackers are crucial to their services. For example, the cookies policy of The New York Times reveals four legitimate purposes of using online trackers: (1) essential operation (e.g., cookies help users to stay logged in); (2) personalized services (e.g., cookies help the website to memorize users'

[16] "Cookies" is an old programming concept that refers to a mechanism to pass data objects between two routines. Netscape, which is a communication corporate, was the first to create "Netscape Cookie" in 1994 worldwide. *See* SIMON ST. LAURENT, COOKIES 15 (1998).

[17] Joanna Geary, *Tracking the Trackers: What are Cookies? An Introduction to Web Tracking*, GUARDIAN (April 23, 2012), https://www.theguardian.com/technology/2012/apr/23/cookies-and-web-tracking-intro [https://perma.cc/E5M3-K4VD] (archived Aug. 23, 2023).

[18] For example, The New York Times employ various tracking methods to track consumers. *See Cookies Policy*, N.Y. TIMES (last updated on Sept. 18, 2021), https://www.nytimes.com/privacy/cookie-policy [https://perma.cc/C8S9-C9D6] (archived Aug. 23, 2023).

[19] *See id.*

choices and settings); (3) performance evaluations (e.g., cookies help the website to record common difficulties of visiting the website); (4) effective marketing/advertising (e.g., cookies contribute to target the specific needs of users and select relevant ads).[20] Such self-disclosures in privacy policies or cookies policies is a popular way for companies to justify their collection and usage of consumer data through online trackers.[21]

B. *Smart Devices as Offline Tracking*

A smart device is an electronic device that connects to the Internet and other devices for data exchange.[22] Based on the interconnectivity, smart devices can fulfill multiple tasks beyond what they were created for. Consider smart mirrors. In addition to reflecting images, smart mirrors can simplify your morning routines by displaying weather, notifications, calendar alerts, news, and TV shows on their surfaces while you brush your teeth or take a shower.[23] You can search Google, check emails, or take selfies and update social media in your bathroom.[24]

Whereas cookies record users' online activities, smart devices collect information about consumers' offline activities. Take Fitbit and Roomba as examples. Fitbit is a popular "fitness and health tracker."[25] For the purpose of measuring and supervising your health status, you need to allow Fitbit to know your body temperature, heart rate, muscle motion, respiration, body and limb motion, affections, desires, and likes.[26] Roomba is iRobot's autonomous vacuum cleaner. Aiming to

[20] *Id.*

[21] There are free templates to generate general privacy policy on the Internet. *See, e.g.*, TERMIFY, https://termify.io/privacy-policy-generator?gclid=EAIaIQobChMIrcOIsePUQIVwcqWCh3T9wLpEAAYASAAEgJcU_D_BwE [https://perma.cc/99LR-FSSJ] (archived Aug. 23, 2023). By reading the privacy policy of *The Times*, we can summarize four parts of a general privacy policy: (1) types of trackers that are used; (2) kinds of data that are collected; (3) purposes of collecting and using these data; (4) legal rights consumers have in different jurisdictions (e.g., the EU citizens and California residents have more legal rights to control their personal data than consumers in other places.). But the legal validity of privacy policy is debatable. *See generally* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013) (arguing that legal solutions to the privacy self-management problem in current privacy policy encounter a consent dilemma that raises its own legal questions).

[22] Manuel Silverio-Fernández, Suresh Renukappa & Subashini Suresh, *What is a Smart Device? – A Conceptualization within the Paradigm of the Internet of Things*, 6 VISUALIZATION ENG'G 1, 8 (2018).

[23] *See, e.g.*, *Makr Mirror Customizable Smart Mirror lets you cater the device to you*, GADGET FLOW, https://thegadgetflow.com/portfolio/customizable-smart-mirror/ [perma.cc/H93M-R7CA] (archived Aug. 23, 2023).

[24] *See, e.g.*, *id.*

[25] *See, e.g.*, *Fitbit.com Updates,* FITBIT https://www.fitbit.com/global/hk/home [https://perma.cc/872D-TMKT] (archived Aug. 23, 2023).

[26] *See, e.g.*, *id.*

clean your home efficiently, Roomba collects "the dimension of a room as well as distances between sofas, tables, lamps and other home furnishings" and maps your physical living environment.[27]

Smart devices track users offline and enable companies to gaze remotely at consumers in their homes, at private times, and during intimate activities. Smart devices are offline trackers that monitor activities outside of cyberspace, including one's conversations, behaviors, and emotions. Like online service providers that use cookies and other online trackers, manufacturers of smart devices justify their collection and usage of consumer data by posting hyperlinks to privacy policies on their websites.[28]

C. *Algorithms as Next-Level Tracking*

Algorithms, also known as big data analyses, have revolutionized the commercial tracking landscape. Supported by the advances in artificial intelligence, algorithms effectively discover patterns of correlations,[29] which help data analysts model typical behaviors[30] by "either identifying individual patterns of behaviour or allocating observed behaviour to a pre-existing category." [31] As a result, algorithms produce mathematical predictions of trends, relationships, and patterns in seemingly random data.[32]

Unlike cookies and smart devices, algorithms do not directly collect information about consumers, be it online or offline. Instead, algorithms analyze data collected by cookies, smart devices, and other emerging technologies and take tracking to the next level. Cookies and smart devices by themselves cannot determine what data are helpful for essential operations, personalized services, marketing, or advertising; it is algorithms that work behind the scenes to process

[27] *See* Jan Wolfe, *Roomba Vacuum Maker iRobot Betting Big on the "Smart" Home*, REUTERS (July 24, 2017), https://www.reuters.com/article/us-irobot-strategy/roomba-vacuum-maker-irobot-betting-big-on-the-smart-home-idUSKBN1A91A5 [https://perma.cc/N8KK-8LEN] (archived Aug. 23, 2023).

[28] For example, Fitbit discloses their data practice in privacy policy to justify their use of consumer data is legitimate. *See, e.g.*, *Fitbit Privacy Policy* (Aug. 16, 2021), https://www.fitbit.com/global/hk/legal/privacy-policy [https://perma.cc/4ELB-TRLT] (archived Aug. 23, 2023) ("Here we describe the privacy practices for our devices, applications, software, websites, APIs, products, and services (the 'Services'). You will learn about the data we collect, how we use it, the controls we give you over your information, and the measures we take to keep it safe.").

[29] Mireille Hildebrandt, *Defining Profiling: A New Type of Knowledge?*, *in* PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINE PERSPECTIVE 17, 19 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

[30] Ann Canhoto & James Backhouse, *General Description of the Process of Behavioral Profiling*, *in* PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINE PERSPECTIVE 47, 47 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

[31] *Id.* at 48.

[32] Bernhard Anrig, Will Browne & Mark Gasson, *The Role of Algorithms in Profiling*, *in* PROFILING THE EUROPEAN CITIZEN: CROSS-DISCIPLINE PERSPECTIVE 65, 65 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

various types of data, interpret the meaning of data, and discover patterns of behaviors.[33] In other words, algorithms are the "brains of tracking" that enable the operation of personalized recommendations, offers, and ads.

Algorithms move the tracking capabilities of commercial surveillance to the next level. With algorithms, tracking becomes inescapable. Saying "no" cannot effectively stop algorithms from watching you. Regardless of what you do, you are watched as long as people around you consent to be watched. [34] At the same time, algorithms make tracking penetrating. Using algorithms enables companies to continuously dig out more new facts about you.[35] These facts may be your innermost thoughts, feelings, desires, or fantasies.[36] Even though algorithms cannot reliably infer human emotions now, interested companies, such as Walmart, are exploring ways to be able to read emotions based on images and other data.[37]

Companies do not disclose the mechanisms of algorithms they use in their privacy policies. Companies keep their proprietary algorithms secret for at least three reasons. The first reason is that algorithms and relevant data practices are trade secrets and are protected for fair competition. [38] Secondly, disclosing algorithms compromises the accuracy of algorithmic outputs. Users may game the system when they know the underlying logic. [39] Thirdly, making algorithms transparent will not increase users' comprehension because algorithms

[33] *See generally* JOHN CHENEY-LIPPOLD, WE ARE DATA: ALGORITHMS AND THE MAKING OF OUR DIGITAL SELVES (2017) (discussing how algorithms interpret our data and define our online identity).

[34] "Only 53 people in Australia installed the This is Your Digital Life app, according to court documents, but it was able to harvest the data of about 311,127 people." Christopher Knans, *Facebook appeal over Cambridge Analytica data rejected by Australian court as 'divorced from reality,'* GUARDIAN (Feb. 6, 2022), https://www.theguardian.com/technology/2022/feb/07/facebook-appeal-over-cambridge-analytica-data-rejected-by-australian-court-as-divorced-from-reality [https://perma.cc/9X7X-6RMV] (archived Sept. 8, 2023).

[35] *See* Sang Ah Kim, *Social Media Algorithms: Why You See What You See*, 2 GEO. L. TECH. REV. 147, 149 (2017) ("During the analysis, the accumulated data can be organized into different categories that each reveal clues about what a user likes to see.")

[36] *See* Melissa Heikkila, *Machines Can Read Your Brain. There's Little That Can Stop Them.*, POLITICO (Aug. 31, 2021, 7:00 AM), https://www.politico.eu/article/machines-brain-neurotechnology-neuroscience-privacy-neurorights-protection/ [https://perma.cc/K99D-C8TJ] (archived Aug. 23, 2023).

[37] Jessica Baron, *Tech Is Already Reading Your Emotions – But Do Algorithms Get It Right?*, FORBES (July 18, 2019, 6:04 PM), https://www.forbes.com/sites /jessicabaron/2019/07/18/tech-is-already-reading-your-emotions-but-do-algorithms-get-it-right/?sh=2af328506fea [https://perma.cc/D2B3-9E62] (archived Aug. 19, 2023).

[38] Charlotte A. Tschider, *Legal Opacity: Artificial Intelligence's Sticky Wicket*, 106 IOWA L. REV. ONLINE 126, 142 (2021).

[39] Ignacio N. Cofone & Katherine J. Strandburg, *Strategic Games and Algorithmic Secrecy*, 64 MCGILL L.J. 635, 647 (2019).

are not intelligible to the majority of the population and even some programmers.[40]

D. *Multilevel Tracking and Privacy Concerns*

Tracking by means of commercial surveillance takes place at multiple levels. Cookies, smart devices, and algorithms manifest at least three types of tracking: online, offline, and next-level tracking. Online tracking monitors individuals' online movements. Offline tracking records individuals' activities disconnected from the Internet. One common limitation of online and offline tracking is that people can avoid them by evading the technologies. Algorithms overcome this constraint by watching individuals through projecting preferences, behaviors, and emotions. Currently, companies are trying to cover as many types of tracking as possible.

Despite their different mechanisms, cookies, smart devices, and algorithms are all tracking technologies that compromise data privacy. Data privacy is threatened and potentially violated when companies digitally track consumers without their knowledge or meaningful consent. These tracking technologies engender privacy concerns worldwide. The industry of commercial surveillance does not only exist in certain regions. Rather, it is a core industry of the digital economy on a global scale.[41] Privacy concerns arise globally when the operation of commercial surveillance is borderless.

III. A COMPARATIVE STUDY OF IDENTIFIABILITY AND BEYOND

Technology companies pervasively compromise data privacy by tracking consumers without meaningful consent, yet there is no consensus on when data privacy laws should intervene. The crux of the problem lies in the standard of identifiability. Identifiability is a legal standard that decides whether data can identify a natural person exclusively. As a general rule, no privacy violation legally occurs unless data are identifiable in the EU, the United States, and China. However, the application of identifiability is utterly challenging. Influenced by many factors, such as data recipients and technological advances, data's identifying capability is ever-changing. Despite this, more and more jurisdictions use identifiability as the pivotal threshold of triggering their domestic data privacy laws. In other words, failure to meet the standard of identifiability will not invoke data privacy regulation at all in many regions.

[40] Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Leaning Algorithms*, 3 BIG DATA & SOC'Y 1, 4 (2016).

[41] *See generally* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019) (examining the centrality of consumer surveillance in the modern digital economy and the moral consequences thereof).

While identifiability plays a boundary role in data privacy regimes, its dual role is seldom discussed. It is not only a domestic, but also a multinational, if not global, gatekeeper to data privacy laws. To depict identifiability as a domestic threshold, a doctrinal analysis of how a particular state defines and applies it within one legal context is necessary. To unveil identifiability as a multinational, or global, gatekeeper, a comparative analysis of how various jurisdictions define and apply it within multiple legal contexts is necessary.

The majority of scholarly discussions focus on the domestic role of identifiability. For instance, in their influential article, *The PII Problem*, Paul Schwartz and Daniel Solove concentrate on the U.S. data privacy jurisprudence and categorize three approaches to apply identifiability in the American data privacy statutes.[42] Likewise, boldly claiming that the EU has "the law of everything" in the data privacy regime, Nadezhda Purtova doctrinally analyzes the standard of identifiability based upon the EU data directives, regulations, guidelines, and the case law of the EU's Court of Justice.[43]

Regarding the comparative role of identifiability, privacy literature provides little insight into the divergence and convergence of identifiability across jurisdictions. Under the current scholarly agenda, conducting identifiability studies from a comparative vantage point seems to be a secondary task. The scant comparative efforts in the literature—such as mentioning the EU's "expansionist" approach to identifiability in *The PII Problem* to function just as a reference point to bolster their framework, PII 2.0[44]—simply serve to support single jurisdiction arguments. In addition, the latest identifiability scholarship pays more attention to the socio-technical aspect than the comparative lens.[45]

While comparatists overlook identifiability specifically, comparative studies are widely recognized in most scholarly and professional subfields of data privacy. Many scholarly efforts are devoted to conducting comparative studies on data privacy topics, including but not limited to privacy rights and obligations, legal mechanisms, and regulatory agencies.[46] Professional organizations,

[42] *See generally* Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814 (2011).

[43] *See generally* Nadezhda Purtova, *The Law of Everything. Broad Concept of Personal data and Future of EU Data Protection Law,* 10 L., INNOVATION & TECH. 40 (2018).

[44] Schwartz & Solove, *supra* note 42, at 1873–75.

[45] *See, e.g.*, Nadezhda Purtova, *From Knowing by Name to Targeting: The Meaning of Identification Under the GDPR*, 12 INT'L DATA PRIV. L. 163, 167 (2022).

[46] *See* Daniel J. Solove, *The Limitations of Privacy Rights,* 98 NOTRE DAME L. REV. 975, 993 (2023); Lindsey Barrett, *Confiding in Con Men: U.S. Privacy Law, the GDPR, and Information Fiduciaries*, 42 SEATTLE UNIV. L. REV. 1057, 1059–60 (2019); Paul M. Schwartz, *The EU-U.S. Privacy Collision: A Turn to Institutions and Procedures*, 126 HARV. L. REV. 1966, 1968 (2013); Alessandra Suuberg, *The View from the*

such as the International Association of Privacy Professionals, provide rich resources to map differences and commonalities between comprehensive data privacy laws across the world.[47] There is no reason to leave out identifiability.

To fill the gap, this Article conducts a comparative study on legal discourses over identifiability in three critical jurisdictions—EU, United States, and China. There are two reasons to choose the EU, United States, and China for comparison. First, these jurisdictions are sufficiently distinct in many aspects.[48] Surveying legal discourses over identifiability in these jurisdictions demonstrates that the convergence over identifiability can transcend vastly heterogeneous jurisdictions. Second, in privacy literature, the importance of comparing EU and U.S. models is well-attested, while China's legal model has been largely under the radar.[49] Understanding China in comparison with the EU and United States will be an efficient way to grasp China's unique characteristics and similarities to the West.

Furthermore, this Article will go beyond a mere comparative study. As a comparative study is by nature descriptive, it does not produce prescriptive results.[50] That a certain measure is commonly used does not necessarily mean that it is a fitting one, let alone the optimal one. In order to mend the gap between descriptive studies and normative judgments, this Article will emphasize the importance of a normative inquiry: What is the normative basis, if any, for adopting identifiability as a common safeguarding concept to privacy regulation? Without reflecting on this normative question, applying identifiability will continue to be elusive.

Crossroads: The European Union's New Data Rules and the Future of U.S. Privacy Law, 16 TUL. J. TECH. & INTELL. PROP. 267, 278 (2013); Paul M. Schwartz & Karl-Nikolaus Peifer, *Prosser's "Privacy" and the German Right of Personality: Are Four Privacy Torts Better than One Unitary Concept?*, 98 CAL. L. REV. 1925, 1929 (2010); James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151, 1155 (2004).

[47] *See Global Comprehensive Privacy Law Mapping Chart*, IAPP (Apr. 2022), https://iapp.org/resources/article/global-comprehensive-privacy-law-mapping-chart/ [https://perma.cc/KC3S-MSWD] (archived Aug. 23, 2023).

[48] *See infra* Part V.

[49] *See* IAPP, *supra* note 47.

[50] Ralf Michaels, *The Functional Method of Comparative Law*, *in* OXFORD HANDBOOK OF COMPARATIVE LAW 339, 374 (2006) ("The Common Core projects look to commonalities among *all* legal orders, but even the fact of commonality (to the extent it exists) does not have intrinsic normative force . . . . The sociologist cannot deduce an 'ought' from an 'is'; comparative material gives no guidelines; even commonality has no independent normative force . . . . [E]quivalence functionalism provides surprisingly limited tools for evaluation.").

IV. LEGAL DISCOURSE OVER IDENTIFIABILITY IN THE EU, U.S., AND CHINA

Part I revealed an inescapable reality: tech companies make massive profits from the business of watching over people and threaten consumer privacy in a novel and unprecedented way. Part II underlined the importance of employing comparative methods in identifiability studies. Part III will first sketch major features of data privacy laws in the EU, United States, and China and then demonstrate how each jurisdiction defines and applies identifiability to trigger regulation on commercial surveillance.

A. *The EU Approach*

The EU has long been regarded as the worldwide leader of legislating for data privacy.[51] In the global arms race to data privacy legislation, the EU plays the role of norm entrepreneurship and provides a global standard for data privacy laws.[52] Data privacy laws in the EU mainly refer to EU's data protection laws, which contain two features.

First, while data protection rights originated from privacy rights, now they are treated as two separate categories of rights.[53] Advocates General once debated heatedly whether there is a distinction between the right to privacy and the right to data protection.[54] In today's

[51] The EU's regulatory power has a significant global influence, which Anu Bradford named the "Brussels Effect." *See* ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD 131 (2020) ("Few regulations have impacted global digital companies or their users more than the EU's 2016 General Data Protection Regulation (GDPR).").

[52] *See generally* Alessandro Martelero, *The Future of Data Protection: Gold Standard vs. Global Standard*, 40 COMPUT. L. & SEC. REV. (2021); Paul M. Schwartz, *Global Data Privacy: The EU Way*, 94 N. Y. U. L. REV. 771, 773 (2019) ("The EU has taken an essential role in shaping how the world thinks about data privacy. Even corporate America draws on EU-centric language in discussing data privacy.").

[53] *Data Protection*, EUR. DATA PROT. SUPERVISOR, https://edps.europa.eu/data-protection/data-protection_en [https://perma.cc/FB3M-R8RT] (archived Aug. 23, 2023) ("The notion of data protection originates from the right to privacy and both are instrumental in preserving and promoting fundamental values and rights; and to exercise other rights and freedoms - such as free speech or the right to assembly."). EU's constitutional document, the Treaty of Lisbon ("EU Treaty"), gives the Charter of Fundamental Rights of the European Union ("EU Charter") the same legal value as the Treaty. The EU Charter enshrines a set of fundamental human rights including the right to data protection and the right to privacy. *See The Treaty of Lisbon,* FACTS SHEETS ON THE EU: EUR. PARL., https://www.europarl.europa.eu/factsheets/en/sheet/5/the-treaty-of-lisbon [https://perma.cc/VR9Y-QHGS] (archived Aug. 23, 2023); *Why do we need the Charter?*, EUR. COMM'N, https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights/why-do-we-need-charter_en [https://perma.cc/PT6E-RARG] (archived Aug. 23, 2023).

[54] Advocates General have an obligation to present opinions with independence and impartiality on cases in the EU court. *See* Philippe Léger, *Law in the European*

Europe, differentiating data protection from privacy is gaining support from scholars[55] and the European Data Protection Supervisor.[56]

Second, EU's data protection laws adopt a rights-based approach. EU citizens are empowered with a set of digital rights (e.g., the right to consent[57] or the right to be forgotten[58]) to actively manage how their data are processed. The underlying philosophy is "foster[ing] self-determination of individuals by granting them enhanced control over their personal data." [59] The EU's data protection laws are comprehensive. Collecting and processing personal data both in the public and private sectors are covered under data protection regulations.

To investigate the EU approach to identifiability, we should start with the concept of Personal Data (PD). PD is a fundamental concept

Union: The Role of the Advocate General, 10 J. LEGIS. STUD. 1, 1 (2004). For example, Advocate General Eleanor Sharpston and Advocate General Pedro Cruz Villalòn consider these two rights distinguishable, while Advocate General Dámaso Ruiz-Jarabo Colomer believed that the right to data protection was subsumed by the right to privacy. *See, e.g.*, Bilyana Petkova & Franziska Boehm, *Profiling and the Essence of the Right to Data Protection*, *in* THE CAMBRIDGE HANDBOOK OF CONSUMER PRIVACY 285, 295 (Evan Selinger, Jules Polonetsky & Omer Tene eds., 2018).

[55] *See generally* Jeanne Pia Mifsud Bonnici, *Exploring the Non-Absolute Nature of the Right to Data Protection*, 28 INT'L REV. L., COMPUT. & TECH. 131, 139 (2014) ("Essentially, both rights (to private life and data protection) are ultimately bound to the right of self-determination: to conduct one's personal life on the one hand and as to use and exposure of personal data, on the other. They are intimately connected but separate and different."). *See also* Orla Lynskey, *Deconstructing Data Protection: The 'Added-Value' of a Right to Data Protection in the EU Legal Order*, 63 INT'L & COMPAR. L.Q. 569, 585 (2014) ("It is possible to think of numerous other day-to-day examples of data processing which falls within the scope of the right to data protection but arguably not privacy."); Juliane Kokott & Christoph Sobotta, *The Distinction Between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR*, 3 INT'L DATA PRIV. L. 222, 225 (2013) ("A first distinction between privacy and data protection lies in the scope of both rights. This begins with the substantive scope, meaning the information covered by the respective right. We have seen that privacy life does not necessarily include all information on identified or identifiable persons. However, data protection covers exactly this information."); Gloria González Fuster & Raphaël Gellert, *The Fundamental Right of Data Protection in the European Union: In Search of an Uncharted Right*, 26 INT'L REV. L., COMPUT. & TECH. 73, 79 (2012) ("As the advent of the EU Charter rendered increasingly untenable the idea that the protection of personal data is only a constituent of the right to privacy, the Court progressively moved towards accepting that personal data protection is a right on its own, but only to describe it as 'closely connected' to the right to privacy.").

[56] EUR. DATA PROT. SUPERVISOR, *supra* note 53 ("Privacy and Data Protection, though connected, are commonly recognised .... as two separate rights.").

[57] *See* Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and the Free Movement of Such Data (General Data Protection Regulation), art. 7, 2016 O.J. (L 119/1) [hereinafter GDPR] (EU).

[58] *Id.* at art. 17.

[59] BRADFORD, *supra* note 51, at 133.

in EU's data protection laws.[60] As a default rule, companies are not allowed to process PD unless they comply with certain legal requirements, whereas they are free to use non-PD. [61] The determination on the legal status of data as PD or non-PD draws the line between regulatory and non-regulatory regimes.[62] Identifiability is one element constituting PD. This section will examine Data Protection Directive, General Data Protection Regulation (GDPR), Article 29 Working Party's opinion, and cases of the Court of Justice of the EU to show how identifiability is defined and applied within the concept of PD.

The EU defines PD in a consistent way. Data Protection Directive 95/46/EC (DPD) refers to "personal data" as: "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."[63]

The latest data protection law, the GDPR, defines PD as:

> [A]ny information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an *identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.*[64]

Both DPD and GDPR define PD as "any information relating to an identified or identifiable natural person."[65] They slightly differ regarding when a person is considered identifiable. The GDPR highlights more types of identifiers that relate to an identifiable

[60] Nadezhda Purtova, *The Law of Everything. Broad Concept of Personal data and Future of EU Data Protection Law,* 10 L., INNOVATION & TECH. 40, 41 (2018) ("The concept 'personal data' determin[es] the material scope of data protection.").

[61] Maria Lilla Montagnani & Mark Verstraete, *What Makes Data Personal?,* 56 U. C. DAVIS 1165, 1177 (2023) ("European data protection law rises and falls with personal data. This is because the GDPR's rights, obligations, and protections only apply to personal data.")

[62] *Id.* at 1169 ("At an implementation level, privacy and data protection statutes depend significantly on an account of personal data to make key normative distinctions—the determination of whether information is personal data distinguishes violations that create liability from innocent disclosures of non-personal information.")

[63] Council Directive 95/46, art. 2, 1995 O.J. (L 281).

[64] *See* GDPR, *supra* note 57, art. 4 (emphasis added).

[65] Council Directive 95/46, *supra* note 63, art. 2; GDPR, *supra* note 57, art. 4.

person.[66] Unfortunately, most elements of PD remain undefined in the DPD and the GDPR. Though the DPD and the GDPR clarify the term "identifiable," such a definition is vague because it is built upon an undefined notion—"identified."

Recital 26 GDPR adopts the reasonable likelihood test to further help judges and regulators to determine when a person is identifiable:

> To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken off all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.[67]

In general, Recitals are not legally binding.[68] But Recitals are considered authoritative interpretations of the GDPR's provisions.[69] For this reason, the reasonable likelihood test, adopted by Recital 26 GDPR, should be applied by judges and regulators.

Recital 26 GDPR points out two considerations. First, whether a natural person is identifiable depends upon the objective possibility of identification.[70] Regulators must consider all the means that are reasonably likely to be used and all objective factors around the means.[71] Even if companies do not have the intention of singling out a natural person, they are processing PD as long as they objectively have the ability to identify a person.[72] Second, the agent of fulfilling

[66] The Directive 95 points out identification numbers, while the GDPR has a broader category: in addition to identification numbers, the GDPR contains names, location data, and online identifiers. *See* Council Directive 95/46, *supra* note 63, art. 2; GDPR, *supra* note 57, art. 4. Recital 30 of the GDPR explains that online identifiers can include "internet protocol addresses, cookie identifiers [and] other identifiers such as radio frequency identification tags." GDPR, *supra* note 57, recital 30.

[67] *See* GDPR, *supra* note 57, recital 26.

[68] *See* Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 194 (2019).

[69] *See* Frederik J. Zuiderveen Borgesius, *Singling Out people Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation*, 32 COMPUT. L. & SEC. REV. 256, 264 (2016) ("Recitals do not have the same legal weight as the provisions of a directive. Nevertheless, when interpreting the provisions of a directive, the Court of Justice of the European Union often considers the recitals. The Court also refers to recitals in data protection cases.").

[70] GDPR, *supra* note 57, recital 26.

[71] *Id.*

[72] Maria Lilla Montagnani & Mark Verstraete, *What Makes Data Personal?*, 56 U. C. DAVIS 1165, 1178–9 (2023) ("Put more concretely, even when the data controller

identification is not limited to data controllers.[73] If anyone on the earth might be capable of identifying an EU citizen in a specific case, the standard of identifiability is met.[74] In light of these two considerations, applying the reasonable likelihood test brings about an extremely broad scope of PD.

In addition to broadness, the reasonable likelihood test causes enormous uncertainty.[75] In essence, the reasonable likelihood test is context-dependent. The objective possibility of identification varies from case to case, which makes PD a flexible and dynamic concept. Besides, technological advances lead to instant change on the legal status of the same piece of data. Since technology is a major factor in deciding the identifying capacity of data,[76] its ever-changing nature inevitably leads to a fluid scope of PD.

The Article 29 Working Party (WP 29), an independent EU advisory board on EU data protection before 2018,[77] gave a non-binding opinion on understanding the concept of PD for the sake of consistent implementation of data protection rules in all the EU member states.[78] Despite being non-binding, WP 29's opinion has "pervasive authority" to apply PD and provides a structural analysis of PD in a doctrinal sense.[79]

has no interest or ability to re-identifying information, it will still be personal data if there are technical measures that allow re-identification.")

[73] GDPR, *supra* note 57, recital 26.

[74] Worku Gedefa Urgessa, *The Protective Capacity of the Criterion of Identifiability under EU Data Protection Law*, 2 EUR. DATA PROT. L. REV. 521, 529 (2016) ("It lays down that, legally decisive is not just the ability of the controller to link a person to data but any person's ability to do so. The implication is that an individual is identifiable if she can be identified, directly or indirectly, by *anybody*.").

[75] *Id.* at 528 (2016) ("Even though, Recital 26 in the preamble to the GDPR makes it clear that identification by *anyone* counts for the purpose of identifiability.... it still creates significant uncertainty for data controllers to determine when data is not identifiable to them.").

[76] *See* GDPR, *supra* note 57, recital 26.

[77] The Article 29 Working Party was established by the DPD and dealt with data privacy issues until 25 May 2018. After that, the European Data Protection Board was established by the GDPR to replace the Article 29 Working Party. *See Legacy: Art. 29 Working Party*, EURO. DATA PROTECTION BOARD, https://edpb.europa.eu/about-edpb/who-we-are/legacy-art-29-working-party_en (last visited October 27, 2023).

[78] *Opinion 4/2007 of the Article 29 Working Party on the "Concept of Personal Data,"* at 3, WP 136 (June 20, 2007) [hereinafter Article 29 Working Party].

[79] Purtova, *supra* note 43, at 43. *See also* Purtova, *supra* note 45, at 173–74 (2022) ("On the one hand, it concerns the concept of personal data in the old DPD and not the GDPR, the Article 29 Working Party itself no longer exists and is substituted by a new advisory authority—the European Data Protection Board (EDPB). Shortly after coming to existence, this functional equivalent of the Article 29 Working Party endorsed a number of Article 29 Working Party opinions, yet WP136 is not among these. On the other hand, an argument can be made that the opinion retained its significance also under the GDPR, since the concept of personal data has not undergone significant changes. While in future the EDPB may choose to issue its own GDPR-specific guidelines on the concept of personal data and take a different view on what identification means, it has not done so yet and its work programme for 2021–22 has given priority to other key data protection concepts such as legitimate interest.").

WP 29 divides the definition of PD into four elements: (1) any information; (2) relating to; (3) identified or identifiable; and (4) natural person.[80] Based upon WP 29, applying PD should contain four issues: (1) whether the data at issue is "information"; (2) whether the data at issue is "relating to" the data subject; (3) whether data subject is "identified" or "identifiable"; and (4) whether the data subject is "a natural person." The second and third issues are pertinent to the application of identifiability. [81] When a person is identified or identifiable by a piece of or a combination of data, the criterion of identifiability is met.[82]

Specifically, the term "any information" signals that the legislature intends a wide interpretation of PD.[83] Be it objective or subjective, true or false, private or professional, photographical or acoustic, any kind of information may be considered PD.[84] The element "relating to" describes the relationship between data and an individual.[85] This element requires a "content," "purpose," or a "result" element to be present.[86] Under this view, non-PD, such as weather data, becomes PD when it is used for the purpose of influencing an individual's behavior or when the result of its usage impacts an individual's behaviors. [87] Regarding the third element, a person is "identified" when data can "distinguishably" identify him or her from others, while a person is "identifiable" when data can possibly single out a person, even though the person is not yet identified.[88] Lastly, the term "a natural person" limits PD to data of human beings and excludes corporations or organizations.[89]

Concerning identifiability, WP 29 also holds that the identifying capacity of data is not fixed but varies across contexts.[90] For example,

[80] Article 29 Working Party, *supra* note 78, at 6.

[81] Nazezhda Purtova claims that the "relating to" element refers to "a relevant relationship between information and an individual," which is different from "a relevant possibility of identification." Purtova, *supra* note 43, at 44. On the contrary, Lee Bygrave thinks both elements indicate the standard of identifiability. BYGRAVE, *supra* note 8, at 129–130 ("From these definitions, we can discern two cumulative conditions for data to be 'personal': first, the data must relate to or concern a person; secondly, the data must enable the identification of such a person. .... [T]he first condition can be embraced by the second.... In other words, the basic criterion appearing in these definitions is that of identifiability—that is, the potential of data to enable identification of a person.").

[82] Purtova, *supra* note 43, at 46.

[83] Article 29 Working Party, *supra* note 78, at 6 ("The term 'any information'....clearly signals the wiliness of the legislator to design a broad concept of personal data.").

[84] Article 29 Working Party, *supra* note 78, at 6–7.

[85] Article 29 Working Party, *supra* note 78, at 9.

[86] *Id.* at 10.

[87] Purtova, *supra* note 43, at 58.

[88] Article 29 Working Party, *supra* note 78, at 12.

[89] *Id.* at 21.

[90] Article 29 Working Party, *supra* note 78, at 13 ("[T]he extent to which certain identifiers are sufficient to achieve identification is something dependent on the context of the particular situation.").

a common name, such as Michael Green, cannot distinguish a person from others in a national census, while the same piece of data can easily single out an employee in a small company. Contrarily, a combination of data, like "a girl wearing Hollister sweater," can identify a student in a contract law class, but it is anonymous information in a party of Hollister fans. Therefore, even for data that has a strong ability to link a person, we can never be absolutely certain that they make people identified in every situation.

The Court of Justice of the European Union (CJEU) has interpreted the scope of PD in many cases. Judgments of the CJEU reveal a broad approach to identifiability.[91] Take IP addresses as an illustration. The CJEU consistently holds that IP addresses are PD. In 2011, the *Scarlet* case held that static IP addresses were PD because they allowed the concerned users to be precisely identified by the Internet service provider (ISP).[92] The *Bonnier Audio AB* case further confirmed the same holding.[93] In 2016, the *Breyer* case held that a dynamic IP address, which changed with each new Internet connection, was PD because the website was reasonably likely to identify users with additional information.[94] Even though the CJEU's judgments are context-dependent, they deliver a message that IP addresses are considered PD in most cases. [95] The European Commission sends the same message by listing IP addresses as examples of PD.[96]

[91] *Worten* held that working time data, including daily work periods and rest periods, constituted PD. *See* C-342/12, Worten — Equipamentos para o Lar SA v. Autoridade para as Condições de Trabalho (ACT), 2013 CJEU Third Chamber, at 19. *IPI* determined that data collected by private detectives were PD. *See* C-473/12, Institut professionnel des agents immobiliers (IPI) v. Geoffrey Englebert, 2013 CJEU Third Chamber. The court in *Ryneš* judged that the image of a person recorded by a camera was PD because it was possible to identify the natural person. *See* C-212/13, František Ryneš v. Úřad pro ochranu osobních údajů, 2014 CJEU Fourth Chamber. *Bara* decided that taxes were PD. *See* C-201/14, Smaranda Bara v. Casa Naţională de Asigurări de Sănătate and Others, 2015 CJEU Third Chamber.

[92] C-70/10, Scarlet Extended SA v. Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM), 2011 CJEU Third Chamber, at 51 ("Those [IP] addresses are protected personal data because they allow those users to be precisely identified.").

[93] C-461/10, Bonnier Audio AB v. Perfect Communication Sweden AB, 2012 CJEU Third Chamber, at 52.

[94] C-582/14, Patrick Breyer v. Bundestrpublik Deutschland, ECLI:EU:C:2016:779 ("Having regard to all the foregoing considerations, the answer to the first question is that Article 2(a) of Directive 95/46 must be interpreted as meaning that a dynamic IP address registered by an online media services provider when a person accesses a website that the provider makes accessible to the public constitutes personal data within the meaning of that provision, in relation to that provider, where the latter has the legal means which enable it to identify the data subject with additional data which the internet service provider has about that person.")

[95] *See* C-342/12, *Worten*; C-473/12, *IPI*; C-212/13, *Ryneš*; C-201/14, *Bara*; C-70/10, *Scarlet Extended*; C-461/10, *Bonnier Audio*; C-582/14, *Breyer*.

[96] *What is personal data?*, EUR. COMM'N, https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en [https://perma.cc/F69H-F2W9] (archived Aug. 14, 2023).

The EU approach to identifiability is extremely broad. As long as anyone on the earth, not just particular data controllers, has the ability to identify a person in a specific case, these data are PD and trigger the EU's data protection regulation. At the same time, the EU approach to identifiability leads to enormous uncertainty. Since applying identifiability is context-dependent, the identifying capacity of data is not fixed. Technological advances can transform a non-PD into a PD in a second.

B. *The US Approach*

The U.S. informational privacy laws have three features. First, unlike its EU counterpart, data protection is not a fundamental human right in the U.S. legal system.[97] Instead, the United States uses informational privacy, a sub-category of privacy, to instruct how companies can legitimately handle data about U.S. consumers.[98] In other words, regulating commercial surveillance is primarily an issue of privacy laws.

Second, unlike the EU's rights-based approach, the United States has much confidence in the "market's invisible hand."[99] The market approach steps up to protect "consumer interest" in the market.[100] The underlying rationale is that American consumers' data are free to be collected, processed, and disseminated by companies in the absence of a law addressing market failures, such as unfairness and deceptions.[101] The passage of the California Consumer Privacy Act (CCPA) in 2018 paved a way for a rights-based approach; that is, empowering California consumers with a set of digital rights to actively manage how their data is processed by businesses.[102]

[97] Paul M. Schwartz & Karl-Nikolaes Peifer, *Transatlantic Data Privacy Law*, 106 GEO. L.J. 115, 132 ("Where the EU views its [data protection] laws as reflecting and making concrete the broader mandates of a fundamental privacy right, the United States anchors its information privacy in the market place.").

[98] *Id.* ("Unlike the EU's data subject, U.S. law does not equip the privacy consumer with fundamental constitutional rights; rather, she participates in a series of free exchanges involving her personal information....Personal information is another commodity in the market, and human flourishing is furthered to the extent that the individual can maximize her preferences regarding data trades. The focus of information privacy law in the United States is policing fairness in exchanges of personal data.").

[99] Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815, 816 (1999).

[100] Schwartz, *supra* note 52, at 773 ("In contrast, the U.S. legal system views information privacy as based largely on a consumer interest.").

[101] *See id.* ("It situates individuals in a data marketplace in which they are to be free to engage in data exchanges, and the law is to police data trades for unfairness, deceptions, and other market failures.").

[102] Anupam Chander, Margot E. Kaminski & William McGeveran, *Catalyzing Privacy Law*, 105 MINN. L. REV. 1733, 1738 (2021) ("Until the CCPA, no state or federal statute in the United States imposed privacy protections across all industry sectors and technologies in the manner that European data protection law had done for decades.").

Third, U.S. informational privacy laws should be understood both horizontally and vertically. One prominent horizontal feature concerns the sector-based approach. As sector-based regulation does not impose uniform requirements, privacy interests are protected only when relevant sectors or fields have privacy legislations. For example, the Video Privacy Protection Act (VPPA) secures privacy interests in the rental or purchase of videotapes or similar audio-visual materials but does not provide general protections or extend to other sectors such as health care or financial services.[103] One well-known drawback of sector-based legislation is that unregulated sectors or fields may become loopholes for informational privacy protections. The sector-based regulatory landscape began to change with the CCPA, which is the first comprehensive and cross-sector data privacy law in the United States.[104]

Vertically, federal and state legislators have relatively independent sovereignty to legislate for informational privacy. In the second half of the twentieth century, due to growing fears of the computer's ability to collect and search personal data, Congress passed several federal laws protecting privacy in various sectors, such as the Family Educational Rights and Privacy Act (FERPA) of 1974[105] and the Cable Communications Policy Act (Cable Act) of 1984.[106] In today's United States, there is no comprehensive data privacy law at the federal level. At the state level, most states lack "omnibus data protection laws" within their own jurisdictions.[107] Only in the last few years have California, Virginia, Colorado, and Utah begun the trend of legislating for informational privacy at state level.[108] The vertical structure is not challenged by state comprehensive data privacy laws.

To trigger informational privacy regulation, data must be Personally Identifiable Information (PII) or Personal Information (PI).[109] Regardless of the different names, PII and PI are substantially

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

[104] Chander, Kaminski & McGevern, *supra* note 102 ("Until the CCPA, no state or federal statute in the United States imposed privacy protections across all industry sectors and technologies in the manner that European data protection law had done for decades.").

[105] *See* 20 U.S.C. § 1232g(b)(2).

[106] *See* Pub. L. No. 98-549, 98 Stat. 2779 (1984).

[107] Paul M. Schwartz, *Privacy and Participation: Personal Information and Public Sector Regulation in the United States*, 80 IOWA L. REV. 553, 605 (1995).

[108] *See, e.g.*, California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.140 (West 2022); Virginia Consumer Data Protection Act of 2021 § 59.1-574; Colorado Privacy Act of 2021 § 6-1-1305(4); Utah Consumer Privacy Act of 2022 § 13-61-101.

[109] At the federal level, Family Educational Rights and Privacy Act (FERPA) of 1974 was the first statute to use PII as the legal threshold to regulate releasing or accessing educational records. *See* 20 U.S.C. § 1232g(b)(2). Cable Communications Policy Act ("Cable Act") of 1984 first considered Fair Information Practices (FIPs) as legal obligations on collecting or processing PII. *See* Pub. L. No. 98-549, 98 Stat. 2779 (1984). At the states' level, California Consumer Privacy Act (CCPA) protects PI and offers

determined by the criterion of identifiability. There are at least four diverging definitions of PII: the tautological approach, the non-public approach, the specific-type approach, and the broad approach. Legal scholars Paul Schwartz and Daniel Solove developed the first three, and the last one derives from the CCPA.[110] At present, some federal statutes adopt the first three definitions and apply identifiability in either an uncertain or restrictive sense.[111] Some newly emerging state statutes adopt the fourth definition and apply identifiability in an uncertain but relatively broad way.[112]

The tautological approach simply treats PII as information that identifies a person. It is so named because it fails to explain the meaning of "identify."[113] The VPPA defines PII as a category that "includes information which identifies a person as having requested or obtained specific video materials or services from a video tape service provider."[114] PII under the VPPA contains three elements: (1) information (2) that identifies (3) a person. The VPPA leaves the first two elements unelaborated and qualifies only the third. Due to the sector-based feature, the main interest of informational privacy legislation is to clarify whom the statute protects. Under the VPPA, only a person in a consumer–vendor relationship with the service provider, once identified, is protected. As a result, people purchasing a YouTube membership, for instance, do not fall under the statute. Regarding "information" and "identify," the statute provides little legislative explanation except their literal meaning.[115]

The lack of qualifiers on "information" and "identify" directly leads to the uncertainty of applying identifiability in the courts. Judges are struggling with applying identifiability under the VPPA. A series of rulings between 2014 and 2017 demonstrate judges' swing attitudes over their understanding of "identify." In 2014, the court in *In re Hulu Privacy Litigation* held that PII only referred to information that can by itself identify a person, such as names.[116] This judgment excluded the possibility of identification based on a combination of data. Many

comprehensive legal protections on consumer privacy. *See* CIV. § 1798.140. Illinois Biometrics Information Privacy Act restricts flows of PII in biometric information. The biometric information is an individual's biometric identifier that is used to identify an individual. *See* 740 ILL. COMP. STAT. ANN. 14/10.

[110] *See generally* Schwartz & Solove, *supra* note 42. The broad approach in the CCPA is one example of EU's influence on the American data privacy legislation. Schwartz, *supra* note 99, at 816–18.

[111] *See supra* discussions on the VPPA, the GLBA, and the COPPA.

[112] *See supra* discussions on the CPPA.

[113] Schwartz & Solove, *supra* note 42, at 1829.

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

[115] *See id.*

[116] *See In re* Hulu Priv. Litig., No. C 11-03764 LB, 2014 WL 1724344 (N.D. Cal. Apr. 28, 2014).

cases around 2014 and 2015 followed this ruling.[117] In 2015, the tide began to shift. The First Circuit, in *Yershov v. Gannett Satellite Info. Network, Inc.* rejected the approach taken in *Hulu*. The court concluded that PII extended beyond a person's name and included "information reasonably and foreseeably likely to reveal which . . . videos [the plaintiff] has obtained." [118] Based on this ruling, any unique identifier—including a smartphone ID and a phone's GPS coordinates—was PII under the VPPA.[119] Conversely, in 2016, the Third Circuit, in *In re Nickelodeon*, proposed an "ordinary person" test, which defined PII as "the kind of information that would readily permit an ordinary person to identify a specific individual's video-watching behavior.[120] In 2017, the California district court in *In re Vizio, Inc., Consumer Privacy Litigation* reemphasized the *Yershov* test and in its vague language seemed to expand the scope of PII.[121]

The non-public approach treats PII as information that is not publicly accessible. The Gramm-Leach-Bliley Act (GLBA) defines PII as non-public personal information (NPI). GLBA refers to NPI as information that is "(i) provided by a consumer to a financial institution; (ii) resulting from any transaction with the consumer or any service performed for the consumer; or (iii) otherwise obtained by the financial institution."[122] NPI consists of information from one of three sources. The first source is consumers themselves. The Federal Trade Commission (FTC) illustrates information provided by consumers with a list, such as "name, address, income, Social Security number, or other information on an application."[123] The second source is transactions between consumers and financial institutions. According to the FTC, NPI from this source covers "the fact that an individual is your consumer or customer, account numbers, payment history, loan or deposit balances, and credit or debit card purchases."[124] The third source is third-party institutions, such as "information from court records or from a consumer report." [125]

The definition of NPI heavily relies on the sources of information. The statute takes for granted that information from specific sources by

[117] *See, e.g.*, Robinson v. Disney Online, 152 F. Supp. 3d 176, 182–83 (S.D.N.Y. 2015); Eichenberger v. ESPN, Inc., No. 14-cv-463 (TSZ), 2015 WL 7252985 at *5–6 (W.D. Wash. May 7, 2015).

[118] *See* Yershov v. Gannett Satellite Info. Network, Inc., 104 F. Supp. 3d 135, 147–148 (D. Mass. 2015), *rev'd in part on other grounds*, 820 F.3d 482, 486 (1st Cir. 2016).

[119] *See id.*

[120] *In re* Nickelodeon, 827 F.3d 262, 290 (3rd Cir. 2016).

[121] *In re* Vizio, Inc., Consumer Priv. Litig., 238 F. Supp. 3d 1204, 1225 (C.D. Cal. 2017).

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

[123] *How to Comply with the Privacy of Consumer Financial Information Rule of the Gramm-Leach-Bliley Act*, FED. TRADE COMM'N, https://www.ftc.gov/business-guidance/resources/how-comply-privacy-consumer-financial-information-rule-gramm-leach-bliley-act [https://perma.cc/DS6B-9VU3] (Aug. 15, 2023).

[124] *Id.*

[125] *Id.*

definition is personally identifiable, so much so it does not warrant an explanation. The statute presumes "non-public" means "information not found within the public domain" and neglects to define it.[126] The FTC gives examples of publicly accessible data under the GLBA:

- Information is generally made lawfully available to the public by the financial institutions.
- Federal, state, or local government records made available to the public, such as the fact that an individual has a mortgage with a particular financial institution.
- Information that is in widely distributed media like telephone books, newspapers, and websites that are available to the general public on an unrestricted basis, even if the site requires a password or fee for access.[127]

The non-public approach considers another condition of PII. A piece of publicly available data is not PII. Consider telephone numbers. Telephone numbers are PII if provided by consumers. But these data lose their non-public character and therefore are non-PI when consumers also allow them to be listed in a public phone directory.[128]

There are two points of confusion underneath the non-public approach. First, whether data are publicly accessible and whether data are personally identifiable are independent from each other. Emphasizing PII's non-public feature does not address what data constitute PII. Hence, the GLBA fails to give explicit ways to define and apply identifiability, and this causes uncertainty. Second, defining PII as non-public personal information may lead to a misconception. That is, privacy interests in the public domain do not deserve legal protections. Plenty of precedents indicate the contrary. The Supreme Court's Fourth Amendment cases demonstrate that where information is collected is not a decisive factor in determining whether the law should intervene. For example, *Katz v. United States* held that telephone conversations in a public phone booth contains privacy interests and are protected by the Fourth Amendment because "[t]he Fourth Amendment protects people, not places." [129] As a result, highlighting the non-public character of PII could be misleading to some degree.

The specific-types approach treats PII as a list of specific types of data. Only those data on the list are regulated. The Children's Online

[126] Schwartz & Solove, *supra* note 42, at 1830.

[127] FED. TRADE COMM'N, *supra* note 123.

[128] *E.g.*, Dunmire v. Morgan Stanley DW, Inc., 475 F.3d 956, 961 (8th Cir. 2007) ("[Customer's] financial information lost its nonpublic character for purposes of GLBA when he filed reparations complaint disclosing his financial situation with Commodities Futures Trading Commission (CFTC).").

[129] 389 U.S. 347, 351 (1967).

Privacy Protection Act (COPPA) regulates certain kinds of data that are collected online, including first and last names, physical addresses, social security numbers, telephone numbers, and email addresses.[130] In 2000, the FTC issued a rule to expand the list of PII under COPPA to include persistent identifiers such as cookies and IP addresses that can track a user across time and across websites.[131]

The specific-types interpretation is convenient because it provides concrete guidance to judges and regulators. Only listed types of data are PII and, thus, require regulation. However, the specific-types approach is limited because it relies on a pre-determined list of finite categories. This backward-looking approach allows little room for judges and regulators to adapt to new and developing technologies. As the approach offers no method to determine whether a specific piece of data belongs to the list,[132] the PII list tends to be either too narrow or outdated.

In addition to the aforementioned three interpretations, a fourth approach—the broad approach—appeared in the United States after Schwartz and Solove published their 2011 article, *The PII Problem*.[133] This approach is broad because it renders more data legally protected. The CCPA is the first state data privacy legislation in the United States to embrace the broad approach to identifiability.[134] As Anupam Chander, Margot Kaminski, and William McGeveran observe, CCPA defines PI, which is interchangeable with PII, in a quite broad way that is "far beyond most existing U.S. privacy laws."[135] Under the CCPA, PI is "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household." [136]

This broad approach has two weaknesses. First, the relationship between data and data subjects, such as "identify," "relate to," "describe," "be reasonably capable of being associated with," and "could reasonably be linked with," remains vague and uncertain as the CCPA does not further spell it out. Second, the statute neglects to define what a household is and under what circumstances a household is identified or identifiable. Applying identifiability under CCPA remains uncertain.

[130] Children's Online Privacy Protection Act of 1998, 15 U.S.C. § 6501 (8).

[131] 16 C.F.R. § 312.2 (2011).

[132] *See* Schwartz & Solove, *supra* note 42, at 1871.

[133] *See id.*

[134] The CCPA is often recognized as a follower of the GDPR. For example, Paul Schwartz believes that the CCPA is inspired by the GDPR and adopts the global standard set up by the GDPR. *See* Schwartz, *supra* note 99. Anupam Chander, Margot Kaminski, and William McGeveran hold an opposite opinion. They believe that the CCPA is fundamentally different from the GDPR and influences the U.S. privacy laws. *See* Chander, Kaminski & McGeveran, *supra* note 102.

[135] *See* Chander, Kaminski & McGeveran, *supra* note 102, at 1750.

[136] CAL. CIV. CODE § 1798.140 (West 2022).

Using the broad approach, the CCPA provides a non-exhaustive list of data that constitutes PI:

- Identifiers such as a real name, alias, postal address, unique personal identifier, online identifier, internet protocol address, email address, account name, social security number, driver's license number, passport number, or other similar identifiers
- Characteristics of protected classification, such as age, gender, race, or religion
- Commercial information, including records of personal property, products or services purchased, obtained, or considered, or other purchasing or consuming histories or tendencies
- Biometric information
- Internet activity information, including, but not limited to, browsing history, search history, and information regarding a consumer's interaction with an internet website, application, or advertisement
- Geolocation data
- Audio, electronic, visual, thermal, olfactory, or similar information
- Professional or employment information
- Education information
- Inferences drawn from any of the information identified in this subdivision to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes.[137]

CCPA shares one similarity with GDPR. That is, applying identifiability is context dependent. In responding to initial comments from the public, the California Attorney General stated that whether data is PI is "a fact-specific and contextual determination."[138]

CCPA differs from GDPR in at least two aspects. Unlike the EU approach, attempts or intentions to identify a person matters to decide the scope of PD.[139] A business can protect itself from liability by proving that it "makes no attempt to reidentify the information."[140] Besides, neither the CCPA's nor the California Attorney General's

[137] *Id.* § 1798.140(v)(1).

[138] *Final Statement of Reasons Appendix A: Summary and Response to Comments Submitted During 45-Day Period*, Row 15 https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/ccpa-fsor-appendix-a.pdf [https://perma.cc/2MJ9-MW8G] (archived Sept. 9, 2023).

[139] *Id.* § 1798.140(v)(3) ("'Personal information' does not include consumer information that is deidentified or aggregate consumer information").

[140] *Id.* § 1798.140(h)(4).

responses explicitly expand the agent of fulfilling identification from the business involved to anyone on the earth. This is also different from the EU approach. As state legislation, the CCPA merely represents one approach to identifiability in the United States. To be precise, the above-mentioned similarities and differences between the CCPA and GDPR are a comparison between California's and the EU's approach. Nevertheless, investigating CCPA is still valuable because it is a pivotal U.S. example and worthy of comparison to the EU.

American informational privacy laws have diverse definitions of PII. The tautological approach neglects to qualify "identify" and causes uncertainty in the application of identifiability. The non-public approach assumes information from specific sources are personally identifiable without explanations and justifications. This also causes uncertainty in applying identifiability. The specific-type approach restrictively applies identifiability and confirms limited types of data as PII. Lastly, the broad approach applies identifiability in a more extensive and case-by-case manner. Uncertainty arises when applying identifiability is context-dependent.

C. *China's Approach*

Data privacy laws in China mainly include data security laws, the Chinese Civil Code, and the Personal Information Protection Law (PIPL). There are two key characteristics of China's data privacy laws.

First, instead of having data protection, China's legislators recognized an independent area of law as personal information protection. Whereas the EU considers data protection to be a matter of fundamental human rights, China considers personal information protection to be a personal interest. After the Chinese Civil Code took effect on January 1, 2021, protecting personal information became an important interest related to but distinct from the civil right to privacy.[141] The Chinese Civil Code distinguishes between private and personal information and indicates that they should be governed by different provisions.[142] Private information is protected by the right to privacy under the Chinese Civil Code, while personal information is protected by the PIPL, which was passed and took effect in 2021.[143] The PIPL is the first comprehensive data privacy law in China. It

[141] *See* Zhonghua Renming Gongheguo Mingfadian (中华人民共和国民法典) [The Civil Code of the People's Republic of China] (promulgated by the Standing Comm. Nat'l People's Cong., May 28, 2020, effective Jan. 1, 2021).

[142] *See id.* art. 1034 ("The provisions on the right to privacy, or, in the absence of which, the provisions on the protection of personal information, shall be applied to the private personal information.").

[143] Zhonghua Renming Gongheguo Geren Xinxi Baohu Fa (中国人民共和国个人信息保护法)[Personal Information Protection Law of the People's Republic of China] (promulgated by the Standing Comm. Nat'l People's Cong., Aug. 20, 2021, effective Nov. 1, 2021).

instructs both the public and private sectors to legitimately handle the personal information of China's citizens.

Second, personal information protection is also an important interest to security. Security has long been treated as an overarching value of Internet governance in China.[144] In 2012, the Decision on Strengthening Information Protection on Networks stated that preventing illegal collections and exchanges of personal information is an important aspect of safeguarding national security and social order, as well as protecting the lawful interests of citizens.[145] Similarly, the Cybersecurity Law of 2016, as the first national legislation for cybersecurity in China, required network operators to obtain consent from individuals to collect and use their personal information.[146] The Cybersecurity Law also secured individuals the right to delete and correct their personal information.[147] The Data Security Law of 2021 once again highlighted that privacy and personal information protection are important to data security.[148] Despite the undivided

[144] Bo Zhao & Yang Feng, *Mapping the Development of China's Data Protection Law: Major Actors, Core Values, and Shifting Power Relations*, 40 COMPUT. L. & SEC. REV. 1, 4 (2021).

[145] *See* Guanyu Jiaqiang Wangluo Xinxi Baohu de Jueding (关于加强网络信息保护的决定) [Decision Concerning Strengthening Network Information Protection] (promulgated by the Standing Comm. Nat'l People's Cong., Dec. 28, 2012, effective Dec. 28, 2012).

[146] Zhonghua Renmin Gongheguo Wangluo Anquan Fa (中华人民共和国网络安全法) [The Cybersecurity Law of the People's Republic of China] (promulgated by the Standing Comm. Nat'l People's Cong., Nov. 7, 2016, effective June 1, 2017), art. 40, CAC (Nov. 7, 2016, 19:38) ("Network operators collecting and using personal information shall abide by the principles of legality, propriety, and necessity; they shall publish rules for collection and use, explicitly stating the purposes, means, and scope for collecting or using information, and obtain the consent of the persons whose data is gathered. Network operators must not gather personal information unrelated to the services they provide; must not violate the provisions of laws, administrative regulations or agreements between the parties to gather or use personal information; and shall follow the provisions of law, administrative regulations, and agreements with users to process personal information they have stored.").

[147] *Id.* at art. 43 ("Where individuals discover that network operators have violated the provisions of laws, administrative regulations, or agreements between the parties to gather or use their personal information, they have the right to demand the network operators delete their personal information; where discovering that personal information gathered or stored by network operators has errors, they have the right to demand the network operators make corrections. Network operators shall employ measures for deletions and corrections.").

[148] *See* Zhonghua Renming Gongheguo Shuju Anquan Fa(中华人民共和国数据安全法)[Data Security Law of the People's Republic of China] (promulgated by the Standing Comm. Nat'l People's Cong., June 10, 2021, effective Sep. 1, 2021), Ch. V, art. 38 ("Where state organs need to collect or use data to perform their statutory duties, they shall collect or use data within the scope as needed for performance of their statutory duties and under the conditions and procedures provided by laws and administrative regulations. They shall, in accordance with the law, preserve the confidentiality of the data accessed in the course of performing their duties, such as personal privacy, personal information, trade secrets, and confidential business information, and shall not divulge such data or illegally provide them to others.").

attention to security, none of these documents explain why protecting personal information is necessary for data security and why the right to consent, deletion, and correction, is a must to enhance data security.

Personal Information (PI) is a fundamental concept in China's data privacy laws. In general, there are two types of definitions of PI: the three-element and the four-element definition.

The CSL and the Civil Code both set a three-element definition of PI, whereas the PIPL uses a four-element definition of PI. The CSL defines PI as:

> [V]arious information that, recorded electronically or through other means, that taken alone or together with other information, identify a natural person's identity, including but not limited to natural persons' full names, birth dates, national identification numbers, personal biometric information, addresses, telephone numbers, and so forth.[149]

The Civil Code defines PI in a similar way:

> [T]he information recorded electronically or in other ways that can be used, by itself or in combination with other information, to identify a natural person, including the name, date of birth, identification number, biometric information, residential address, telephone number, email address, health information, whereabouts, and the like, of the person.[150]

Definitions of PI under the CSL and the Civil Code share the same structure: (1) [various] information that (2) identifies (3) a natural person['s identity]. Neither the CSL nor the Civil Code explain the exact meaning of each element. According to the wordings, we can infer an underlying assumption of each element. The first element, "[various] information," delivers a clear message that, regardless of its format and content, any information, independently or collectively, can fall into the personal information protection. The second element, "identify," refers to a relationship between information and individuals. The third element, "a natural person ['s identity]," is in contrast to legal persona.

In two aspects, the Civil Code defines PI slightly differently from the CSL. First, the Civil Code changes "a natural person's identity" to "a natural person." Such a change expands the scope of PI from

[149] Cybersecurity Law of the People's Republic of China, art. 76, CAC (Nov. 7, 2016, 19:38) (author's translation).

[150] The Civil Code of the People's Republic of China (promulgated by the Standing Comm. Nat'l People's Cong., May 28, 2020, effective Jan. 1, 2021), art. 1034.

information about one's identity to information about a person. Second, the Civil Code adds "email address, health information, [and] whereabouts" as examples of PI in comparison with the CSL's list of PI. Such a change demonstrates that the legislature intended to expand the scope of PI.

The PIPL resembles the EU definition of PI: "'Personal information' refers to various information related to an identified or identifiable natural person recorded electronically or by other means, but does not include anonymized information."[151] The definition of PI under the PIPL contains four elements: (1) information (2) related to (3) an identified or identifiable (4) natural person. This structure is almost identical to that of PD under the GDPR.[152] Since the PIPL does not define each element nor provide guidance for practitioners to decide the scope of PI, we can only infer that China's legislators consider the EU's broad approach desirable and chose to define PI in a similar way.[153]

Despite China's consistent definition of PI, these statutes give no instructions on how to apply the concept of PI. The latest version of the Information Security Technology–Personal Information Security Specification (GB/T 35273-2020) ("2020 Specification") [154] was released in 2020 to clarify legitimate PI processing activities and help authorities manage, supervise, and evaluate PI processing activities.[155] The 2020 Specification is a recommended guidance, and it is non-mandatory; it may become a binding document if it is adopted as a regulatory standard by authorities.[156]

[151] Personal Information Protection Law of the People's Republic of China (promulgated by the Standing Comm. Nat'l People's Cong., Aug. 20, 2021, effective Nov. 1, 2021), art. 4.

[152] *See supra* Part III.A.

[153] Gil Zhang & Kate Yin, *A Look at China's Draft of Personal Information Protection Law*, IAPP (Oct 26, 2020), https://iapp.org/news/a/a-look-at-chinas-draft-of-personal-data-protection-law/ [https://perma.cc/9UVG-TWVV] (archived Oct. 17, 2023) ("Taking a closer look at the draft PIPL, it is easy to see many provisions in it are inspired by the EU General Data Protection Regulation.").

[154] Xinxi Anquan Jishu Geren Xinxi Anquan Guifan (信息安全技术 个人信息安全规范) [Information Security Technology – Personal Information (PI) Security Specification] (promulgated by the Standing Comm. Nat'l People's Cong., Mar. 6, 2020, effective Oct. 1, 2020) GB/T 35273-2020, NAT'L STANDARD P.R.C. (China) [hereinafter 2020 Specification].

[155] *See China Issues New Personal Information Security Specification*, WILMERHALE (March 24, 2020), https://www.wilmerhale.com/en/insights/client-alerts/20200324-china-issues-new-personal-information-security-specification [https://perma.cc/42GV-M3SC] (archived Sept. 12, 2023).

[156] In 2019, authorities in China adopted the old version of the Specification to regulate apps. *See* Minghe Hu, *China Issues Rules to Sop Apps from Abusing User's Personal Information in Latest Data Privacy Effort*, S. CHINA MORNING POST (Dec. 31, 2019), https://www.scmp.com/tech/apps-social/article/3044051/china-issues-rules-stop-apps-abusing-users-personal-information?module=perpetual_scroll_0&pgtype [https://perma.cc/5GPL-CPLZ] (archived Sept. 9, 2023).

The 2020 Specification defines PI as "any information that is recorded, electronically or otherwise, that can be used alone or in combination with other information to identify a natural person or reflect the activity of a natural person.[157] The 2020 Specification sets two criteria for determining whether a piece of information is PI—one is the criterion of identification and the other is the criterion of association. A piece of data is considered PI when either criterion is met. The criterion of identification is met if information identifies individuals. Based on the criterion of identification, "PI is the information that could help one identify a specific natural person through the specificity of the information."[158] This is identical to the CSL, Civil Code, and PIPL. The 2020 Specification leaves unexplained under what circumstances information is capable of identifying a specific person. As a result, the criterion of identification remains uncertain.

The criterion of association is met if the information is associated with individuals. Based on the criterion of association, "the information generated in the activities of a known natural person (such as the person's location information, call logs and browsing history) is PI."[159] This renders all human-generated data PI because they are derived from a person's activities. This criterion significantly expands the scope of PI to include information that is generally regarded as non-identifiable, such as longitude and latitude and even metadata (i.e., data that describe personal communications).[160]

The 2020 Specification adopts an uncertain and broad approach to identifiability, just like the GDPR and the CCPA. Its broadness, however, is unique in at least two aspects. First, while the GDPR and CCPA both highlight the importance of contextual analysis in applying identifiability, the 2020 Specification pays little attention to context. Second, the 2020 Specification considers some non-PI under the GDPR and CCPA to be PI (e.g., metadata, longitude and latitude). At the same time, the 2020 Specification skips inference data (i.e., profiling users' preferences and behaviors), which is regarded as identifiable under the GDPR and CCPA.

In China, judges were once inconsistent in applying identifiability. The case of *Zhu Ye v. Beijing Baidu Netcom Science & Tech. Co., Ltd.* epitomizes this inconsistency. A woman sued Baidu, the leading search engine in China, for privacy violations upon finding out that the pop-up advertisements from other websites were recommended on the basis of her search queries.[161] The district court and the appellate court delivered opposite rulings. The district court held that traces of online

[157] 2020 Specification art. 3.1.

[158] *Id.* annex A.

[159] *Id.*

[160] *Id.*

[161] Zhu Ye v. Beijing Baidu Netcom Sci. & Tech. Co., Ltd., CLI.C. 8917452.

activities fell into the scope of privacy because these data could reflect individual online preferences, interests, and needs. By contrast, the appellant court acknowledged the "privacy attributes" of online activities while considering them non-PI due to their anonymity.[162]

This murky situation started to clear up in July 2020 with *Huang v. Tencent Tech. Co., Ltd.*, and *Ling v. Beijing Microlive Vision Tech. Co., Ltd.* In *Huang*, the Beijing Internet Court ("the Court") first adopted two criteria of identifiability under the Specification[163] and held that WeChat friend lists and WeChat reading information were PI.[164] The Court claimed that the scope of PI was determined by two kinds of identifiability: identification of one's identity (i.e., who the person is) and identification of one's characteristics (what kind of person he or she is).[165] These two kinds of identifiability were in line with the two criteria under the 2020 Specification.[166] With reference to the criterion of identification, WeChat friend lists and WeChat reading information were reasonably likely to identify a person in combination with other information. By applying the criterion of association, WeChat friend lists and WeChat reading information reflected characteristics of a person. These data were undoubtedly PI, as both criteria were met.

In the case of *Ling v. Beijing Microlive Vision Tech. Co., Ltd.*, the Beijing Internet Court held that contact information was PI because it met the criterion of association. That is, a natural person's contact information reflected one's social life and relationship.[167] In addition, the Court stated that applying identification was not merely asking whether data in isolation could identify a person. Instead, the Court shall consider whether a combination of data met the criterion. Hence, geolocation, regardless of its accuracy, was PI when it was combined with one's phone number.[168]

The case of *Ling* also delivered that judgments of PI were contextualized. The plaintiff argued that names, telephone numbers, contacts, and geolocations were PI because they were listed as PI both under the CSL and the Specification.[169] The Court, at the same time, refused to decide that those data were PI under all circumstances and claimed that the Court's determination on PI was context-

[162] *Id.*

[163] The case cited an early version of the Specification (2017 Specification) because the 2020 Specification was not released when the judgment was made. The 2017 Specification defines identifiability the same manner as the 2020 Specification. *See* Huang v. Tencent Tech. Co., Ltd., Beijing 0491 Min Chu 16142, 27 (Beijing Internet Civ. Ct. 2019).

[164] *Id.* at 28.

[165] *Id.*

[166] *Id.*

[167] *See* Ling v. Beijing Microlive Vision Tech. Co., Ltd., Beijing 0491 Min Chu 6694, 3 (Beijing Internet Civ. Ct. 2019).

[168] *Id.* at 48.

[169] *Id.* at 4.

dependent.[170] As discussed before, the CSL, Civil Code, and 2020 Specification do not explicitly mention the importance of contextual analysis in deciding PI. But this case demonstrates that contextual analysis is playing a role in applying identifiability.

China's approach to identifiability may be evaluated from two perspectives. From the perspective of legislators, the definition of PI is consistent, but PI's elements lack further clarification. Hence, applying identifiability remains vague and uncertain. From the perspective of regulators and judges, the application of identifiability boils down to two specific criteria: identification and association. The former is keeping with the conventional understanding of identifiability, while the latter is an innovative standard that is not found in the EU and U.S. data privacy laws. Applying identification and association is still uncertain and broad.

V. SIMILARITY IN DIFFERENCE: THE IDENTIFIABILITY PROBLEM

Part III surveyed legal discourses over identifiability in the EU, United States, and China. This Part will first summarize the findings of Part III to show the homogeneity of the ways identifiability is inadequately defined and inconsistently applied. Then this Part will generalize such transnational homogeneity as "the identifiability problem." Recognizing the identifiability problem is significant because it reveals an overlooked common phenomenon across jurisdictions. On the other side of the coin, it lays a factual foundation for synthesizing transnational efforts into solving the identifiability problem from a doctrinal perspective.

A. *Similarity*

Four convergences emerge from the examination of legal discourses over identifiability:

(1) Identifiability functions as one essential element of legally protected data and thereby functions as a critical gatekeeper to the transnational privacy regulation.
(2) Identifiability is legally ill-defined.
(3) Identifiability is assumed to be closely related to the identification of a person.
(4) The application of identifiability lacks certainty and leads to the problem of judicial unpredictability.

First, all three jurisdictions adopt identifiability as one essential element of legally protected data. Be it qualifying data subject or the relationship between data and data subject, identifiability is an

[170] *Id.* at 52.

integral part of Personal Data in the EU, PII/Personal Information in the United States, and Personal Information in China. Since legally protected data in each jurisdiction is a fundamental concept in data privacy law, identifiability serves to be a vital gatekeeper for the scope of data privacy regulation.

Second, none of the three jurisdictions clearly explains the meaning of identifiability. EU legislators shed certain light on the meaning of "identifiable" but leave the term "identified" unclear.[171] U.S. and Chinese lawmakers seem to treat the nominal meaning of identifiability as adequate and, therefore, do not define it.[172] From a doctrinal perspective, the concept of identifiability is rarely articulated at the legislative level.

Third, all three jurisdictions rely on a presumed relationship between identifiability and the identification of a person.[173] WP 29 simply uses the notion of identification to articulate the basic meaning of identifiability under the EU data protection law.[174] A U.S. court understood identifiability interchangeably with identification.[175] In China, the 2020 Specification treats the criterion of identification as one approach to applying identifiability. [176] Highlighting the identification of a person implies that data privacy is legally understood as a value pertinent to individuals.[177] Even though data privacy has societal benefits, they are overshadowed by the values of individuals.[178]

[171] *See infra* discussions on Part IV Section A.

[172] *See infra* discussions on Part IV Section B and C.

[173] *See* BYGRAVE, *supra* note 8, at 129–30 ("From these definitions, we can discern two cumulative conditions for data to be 'personal': first, the data must relate to or concern a person; secondly, the data must enable the identification of such a person....the first condition can be embraced by the second, in the sense that data will normally relate to, or concern, a person if it enables that person's identification. In other words, the basic criterion appearing in these definitions is that of identifiability—that is, the potential of data to enable identification of a person.").

[174] *See supra* Part IV.A.

[175] *See In re* Hulu Privacy Litig., No. C 11-03764 LB, 2014 WL 1724344, at *11 (N.D. Cal. Apr. 28, 2014).

[176] *See supra* Part IV.C.

[177] Even though the CCPA allows an identification of household as an extra gatekeeper to initiate legal regulation, such requirement is rare globally speaking. *See* BYGRAVE, *supra* note 8, at 135.

[178] Protection of individual autonomy and human dignity are often cited in privacy laws and legal scholarships to justify emerging legislations of data privacy. For example, The European Court of Human Rights stresses that developing one's identity autonomously is an important aspect to fulfill the fundamental right to respect for private life. *See, e.g.*, EUR. CT. OF H.R., GUIDE ON ARTICLE 8 OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS: RIGHT TO RESPECT FOR PRIVATE AND FAMILY LIFE, HOME AND CORRESPONDENCE 66–68 (updated on Aug. 31, 2022), https://www.echr.coe.int/documents/guide_art_8_eng.pdf [https://perma.cc/Q6KT-JRD5] (archived Sept. 9, 2023). In the US, the emergence of several states informational privacy statues emphasizes autonomous personhood argument. One common effect of those state statutes is strengthening consumers' controlling power over their personal data. For

Fourth, all three jurisdictions apply identifiability in unpredictable and inconsistent ways. As mentioned before, most approaches lack sufficient qualifications on what identifiability entails. The vague definition of identifiability directly leads to uncertainty in its application. In addition, some approaches apply identifiability in a case-by-case manner, as they acknowledge that the identifying capacity of data is highly contextual and unstable.[179]

Two contextual factors can substantially influence the legal status of data: data recipients and technological advances. Data recipients lead to the audience challenge, while technological advances bring about the re-identification challenge and aggregation challenge.

The audience challenge denotes that different data recipients comprehend information differently based on their distinct prior knowledge. The same piece of data "can be anonymous for one controller, while identifiable for another." [180] Therefore, the legal status of data varies from audience to audience.

The re-identification challenge derives from re-identification techniques. Re-identification techniques make use of anonymized data to establish a unique link to individuals. These techniques can be easily employed by experts and even learned amateurs to render identifiability a moving target.[181] With these techniques, "[t]he same

example, the Colorado Privacy Act explicitly seeks to create personal data privacy rights, including the consumer's right to opt out of the processing of their personal data; access, correct, or delete the data; or obtain a portable copy of the data. *See* Colorado Privacy Act of 2021 § 6-1-1302. In China, the PIPL connects the value of dignity and sensitive data protection. Personal Information Protection Law of the People's Republic of China (promulgated by the Standing Comm. Nat'l People's Cong., Aug. 20, 2021, effective Nov. 1, 2021), art. 28 (Sensitive personal information is personal information that, once disclosed or used in an illegal way, can cause discrimination against individuals or lead to serious harms on persons or prosperities, including race or ethnic origin, religious beliefs, personal biological features, medical health, financial accounts, personal behaviors, etc.). *See also How does the Personal Information Protection Law protect the security of personal information*, NAT'L PEOPLE'S CONG. P.R.C. (Sept. 14, 2021), http://www.npc.gov.cn/npc/c30834/202109/09f2056a57fd4ff0a3cc9ae23c1cbb27.shtml [perma.cc/EP8G-H2L5] (date archived Sept. 9, 2023).

[179] The EU approach is context-dependent. In the U.S. legal context, only the broad approach highlights that applying identifiability requires contextual interpretations. In China, the Beijing Internet Court emphasizes that interpreting identifiability is a case-by-case analysis.

[180] Urgessa, *supra* note 74, at 528.

[181] For example, after America Online (AOL) released anonymized 20 million search queries for 650,000 users of AOL's search engine for research purpose, two reporters from *New York Times* immediately identified User 4417729's identity: a 62-year-old widow, Thelma Arnold, from Georgia. *See, e.g.*, Michael Barbaro & Tom Zeller Jr., *A Face is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES (Aug. 9, 2006), https://www.nytimes.com/2006/08/09/technology/09aol.html. A computer scientist found that 87.1% of people in the US could be uniquely identified merely based on ZIP code, birth date, and sex. *See* Latanya Sweeney, *Simple Demographics Often Identify People Uniquely*, (Carnegie Mellon U., Data Privacy Working Paper 3, Pittsburgh 2000) https://dataprivacylab.org/projects/identifiability/paper1.pdf [https://perma.cc/4GXB-

piece of data can be anonymous at the time of collection, but turn into personal later, just sitting there, simply by virtue of technological progress."[182] In light of re-identification, the identifiability of a piece of data is dynamic and the notion of identifiability becomes futile.

The aggregation challenge derives from aggregation techniques. By combining seemingly unrelated pieces of information, companies learn new facts about a person in an extensive and penetrating way.[183] The employment of aggregation techniques has no correlation to the degree of identification.[184] Companies may have aggregated information about one person, but this aggregated information is rarely connected to a specific person.[185] Low identifiable and aggregated information causes privacy concerns but may escape privacy regulation.

Two kinds of legislative efforts attempted to overcome the uncertainty of identifiability, but both efforts have generated unsatisfactory results.

One kind of effort applies identifiability strictly and maintains a narrow scope of data privacy laws. The specific-types approach in the United States is of this sort. With the specific-types approach, data privacy laws explicitly list certain types of data for regulation. Unlisted data are free for commercial use. The downside of this kind of approach is that the application of identifiability may become too narrow. When this is the case, the problem of under-regulation arises. The law does not intervene in cases of data collection and processing when it should.

Z6JF] (archived Sept. 9, 2023). Technicians easily discovered users' identities and all movie-watching preferences once connecting anonymous Netflix rating data and Internet Movie Database (IMDb) rating data. *See* Dan Jackson, *The Netflix Prize: How A $1 Million Contest Changed Binge-Watching Forever*, THRILLIST (July 7, 2017), https://www.thrillist.com/entertainment/nation/the-netflix-prize [https://perma.cc/VZL4-LWE6] (archived Sept. 9, 2023). The processing of credit cards records (i.e., merely accessing amounts spent, shop type and a code representing each person) can uniquely re-identify 90% individuals who provide 3-month credit card transactions. *See* Yves-Alexandre de Montjoye, Laura Radaellim Vivek Humar Singh & Alex Pentland, *Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata*, 347 SCIENCE 536, 541 (2015), https://science.sciencemag.org /content/347/6221/536 [https://perma.cc/G83P-3RW7] (archived Sept. 21, 2023). The Australian Office of Information Commission even introduced the Privacy Amendment (Re-identification Offence) Bill in 2016 after researchers from the University of Melbourne re-identify every individual health practitioners in the Medical Benefit Scheme datasets by just combining the de-identified datasets with publicly available data. *See Re-Introducing the Re-Identification Offence Bill: The Dumbest Privacy Idea This Year?*, PRIVACY108 (Dec. 29, 2021), https://privacy108.com.au/insights/re-identification-offence-bill-the-dumbest-privacy-idea-this-year/ [https://perma.cc/XG94-EBPU] (archived Oct. 17, 2023).

[182] Purtova, *supra* note 43, at 44.

[183] Algorithms are used to aggregate massive information to learn new facts about users. *See supra* Part I.C.

[184] *See* DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 123 (2008).

[185] Solove names such a phenomenon as "high aggregation and low identification." *See id.*

The other kind of effort swings the pendulum to the other end to apply identifiability broadly and allow an extensive scope of data privacy laws. The EU approach is of this sort. Purtova predicted that the GDPR would be "the law of everything" in the near future. [186] When the application of identifiability is too broad, the problem of over-regulation arises. Arguably, all data is identifiable and deserves regulation. The encompassing scope of identifiable data leads to an extensive regime of regulation, which could be impractical and unfeasible.[187] Besides, if any data can arguably be identifiable and deserves legal protections, identifiability loses its function as a gatekeeper of regulation. After all, "[a] legal concept will do us little good if it expands like a gas to fill up the available space."[188]

B. *Difference*

The EU, the United States, and China are vastly distinct in three senses. Broadly speaking, these three entities are different from each other in terms of legal systems, data privacy philosophy, and data privacy regulation.

(1) Regarding legal systems, most EU Member States adopt civil law systems, while the U.S. belongs to the common law system. China is a mixed legal system, which contains Germanic civil law and socialist law. Different legal systems usually are operated via different legal mechanisms.
(2) Regarding data privacy philosophy, the EU safeguards human dignity from shame and humiliation in its data protection laws.[189] The United States protects privacy on a different ground: freedom. American scholarly writings and court doctrines constantly treat the state as the main enemy of invading individual privacy and liberty.[190] Influenced by the global movement of data privacy but going in a different direction, China's data privacy laws value security.[191]
(3) Regarding data privacy framework of regulation, the EU adopts the rights-based approach. The United States manifests a coalescence of two approaches. On one hand, the United States

[186] Sooner or later our life will be fully digitalized, and any data can influence personal life. Purtova, *supra* note 43, at 72–75.

[187] Purtova is concerned that "a highly intensive and non-scalable regime of rights and obligations that results from the GDPR cannot simply be upheld in a meaningful way." *Id.* at 42.

[188] Tom Gerety, *Redefining Privacy,* 12 HARV. CIV. RTS.–CIV. LIBERTIES L. REV. 233, 234 (1977).

[189] James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151, 1164 (2004).

[190] PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY 225 (1995).

[191] *See* Zhao & Feng, *supra* note 144, at 4.

historically prefers a market-based approach. On the other hand, a rights-based approach is employed in some states' data privacy laws. China's data privacy framework is still under development. A full-blown overarching regulation is yet to take shape.

In a narrow sense, there are two major differences among how the EU, United States, and China define and apply identifiability. First, identifiability is used dissimilarly. The EU uses identifiability as a qualifier on data subjects. Rather than all individuals, only identified or identifiable individuals are protected by the law. On the contrary, the United States uses identifiability as one kind of relationship. In general, information that identifies individuals is legally protected. In China, both usages of identifiability exist in various laws. Second, identifiability is interpreted differently. The EU uses the reasonable likelihood test, whereas the United States adopts a variety of approaches. China coins a technique—the standard of association—to supplement the criterion of identification.

C. *Similarity in Difference*

The EU, United States, and China coincide in treating identifiability as the critical gatekeeper to data privacy; they also encounter shared struggles against the uncertainty over identifiability. This transnational convergence is what I call the "the identifiability problem": the set of common difficulties in implementing data privacy regulation caused by the uncertainty of whether data is identifiable. To accurately understand the identifiability problem, there are three clarifications.

First, the generalization is mainly based upon three jurisdictions and, therefore, does not intend to embrace all jurisdictions. However, the generalization demonstrates that the emergence of the identifiability problem transcends legal systems, legal cultures, regulatory frameworks, and interpretations of identifiability because the EU, United States, and China are vastly distinct in these aspects.

Second, some evidence suggests that a wide range of jurisdictions, not limited to the EU, United States, and China, also fit into the generalization of the identifiability problem.[192] While this Article

[192] Bygrave summarizes six common issues of applying identifiability at an international level: "(1) What exactly is meant by identification/identifiability? (2) How easily or practicably must a person be identified from data in order for it to be regarded as "personal"? (3) Who is the legally relevant agent of identification (that is, the person is to carry out identification)? (4) To what extent must the link between a set of data and a person be objectively valid? (5) To what extent is the use of auxiliary data or information permitted in the identification process? Can data be 'personal' if it allows a person to be identified only in combination with other (auxiliary) data or information? (6) to what extent must data be linkable to just *one* person in order to be 'personal'?" *See* BYGRAVE, *supra* note 8, at 130.

surveys data privacy laws in three critical jurisdictions, any jurisdiction that adopts the criterion of identifiability as the critical condition to trigger privacy regulation must counter the identifiability problem.

Third, the above-mentioned similarities are relative, rather than absolute. In comparative law, similarity always "entails the possibility of difference."[193] As Part II asserts, this Article emphasizes "similarity in difference." Based upon the transnational similarity, the multinational, if not global, role of identifiability is displayed. By comparing legal discourses over identifiability in the EU, United States, and China, this Part demonstrates the transnational existence of the identifiability problem. Depicting such an overlooked phenomenon across jurisdictions establishes a factual foundation for integrating multiple doctrinal attempts to address the threshold issue of data privacy regulation.

VI. CURRENT SOLUTIONS AND A NEW DIRECTION

Rather than recognizing the identifiability problem in transnational privacy regulation, current identifiability studies focus on determining whether identifiability is useful. Current scholarship takes one of two extreme approaches to identifiability: revision or abandonment. Scholars who choose the former propose various ways that seek to re-interpret identifiability; other scholars who pursue the latter propose alternative thresholds to replace identifiability. This Part will briefly describe representative approaches that call for revising and abandoning identifiability, critique them, and articulate a new direction.

A. *The Revision of Identifiability*

In the three jurisdictions, identifiability is the indispensable condition to trigger privacy regulation.[194] Addressing challenges on applying identifiability is necessary. PII 2.0 and group privacy theory are two representative solutions. PII 2.0 creates three categories of identifiability, which makes the application of identifiability more functional. Group privacy theorists argue that the bar of identifiability needs to be lowered. Privacy law should intervene once a group of people, not merely a natural person, is identified. By evaluating the PII 2.0 and group privacy approaches, this Section demonstrates that neither strategy helps the jurisdictions completely out of the identifiability predicament.

[193] Nils Jansen, *Comparative Law and Comparative Knowlege*, *in* THE OXFORD HANDBOOK OF COMPARATIVE LAW 291, 296 (Mathias Reimann & Reinhard Zimmermann eds., 2d ed. 2019).

[194] *See supra* Part V.A.

1. PII 2.0

Paul Schwartz and Daniel Solove recognize that identifiability is a concept of degree. As the application of identifiability is not a binary choice, they reject the binary understanding of PII/non-PII. Between PII and non-PII is an intermediate category that covers data that may potentially identify individuals and may deserve a certain degree of regulation.

Based on a continuum notion of the risk of re-identification, Schwartz and Solove interpret PII to have three categories: (1) identified, (2) identifiable, or (3) non-identifiable information.[195] Identified information refers to data that can single out a specific person, such as names and SSNs.[196] Non-identifiable information is data that is the least likely to identify a natural person, such as data on the population in the United States.[197] In between the identified and non-identifiable information is a new category, identifiable information. Identifiable information denotes data that can possibly, but not "significantly probably," become identified.[198] Aggregated information falls into this category, as it directly and uniquely links to a group of people, rather than a specific person. It may become personally identified when it is combined with other information.

Schwartz and Solove argue that different categories of data should bear different levels of obligations on data holders.[199] They illustrate their approach with the OECD's version of Fair Information Practices (FIPs).[200] The three categories of PII correspond to different subsets of FIPs obligations on data holders.[201] When a person is identified, full FIPs should apply. When a person is identifiable, three FIPs (out of seven)—data quality, security, and openness principles—should apply, but other principles—collection limitation, purpose specification, use limitation, and individual participation—do not apply. When a person is non-identifiable, no FIPs should apply.

PII 2.0 breaks down the either-or application of identifiability. Rather than being forced to choose between PII and non-PII, judges and regulators now have a fallback option–the identifiable one. At the

[195] Schwartz & Solove, *supra* note 42, at 1877.

[196] *Id.*

[197] *Id.* at 1878.

[198] *Id.*

[199] *Id.* at 1883.

[200] ROBERT GELLMAN, FAIR INFORMATION PRACTICES: A BASIC HISTORY 11 (Version 2.22, Apr. 6, 2022), https://bobgellman.com/rg-docs/rg-FIPshistory.pdf [https://perma.cc/6TNQ-7K64] (archived Aug. 23, 2023). Schwartz and Solove use OECD's version as an illustration: (1) collection limitation principle; (2) data quality principle; (3) purpose specification principle; (4) use limitation principle; (5) security safeguard principle; (6) openness principle; (7) individual participation principle; (8) accountability principle. *See* OECD Privacy Principles, OECD, http://oecdprivacy.org/ [https://perma.cc/YUJ9-XT3V] (archived Aug. 23, 2023).

[201] *See* Schwartz & Solove, *supra* note 42, at 1877.

same time, by distributing differential moral obligations on identified, identifiable, and non-identifiable categories, PII 2.0 is less burdensome and more feasible to comply with than the GDPR.

Unfortunately, PII 2.0 faces two conceptual challenges. First, if a tertiary taxonomy of PII, i.e., PII 2.0, better captures the continuum of the risk of re-identification than a binary dichotomy, it begs a series of questions: Is a quaternary or a quandary taxonomy of PII even more desirable than PII 2.0? If so, why not go for four, five, or more categories? Why do we stop at three?

Second, there is a conceptual flaw underlying PII 2.0. Most people presuppose that anonymization preserves privacy. Should this be the case, when privacy is violated, data involved must be personally identifiable. But this does not mean that identifiable data must cause privacy violations. Put differently, the former denotes that identification of a person is a necessary condition of privacy violations, while the latter indicates that identification is a sufficient condition. Current legal practices of treating identifiability as the sole gatekeeper to data privacy regulation essentially support the latter view, which lacks justification and is therefore questionable. Since PII 2.0 is constructed on the risk of re-identification, it is important to check its theoretical foundation from the outset: Why does the risk of re-identification outweigh other factors that could contribute to privacy violations?

2. Group Privacy Approach

Some EU scholars find the aggregation challenge particularly disturbing. For them, privacy violations do not occur only when PII are accessed.[202] Privacy concerns also arise when "data is no longer gathered about one specific individual or a small group of people, but rather about large and undefined groups."[203] Tech companies possess many new facts about a natural person without singling out any person in their algorithmic profiles.[204] These new facts are not expected "when the original, isolated data [is] collected."[205] Solove sounds the privacy alarm about "high aggregation low identification."[206] To address the

[202] *See* Linnet Taylor, *Safety in Numbers? Group Privacy and Big Data Analytics in the Developing World*, *in* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES 13, 16 (Linnet Taylor, Luciano Floridi & Bart van der Sloot eds., 2017).

[203] *See* Linnet Taylor, Luciano Floridi & Bart van der Sloot, *Introduction: A New Perspective on Privacy*, *in* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES 1, 5 (Linnet Taylor, Luciano Floridi & Bart van der Sloot, eds., 2017) [hereinafter Taylor, Floridi & van der Sloot].

[204] *See* Lanah Kammourieh, Thomas Baar, Jos Berens, Emmanuel Letouzé, Julia Manske, John Palmer, David Sangokoya & Patrick Vinck, *Group Privacy in the Age of Big Data*, *in* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES 37, 48 (Linnet Taylor, Luciano Floridi & Bart van der Sloot eds., 2017).

[205] *See* SOLOVE, *supra* note 184, at 118.

[206] *Id.* at 123.

aggregation challenge, these scholars argue that the application of identifiability should not only cover data that can exclusively identify individuals.[207] Identifying a group of people should be a new interest of law. Accordingly, the right to privacy should be considered as a group right.

Though unspoken, the group privacy approach in fact departs from the majority legal practice of treating the identification of a person as the singular gatekeeper to data privacy regulation. The group privacy approach considers the risk of re-identification as an important, but not the sole, factor that constitutes privacy violations. Privacy violations can occur even when data is at a low degree of identification. This fresh understanding of what constitutes data privacy violations is an important contribution of the group privacy approach.

Unfortunately, group privacy theories are underdeveloped for two reasons. First, there is no unified terminology among group privacy scholars. Scholars are still trying to propose clear ways to articulate what a group is and what a group privacy right is.[208] Second, members of the groups normally do not know of the existence of the group, let alone their membership and the criterion of grouping.[209] This leads to the following difficulties in advancing group privacy regulation: What is the appropriate basis for defining groups? What should be the legal status of a group (e.g., treated as a legal person like corporations or on the basis of a cluster of individual rights)? Who has the right to claim group privacy (e.g., the individuals belonging to the group or representatives of the groups)?

One problem with group privacy theories is that the proponents assume all grave consequences caused by the aggregation technique are privacy harms. Group privacy scholars fail to explain why all harms caused by the aggregation challenge concern privacy. Some scholars equate privacy harms to any ethical challenge of misusing data.[210] Indeed, the use of aggregation techniques threatens privacy, but it is debatable whether privacy is the only value that is at stake. Security, justice, and fair competition are other values that can be jeopardized. Some, therefore, claim that group privacy theory should

[207] *See generally* Taylor, Floridi & van der Sloot, *supra* note 203.

[208] *See* Linnet Taylor, Luciano Floridi & Bart van der Sloot, *Conclusion: What Do We Know About Group Privacy?*, *in* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES 279, 280 (Linnet Taylor, Luciano Floridi & Bart van der Sloot eds., 2017).

[209] *See id.* at 281.

[210] *See* Taylor, *supra* note 202, at 31–2 ("The potential risks and harms outlined in this analysis all relate to the consequences of drawing conclusions about a given group based on assumptions drawn from other groups. As such, they are problems with treating the group as a category....most [generalizations] raise issues of both privacy and data protection because they incorporate problems both of visibility and identification, and protection from intervention. These problems point to the need for a new ethical approach to research with regard to group-level information.").

solve injustice and discrimination problems, not only privacy harms.[211] Group privacy theorists, without differentiating privacy from other values, face further challenges: Why are all ethical issues of using data essentially privacy concerns? If we adopt group privacy theory, will all grave consequences mentioned be avoided?

B. *The Abandonment of Identifiability*

Some scholars consider identifiability inadequate as a threshold for regulating data privacy and call for abandoning it altogether.[212] Abandonment has been argued on two grounds, one being technical, and the other being conceptual. Technologically speaking, the notion of identifiability may be futile since all data are potentially identifiable. Hence, it is increasingly difficult to categorize data as identifiable or non-identifiable. One well-known proponent of this view is Paul Ohm, who rejects identifiability due to the re-identification challenge and proposes a utility approach as a replacement for identifiability.[213] Conceptually speaking, some contend that there is no conceptual relation between identifiability and privacy regulation, so it is a flaw to take the former as the sole condition to the latter. One well-known proponent of this view is Helen Nissenbaum, who believes that whether data is identifiable or not has no bearing on privacy.[214] Instead of identifiability, she proposes a "contextual integrity" approach that focuses on contextual factors of defining privacy.[215] By evaluating the utility approach and contextual integrity approach, this Section will demonstrate that both approaches generate extra difficulties with delimiting the scope of data privacy regulation.

1. Utility Approach

Paul Ohm passionately argues for the abandonment of identifiability. He finds the re-identification challenge unavoidable and impossible to overcome. With the advances in re-identification techniques, anonymized data can easily be transferred into identifiable data.[216] The easy transformation from PII to non-PII renders PII-centric privacy regulation futile. Ohm points out that our legal

[211] *See* Kammourieh, Baar, Berens, Letouzé, Manske, Palmer, Sangokoya & Vinck, *supra* note 204, at 48.

[212] *See generally* Solon Barocas & Helen Nissenbaum, *Big Data's End Run Around Anonymity and Consent*, *in* PRIVACY, BIG DATA AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT (Julia Lane, Victoria Stodden, Stefan Bender & Helen Nissenbaum eds., 2015).

[213] *See generally* Paul Ohm, *Broken Promises of Privacy: Responding to the Suppressing Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010).

[214] *See generally* Barocas & Nissenbaum, *supra* note 212 (arguing that anonymity is the means of protecting privacy, not privacy itself).

[215] *See id.* at 46.

[216] *See* Ohm, *supra* note 213, at 1744.

communities underestimate this technical reality and, consequently, improperly put faith in identifiability.[217] His view on the failure of anonymization has gained wide acceptance among privacy scholars.[218]

Ohm urges regulators to adopt a utility approach to determine the material scope of privacy regulation: in every case the regulator should assess the benefits and the costs of free flow of information.[219] Due to re-identification, Ohm forfeits the legal binary understanding of identifiability as the critical gatekeeper to data privacy regulation.[220] Instead, he embraces the technological understanding of identifiability (i.e., the risk of re-identification) as one cost of the free flow of personal data. A conclusive decision on regulation should be made after balancing three factors: (1) the risk of re-identification; (2) the sensitivity of data; [221] and (3) the benefits of the free flow of information. [222] Among these factors, the risk of re-identification should be calculated based on five factors—data-handling techniques,[223] private or public release,[224] quantity,[225] motive,[226] and trust.[227]

Ohm contributes to bursting the bubble of anonymization. Faith in anonymization is flawed in light of re-identification techniques. From a technological perspective, identification is always possible, as it gets increasingly easy to link from data to a natural person. For Ohm, it is useless to adopt identifiability as a watershed of regulatory and non-regulatory regimes. But he does not deny the importance of identifiability in judgments of privacy violations. In his approach, identifiability (i.e., the risk of re-identification) is one factor contributing to privacy violations.[228]

Unfortunately, Ohm's utility approach is unfeasible for two reasons. First, deciding the basic unit of the calculation is next to impossible. On the one hand, if every piece of data must be evaluated, it would be impractical, if not impossible, to evaluate the risk of re-identification and the benefit of the free flow of information. The

[217] *See id.* at 1759.

[218] *See* HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY AND THE INTEGRITY OF SOCIAL LIFE 27 (2009); Schwartz & Solove, *supra* note 42; Purtova, *supra* note 43.

[219] *See* Ohm, *supra* note 213, at 1759.

[220] *Id.* at 1705.

[221] *Id.* at 1768.

[222] These include better medical researches and treatments, better search tools and products, and fewer identity thefts. *Id.*

[223] One such technique is a rough evaluation without a mathematical precession. *Id.* at 1765.

[224] The private release has a lower risk than the public release. *Id.*

[225] Such questions include how much data to release and how long to retain matters. *Id.* at 1766.

[226] *E.g.*, *id.* at 1767 (whether they have the motive to re-identify).

[227] For example, regulators should craft mechanisms to build trust between data subjects and data administrators. *E.g.*, *id.* at 1767–68.

[228] *See id.* at 1765.

workload will be too demanding, and the risk and the benefit can be too subtle to compare. On the other hand, if a combination of data should be evaluated, it opens another can of worms. How many pieces of data constitute the right combination of data? Two, three, or even more? Is there a magic number, after all? Second, the utility approach is more complicated than simply applying a mathematical formula. It requires the exercise of normative judgment for which Ohm offers no guidance. Each of the three factors is both quantitative and qualitative. For example, the sensitivity of data is a combination of a quantitative judgment (e.g., mathematical predictions of financial harms of unwanted disclosures) and a qualitative determination (e.g., unwanted disclosures of medical data are bad). Assigning values to the costs and the benefits in each case significantly increases the complexity of applying the utility approach.

2. Contextual Integrity Approach

Like Ohm, Helen Nissenbaum argues for the abandonment of identifiability, but on a different ground. Nissenbaum notes that the disclosure of PII in the public place is not necessarily a violation of privacy. [229] She thereafter argues that identifiability is not conceptually relevant to the definition of privacy. Privacy is intact as long as the flow of personal information is "reasonable." What falls in the realm of reasonable flow is determined by contextual informational norms.[230] In other words, privacy "is preserved when informational norms are respected and violated when informational norms are breached."[231]

To distill contextual informational norms, we should consider three essential components in the flow of data—actors (senders, recipients, and subjects), attributes (types of data), and transmission principles (how to transfer data). [232] Nissenbaum formulates contextual informational norms: "In a context, the flow of information of a certain *type* (attributes) about a *subject* (acting in a particular capacity/role) from a *sender* (possibly the subject, acting in a particular capacity/role) to a *recipient* (acting in a particular capacity/role) is governed by a particular *transmission principle*."[233] The purpose of privacy regulation is to preserve the integrity of contextual norms once it is identified.

[229] NISSENBAUM, *supra* note 218, at 113.

[230] *See id.* at 165.

[231] *Id.* at 140.

[232] *See id.* at 129.

[233] Deirdre K. Mulligan & Helen Nissenbaum, *Background for Concepts of Privacy Exercise*, CCC PRIVACY BY DESIGN 11 (Feb. 5 & 6, 2015), http://archive2.cra.org/ccc/files/docs/meetings/Privacy/Background%20for%20Concepts%20of%20Privacy%20Exercise.pdf [https://perma.cc/JS4K-7NZA] (archived Aug. 23, 2023) (emphasis added).

Nissenbaum's theory appears to be promising. The formula she offers gives the impression that norms may be identified mathematically. It is very appealing, especially to computer scientists and technicians, who can realize the model by building computational models with variables and parameters.[234] Once norms are identified, determining whether they are complied with will be straightforward, if not fall readily into place. In addition, among all approaches this Article examined, contextual integrity is the only one that expressly touches on the conceptual relationship between identifiability and privacy. For Nissenbaum, identifiability is conceptually irrelevant to privacy and thereby we should give up resolving problems surrounding this wrong standard and turn to other elements that conceptually constitute privacy violations.

However, the application of contextual integrity is limited. Not every context readily has well-established informational norms. It may be difficult to find a norm, let alone an agreed-upon one, for a newly emerging context (e.g., artificial intelligence or digital devices "jailbreak") or a context that is under huge transition (e.g., Brexit or facial recognition techniques used in an educational context).[235] Nissenbaum concedes that, even in conventional contexts, some informational norms are not articulated well enough for legal regulation or public policy, such as informational norms for friendship, courtship, kinship, marriage, and religion.[236] In fact, entrenched informational norms may only be found in very limited contexts.

Contextual integrity also faces a normative challenge. In essence, informational norms are shaped by value systems, rather than contexts. Clarifying details of contexts contributes to making normative decisions on privacy protection in specific situations, but contexts alone, however clearly they are articulated, cannot resolve disagreements on norms. At its best, the contextual integrity approach provides descriptive clarification on contexts, rather than normative guidance.

[234] *See* Jane Henriksen-Bulmer, Shamal Faily & Sheridan Jeary, *Privacy Risk Assessment in Context: A Meta-Model Based on Contextual Integrity*, 82 COMPUT. & SEC.270, 271 (2019); Michael Zimmer, *Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity*, SOC. MEDIA + SOC'Y (Apr.–June 2018), https://doi.org/10.1177/2056305118768300 [https://perma.cc/JLJ2-65BP] (archived Sept. 21, 2023); Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman & Nick Feamster, *Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity*, 2 PROC. ACM ON INTERACTIVE, MOBILE, WEARABLE & UBIQUITOUS TECHS. 1, 14 (2018); Sebastian Benthall, Seda Gürses & Helen Nissenbaum, *Contextual Integrity through the Lens of Computer Science*, 2 FOUNDS. & TRENDS IN PRIV. & SEC. 1, 2 (2017), https://sbenthall.net/papers/3300000016-Benthall-Vol2-SEC-0016.pdf [https://perma.cc/8UBY-BM7W] (archived Aug. 23, 2023).

[235] *See* Neil Connor, *Chinese school uses facial recognition to monitor student attention in class*, TELEGRAPH (May 17, 2018, 11:23 AM), https://www.telegraph.co.uk/news/2018/05/17/chinese-school-uses-facial-recognition-monitor-student-attention/ [https://perma.cc/5UWS-2XAX] (archived Aug. 23, 2023).

[236] NISSENBAUM, *supra* note 218, at 236.

Even in a particular context where all necessary components can be precisely identified, specifying an informational norm may still prove to be elusive. As an example, take students' attendance records at a state university. It is debatable whether selling this data constitutes a departure from an informational norm in an educational context. Consider a hypothetical situation: some people (30 percent) believe that a state university should not transfer any academic records to advertisers for profit. Other people (30 percent) feel that transactions related to attendance records are morally permissible. Still, others (40 percent) are neutral on this issue. Specifying an informational norm is difficult under such a situation, even though we clearly know all necessary components—data subjects (i.e., students), data senders (i.e., a state university), data recipients (i.e., advertisers), types of data (i.e., records of attendance), and transmission principle (i.e., selling for profits) in a specific context (i.e., education). In this kind of situation, Nissenbaum's version of the contextual integrity approach does not provide a way towards a consensus on a default norm.

C. *A Normative Inquiry*

Scholars have strenuously sought solutions that deal with the threshold issue in data privacy law.[237] Previous attempts point in opposite directions—one has to decide to either revise identifiability or abandon it. The depiction of the identifiability problem in transnational privacy regulation calls forth an urgent inquiry. Rather than hastily choosing between revision and abandonment, we need to resolve a normative question: Can identifiability be justified? In other words, what is the normative foundation of using identifiability as a gatekeeper to data privacy law? Addressing such a normative problem substantially informs our decision on whether to revise or abandon identifiability.

The identifiability problem reveals a baffling phenomenon that stems from unreflective assumptions. Identifiability is notoriously difficult to apply from a legal perspective. New tracking technologies and innovative business models keep reminding us of the failure of identifiability. The industry of commercial surveillance constantly profits from tracking people without identifying a single person. Despite this, identifiability is becoming a global standard that transcends legal systems, legal cultures, and regulatory frameworks and mechanisms. Currently, more than 140 countries have started to legislate EU-style data privacy standards and laws.[238] With so many jurisdictions embracing identifiability as the threshold of data privacy law, the international convergence on defining and applying

[237] *See, e.g.*, Graham Greenleaf & Bertil Cottier, *2020 Ends a Decade of 62 New Data Privacy Laws*, 163 PRIV. L. & BUS. INT'L REP. 24, 24–25 (2020).

[238] *Id.*

identifiability is seemingly becoming inevitable. A startling, yet overlooked, fact is that none of these jurisdictions gives any reason for taking up identifiability as the legal mechanism. The global community unquestioningly makes use of an unjustified concept to determine the scope of data privacy protections and consequently suffers from the uncertainty of identifiability. We are unaware of a collective blind spot, to say the least.

To eliminate this blind spot, we shall focus on two "why" questions: 1) Why should the law use identifiability to safeguard data privacy? 2) Why should the law pass over other standards? The answer to these questions inherently hinges on the conceptual relationship between identifiability and data privacy. Take the concept of fruit as an example. If the color of an object does not make it fruit, then it is unreasonable to categorize a tomato as a fruit simply because of its red color. Color should not become a decisive threshold of fruits due to the lack of a conceptual relationship. Clarifying the relationship between identifiability and data privacy is the first step toward a sustainable decision between keeping identifiability and abandoning it as the primary gatekeeper.

Identifiability may be irrelevant or relevant to data privacy. The understanding of the relationship between identifiability and data privacy reveals the role identifiability should play in data privacy regulation.

(1) Identifiability is irrelevant to data privacy (i.e., the identification of a person does not trigger data privacy violations).

(2) Identifiability is relevant to data privacy. Based on how relevant identifiability is to data privacy, the relationship can further be divided into two kinds.

   A. The identification of a person partially contributes to data privacy violations, but it is not the only factor.

   B. The identification of a person decisively triggers data privacy violations.

For those who support (1), identifiability should be abandoned. If the identification of a person cannot lead to privacy violations, the law shall not take it seriously in the judgment of privacy violations. For those who agree with (2), identifiability should remain a factor in data privacy regulation. Since the identification of a person can trigger privacy violations, the law should consider it in the judgment of privacy violations. For those who stand on (2A), identifiability is a necessary but insufficient condition for privacy violations. As a result, other factors should be considered in the judgment of privacy violations. For those who support (2B), identifiability is a necessary and sufficient condition; therefore, it should be the dominant gatekeeper in data privacy regulation.

All of the aforementioned solutions, either revising or abandoning identifiability, are rooted in assumed relationships between identifiability and data privacy. The contextual integrity theory is based on (1). As mentioned before, since disclosing identifiable data does not always trigger privacy violations, Nissenbaum asserts that identifiability is not conceptually relevant to the definition of privacy and calls for contextually informational norms to be the new gatekeeper.[239] Unfortunately, Nissenbaum makes no effort to justify her statement that identifiability is conceptually irrelevant to the concept of privacy.[240] On the basis of (2A), scholars argue that privacy violations are caused by multiple factors and promote the utility approach or group privacy theory. The utility approach regards the sensitivity of data as another vital proxy of privacy threats, whereas the group privacy approach considers identifying a group of people as compromising privacy, as well as identifying individual persons. Scholars who favor PII 2.0 accept (2B). For Schwartz and Solove, no other factors should be considered for privacy violations except identifiability.[241]

Clarifying the relationship between identifiability and data privacy is an urgent task that takes precedence over other issues for two reasons. First, such a clarification helps to rationalize an unjustified legal mechanism and eliminate a global blind spot regarding identifiability. Initially, it could be a subconscious choice, influenced by an unchecked intuition that identifiability matters to data privacy. However, a subconscious choice of gatekeeper to human welfare is not accountable lawmaking. In the age of commercial surveillance, protecting data privacy is becoming a significant aspect of human flourishing and deserves deliberate consideration of the threshold issue.

Second, clarifying the relationship between identifiability and data privacy increases the plausibility of current solutions to the threshold issue of data privacy law. Though previous solutions do not emphasize or even mention the significance of clarifying the relationship between identifiability and data privacy, they are all inescapably built upon certain unexamined assumptions of such a relationship. Moreover, those solutions are substantially determined by their assumptions. Unless scholars face their assumptions squarely, they will not be able to properly respond to the normative challenges.[242] Justifying these unfounded assumptions is the best strategy for current proposals.

[239] NISSENBAUM, *supra* note 218, at 113.

[240] *See supra* Part V.B.

[241] Schwartz & Solove, *supra* note 42, at 1879.

[242] PII 2.0 fails to justify why the risk of re-identification plays a decisive role in privacy regulations. *Contra* Schwartz & Solove, *supra* note 42. This directly links to the normative role of identifiability in data privacy. Group privacy approaches need to

VII. CONCLUSION

The world economy is evolving towards accelerating globalization and digitalization. Companies employ various tracking technologies to collect and process data in a borderless realm. They have made billions of dollars by digitizing every layer of the social and economic interactions of citizens from the global valley. Due to the lack of global regulation, each state has no choice but to unilaterally implement and, in isolation, faces a threshold issue: when should data privacy laws intervene?

It is not a challenge unique to a certain jurisdiction. After investigating data privacy laws in the EU, the United States, and China, this Article demonstrates that the common crux lies in the standard of identifiability. In these jurisdictions, regulators adopt the underlying identifiability of information as the critical gatekeeper to data privacy.[243] They stumble by shared judiciary obstacles in defining and applying identifiability. This "identifiability problem" may well exist everywhere that uses identifiability as the threshold.

Recognizing the identifiability problem fills at least three gaps in the scholarship. First, the recognition reveals a common but overlooked phenomenon in transnational data privacy regulation. Scholars have an unfounded impression that identifiability is a global standard. The existence of the identifiability problem offers a descriptive foundation for such an impression. Second, this recognition is a factual base for synthesizing global efforts to deal with doctrinal challenges on identifiability. States can learn from one another how to solve the problem doctrinally. Third, the recognition leads to a normative question: Why does identifiability matter to data privacy? If the gatekeeper cannot stably and reliably fulfill its function, what are the reasons for us to stick with it? If applying identifiability becomes too arduous in multiple jurisdictions, is it worth having it? Before proposing additional solutions, the first and foremost task is to clarify the normative role of identifiability in privacy. This must be a threshold goal of any functional transnational privacy regime.

answer why the identification of a group of people violates privacy. *See generally* Taylor, *supra* note 187. Essentially, it is a question about what kind of identifiability constitutes privacy violations. This question stems from the relationship between identifiability and privacy. The utility approach lacks a normative guidance on how to assign values to costs and benefits of a free flow of data. People in favor of the utility approach desire to know how many normative weights should be put on three factors—the risk of re-identification, sensitivity of data, and benefits of a free flow of data. *See generally* Ohm, *supra* note 213. Such a question also involves a response to the relationship between identifiability and privacy. Shedding light on the normative position of identifiability in privacy helps enhance the feasibility of these solutions.

[243] *See supra* Part IV.