Michigan Technology Law Review

Volume 26 Article 3

2019

Privacy Preserving Social Norm Nudges

Yifat Nahmias
Bar-Ilan University

PRIVACY PRESERVING SOCIAL NORM NUDGES

Yifat Nahmias
Oren Perez
Yotam Shlomo
*Uri Stemmer**

Nudges comprise a key component of the regulatory toolbox. Both the public and private sectors use nudges extensively in various domains, ranging from environmental regulation to health, food and financial regulation. This article focuses on a particular type of nudge: social norm nudges. It discusses, for the first time, the privacy risks of such nudges. Social norm nudges induce behavioral change by capitalizing on people's desire to fit in with others, on their predisposition to social conformity, and on their susceptibility to the way information is framed. In order to design effective social norm nudges, personal information about individuals and their behavior must be collected, processed, and later disseminated (usually in some aggregated form). Thus, the use of social norm nudges opens up the possibility for privacy threats. Despite the significant privacy concerns raised by social norm nudges, research on the topic has been scarce. This article makes two contributions to the understanding of the privacy risks underlying the use of social norm nudges. The first contribution is analytic: it demonstrates that using social norm nudges can pose a threat to individuals' privacy through re-identification of anonymized data. This risk was demonstrated in other contexts (e.g. Netflix recommendation contest). The second contribution is policy oriented: it argues that the strategy of differential privacy can be used to mitigate these privacy risks and offer a way to employ social norms nudges while protecting individuals' privacy.

---

* Prof. Oren Perez is the Dean of the Faculty of Law at Bar-Ilan University ("BIU"). Dr. Yifat Nahmias has recently received her PhD from the BIU Faculty of Law. Yotam Shlomo is the manager of the Environmental Regulation Clinic at the BIU Faculty of Law. Dr. Uri Stemmer is an Assistant Professor at the Department of Computer Science, Ben-Gurion University. We thank Benny Pinkas and Eyal Peer for their comments on previous versions of this paper. We are grateful for the support of the BIU Center for Research in Applied Cryptography and Cyber Security in conjunction with the Israel National Cyber Bureau in the Prime Minister's Office.

TABLE OF CONTENTS

I. INTRODUCTION

One of the most important trends in contemporary regulatory practice has been the attempt to incorporate insights from behavioral economics,[1] especially nudges and de-biasing mechanisms, into the regulatory toolbox.[2] For instance, in July 2010, the United Kingdom established the Behavioral Insights Team ("BIT") – otherwise known as the "Nudge Unit" – which is dedicated to developing public policy solutions based on insights from psy-

1. PETE LUNN, REGULATORY, POLICY AND BEHAVIORAL ECONOMICS 25–55 (2014).

2. Oren Perez, *Can Experts Be Trusted and What Can Be Done About It? Insights from the Biases and Heuristics Literature, in* NUDGING AND THE LAW: A EUROPEAN PERSPECTIVE 115 (Alberto Alemanno & Anne-Lise Sibony, eds., 2015); David Tannenbaum, Craig R. Fox & Todd Rogers, *On the Misplaced Politics of Behavioural Policy Interventions*, 1 NATURE HUM. BEHAV. 1, 1 (2017); *see also* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 6–9, 17–39 (2008) [hereinafter THALER & SUNSTEIN, NUDGE].

chology and behavioral economics.[3] Three years later, then-U.S. President Barack Obama signed an executive order establishing the White House Social and Behavioral Sciences Team, based on the understanding that "[a]dopting the insights of behavioral science will help bring our government into the 21st century in a wide range of ways."[4] These examples, among others, demonstrate how policy makers incorporate social norm nudges into the regulatory toolbox and how they use them in diverse domains, from environmental regulation and health to food and financial regulation.[5]

A nudge can be defined as any aspect of choice architecture that alters an individual's behavior in a predictable way, without limiting his choices or significantly changing incentives.[6] Using this broad definition, Thaler and

---

3. It is important to note that in 2014, the BIT became a social purpose company, partially owned by the Cabinet Office. The aims of the BIT, however, remain the same, namely: (1) making public services more cost-effective and easier for citizens to use; (2) improving outcomes by introducing a more realistic model of human behaviour to policy; and (3) enabling people to make better choices for themselves. *See generally About Us*, BEHAVIOURAL INSIGHTS TEAM, https://www.behaviouralinsights.co.uk/about-us/ (last visited Feb. 4, 2019).

4. USING BEHAVIORAL SCIENCE INSIGHTS TO BETTER SERVE THE AMERICAN PEOPLE, EXEC. ORDER NO. 13707, 80 FED. REG. 181, 1 (Sept. 15, 2015); Press Release, The White House, Office of the Press Secretary, FACT SHEET: New Progress on Using Behavioral Science Insights to Better Serve the American People (Sept. 15, 2016), https://obamawhitehouse.archives.gov/the-press-office/2016/09/15/fact-sheet-new-progress-using-behavioral-science-insights-better-serve.

5. For an extensive, although partial, summary of empirical "nudges" and interesting behavioral change interventions that have been deployed around the world in the last decade, *see* Mark Egan, *Nudge Database v1.2*, https://www.stir.ac.uk/media/stirling/services/faculties/social-sciences/research/documents/Nudge-Database-1.2.pdf (last visited Sept. 28, 2019).

6. THALER & SUNSTEIN, NUDGE, *supra* note 2 at 6, 105-10. Changing the default rule in a pension plan is an example of a nudge. Redesigning the 401(k)-retirement saving plan enrollment default rule from opt-in to opt-out dramatically increased employees' participation in retirement saving plans. Under an automatic enrollment default rule, employees are automatically enrolled in their company's 401(k) plan unless they elect to opt out of it. *See* James J. Choi, David Laibson et al., *For Better or For Worse: Default Effects and 401(k) Savings Behavior*, *in* PERSPECTIVES IN THE ECONOMICS OF AGING 81-121 (David A. Wise, ed., 2004); Brigitte C. Madrian & Dennis F. Shea, *The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior*, 116 Q.J. ECON. 1149, 1149–87 (2001). On the effect of opting-in or opting-out from default rules, *see* Alberto Abadie & Sebastien Gay, *The Impact of Presumed Consent Legislation on Cadaveric Organ Donation: A Cross-Country Study*, 25 J. HEALTH ECON. 599, 610 (2006); Eric J. Johnson & Daniel Goldstein, *Do Defaults Save Lives?*, 302 SCIENCE 1338 (2003); Eric J. Johnson et al., *Framing, Probability Distortions and Insurance Decisions*, 7 J. RISK & UNCERTAINTY 35, 46-48 (1993); Daniel Pichert & Konstantinos V. Katsikopoulos, *Green Defaults: Information Presentation and Pro-Environmental Behaviour* 28 J. ENVTL. PSYCHOL. 63, 67–69 (2008); William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 7 J. RISK & UNCERTAINTY 35 (1988). For additional examples, *see* Gerd Bohner & Lena E. Schlüter, *A Room with a Viewpoint Revisited: Descriptive Norms and Hotel Guests' Towel Reuse Behavior*, PLOS ONE 1 (Aug. 2014); Noah J. Goldstein et al., *A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels*, 35 J. CONSUMER RES. 472 (2008).

Sunstein have identified a broad range of nudges, including ones that take advantage of an individual's sensitivity to social norms.[7]

"Social norm nudges" inform people about what others are doing and thereby induce them to alter their behavior.[8] The success of a "social norm" nudge is attributed to people's propensity to use others' behavior as an appropriate guide for their own, and it is most influential when it is as personally relevant and as specific as possible.[9] The advent of social nudges represents a new way for policymakers to promote a range of welfare-enhancing behaviors. The most sophisticated and effective nudges are those that are tailored to the unique characteristics of individual citizens. However, as individual tailoring often depends on the collection and processing of personal information, privacy concerns become a fundamental problem for the use of social nudges. To date, the potential ramifications of those privacy threats have not received sufficient attention by scholars and policymakers.[10]

The most common approach for protecting individuals' privacy concerns when there is a need to use private information for public policy goals is *de-identification*. De-identification generally involves the removal of specific identifying marker, such as, but not limited to, names, e-mail addresses, social security numbers, etcetera.[11] De-identification techniques suffer, however, from significant limitations. These limitations were exposed by the famous incident in which Netflix released supposedly anonymized information regarding the viewing habits of their subscribers; the data was cross-referenced with the open dataset of IMDB reviews (an online database of media-related information), leading to the re-identification of some individuals.[12] Over the past decade, the inherent limitations of common de-identification techniques as a tool for protecting people's privacy were highlighted by various writers.[13] These limitations and the risks they pose for privacy are also relevant to social norm nudges which are based on the dissemination of private data (albeit in a processed form).

This article does not claim that such privacy risks should completely prevent regulators from using social norm nudges. Rather, this article asserts that to allay these privacy concerns, policymakers should adopt a *differential privacy* framework, which offers better protection to privacy than exist-

---

7. *See generally* THALER & SUNSTEIN, NUDGE, *supra* note 2.

8. Using nudges as a form of regulatory intervention is generally referred to as "libertarian paternalism" or "soft paternalism."

9. THALER & SUNSTEIN, NUDGE, *supra* note 2 at 53–74. *See also infra* notes 31–41 and accompanying text.

10. Andreas Kapsner and Barbara Sandfuchs' article titled "Nudging as a Threat to Privacy" is a rare exception. *See* Andreas Kapsner & Barbara Sandfuchs, *Nudging as a Threat to Privacy,* 6 REV. PHIL. PSYCH. 455 (2015).

11. *See infra* notes 146–56.

12. *See infra* notes 172–74.

13. *See infra* Part IV.

ing *ad-hoc* de-identification techniques. This framework also offers a way to balance individuals' privacy concerns with the policy advantages of social nudges.

Differential privacy is a formal mathematical framework for guaranteeing privacy protection in data analysis.[14] Differential privacy is not a single method, but rather a definition or a standard that states the requirements for any proposed privacy preserving solution.[15] It is based on the idea of introducing "noise" into data,[16] thus making it harder to de-identify any specific individual.[17] Under the framework of differential privacy, privacy is not just a property of the output, but rather a property of the computation that generates the output.[18] Thus, this technique is considered to be a relatively effective method to protect against privacy breaches and attacks.[19]

The idea that differential privacy methods can be used to mitigate privacy threats in the context of social norm nudges has not been discussed previously in the legal literature.[20] This article seek to fill this void by demonstrating, both theoretically and empirically, how regulators can use the differential privacy technique in order to protect individuals' privacy in the context of social norm nudges.

---

14. Cynthia Dwork, *A Firm Foundation for Private Data Analysis*, 54 COMM. OF THE ACM 86, 91 (2011); *see also* Alexandra Wood et al., *Differential Privacy: A Primer for A Non-Technical Audience*, 21 VAND. J. ENT. & TECH. L. 209, 219–20 (2018). Namely, a randomized algorithm A : Xn • Y is (•,) differentially private if for every two databases S, S' • Xn that differ on one row, and every set T • Y , we have Pr[A(S) • T] • e• · Pr[A(S') • T] + D.

15. Erica Klarreich, *Privacy by the Numbers: A New Approach to Safeguarding Data*, QUANTA MAG. (Dec. 10, 2012), https://www.quantamagazine.org/a-mathematical-approach-to-safeguarding-private-data-20121210/.

16. *See generally* Cynthia Dwork, *Differential Privacy: A Survey of Results*, THEORY & APPLICATIONS OF MODELS OF COMPUTATION 1, 1, 3 (Apr. 2008).

17. *See* Wood et al., *supra* note 14 at 223; *see also* Yehuda Lindell & Eran Omri, *A Practical Application of Differential Privacy to Personalized Online Advertising* 3 (Jan. 2011) (unpublished manuscript) (on file with the Cryptology ePrint Archive) ("differential privacy guarantees that hardly any information is gained on individual records upon seeing the output of a computation").

18. Wood et al., *supra* note 14 at 221.

19. *Id.* at 235. Although, as will be further discussed below, the level of protection may vary. *See infra* Part IV.

20. The references we were able to find only discuss the use of differential privacy in general. *See, e.g.*, Andrew Chin & Anne Klinefelter, *Differential Privacy As A Response to the Reidentification Threat: The Facebook Advertiser Case Study*, 90 N.C. L. REV. 1417, 1427 (2012); Anna Myers & Grant Nelson, *Differential Privacy: Raising the Bar*, 1 GEO. L. TECH. REV. 135, 139 (2016); Kobbi Nissim et al., *Bridging the Gap Between Computer Science and Legal Approaches to Privacy,* 31 HARV. J.L. & TECH. 687, 690 (2018); Andrea Scripa Els, *Artificial Intelligence as a Digital Privacy Protector*, 31 HARV. J.L. & TECH. 217, 218 (2017); Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. COLO. L. REV. 1117, 1137–40 (2013).

Using differential privacy algorithm involves, however, some tradeoffs between accuracy and privacy.[21] More "noise" yields better privacy but also less accurate computation (and, hence, poorer social norm nudges). Thus, policymakers should keep in mind that privacy comes at a cost, and that achieving "perfect" privacy is not always an option. Notwithstanding the above, this framework should enable regulators to maintain the advantages of social norm nudges while simultaneously protecting the privacy of the individuals who make up the sample.

The article is organized as follows. Part II begins with a brief introduction to the theory of nudges, focusing especially on social norm nudges. Part III develops a framework for evaluating privacy harms that could ensue when implementing social norm nudges and explores the privacy risks that may result from the use of social nudges at each and every step of the nudging process. Moreover, it focuses on privacy threats raised during, and as a result of, the communication of information, and demonstrate that these risks are likely to evolve in unforeseen ways over the course of the next few years. Part IV first examines the use of de-identification techniques in the legislative efforts to balance the utility of data with individual privacy interests. It then demonstrates that regulation that is more efficient is necessary and recommends the adoption of solutions based on differential privacy models. Finally, it discusses possible tradeoffs between privacy and the efficacy of nudges. To illustrate the threats to privacy, as well as the advantages of the differential privacy framework, this article uses both a hypothetical scenario and real data relating to water consumption. Lastly, Part V summarizes important points and concludes the article.

II. NUDGES: SETTING THE SCENE

The term "nudge" was coined by Richard Thaler and Cass R. Sunstein in their book, *Nudge: Improving Decisions About Health, Wealth and Happiness*.[22] Thaler and Sunstein define a nudge as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives."[23] In

21. It is important to emphasize that using differential privacy has its costs on the accuracy of the information that is disseminated. *See infra* Part IV.

22. *See* THALER & SUNSTEIN, NUDGE, *supra* note 2, at 6. Thaler and Sunstein are among the leading researchers in the field. Noteworthy, however, are other concepts such as asymmetry paternalism and liberal paternalism which suggest using similar instruments to "push" people towards the required direction. *See, e.g.*, Daniel Kahneman and Amos Tversky, *Judgment Under Uncertainty: Heuristics and Biases* 185 SCI. 1124, 1124 (1974); Daniel Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics*, 93 AM. ECON. REV. 1449, 1450 (2003).

23. THALER & SUNSTEIN, NUDGE, *supra* note 2, at 6.

other words, nudge theory is concerned with how to achieve predictable changes in people's behaviors, taking into account known heuristics and biases.[24] To illustrate the steering potential of nudges, Thaler and Sunstein use the example of a cafeteria.[25] They argue that people's choices are as much a result of their preferences as they are a result of the cafeteria's setup. In fact, they demonstrate that when the placement of products is altered, there is a change in the food products people will choose.[26]

The term "nudges" is sometimes used loosely. However, to be correctly characterized as a nudge, a policy tool cannot significantly change the material incentives facing the individuals (e.g., using tax or sanction),[27] nor force the individual to follow a particular path using regulatory instruments such as mandates or bans.[28] In other words, although nudges aim to push people toward a particular direction, they do not limit the original set of choices available; therefore, they are choice-preserving. A public law constraint on the use of nudges is that their use should increase the welfare of the individual being nudged or that of society as a whole.[29]

---

24. Pelle Gulddbotg Hansen, *The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?* 1 EUR. J. RISK REG. 1, 4 (2016) ("A nudge is a function of (I) any attempt at influencing people's judgment, choice or behaviour in a predictable way (1) that is made possible because of cognitive boundaries, biases, routines and habits in individual and social decision-making posing barriers for people to perform rationally in their own declared self-interests and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts.")

25. THALER & SUNSTEIN, NUDGE, *supra* note 2, at 1–3.

26. Making certain foods easily accessible, aesthetically pleasing, or placing them right near the entrance all tend to significantly increase consumption rates of certain foods. Cafeteria owners can use this information to arrange foods to achieve various goals: to encourage healthy eating, to get kids to pick the same foods that they would on their own, to maximize sales of certain foods based on suppliers, or to maximize profit. *Id.*

27. Therefore, taxes, subsidies or fines are not considered nudges. *See* Cass R. Sunstein, WHY NUDGE?: THE POLITICS OF LIBERTARIAN PATERNALISM 57-59 (Yale Univ. Press 2014); Cass R. Sunstein, *The Ethics of Nudging*, 32 YALE J. ON REG. 413, 417 (2015) ("[i]f an intervention imposes significant material costs on choosers, it might of course be justified, but it is not a nudge.")

28. Todd Haugh, *Nudging Corporate Compliance*, 54 AM. BUS. L.J. 683, 690 (2017); Cass R. Sunstein, *Nudging: A Very Short Guide*, 37 J. CONSUMER POL'Y 583, 583–584 (2014); *see also* Colin Camerer et al., *Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism"*, 151 U. PA. L. REV. 1211, 1211–12 (2003).

29. This assertion is, of course, not free of criticism. Further, "this begs the question of who makes the determination of what is good and by what measure, but as a general matter, nudges are not to be 'employed to sway people to make bad decisions that they will later regret'." *See* Haugh, *supra* note 28, at 690–91 (citing Richard H. Thaler, *The Power of Nudges, for Good and Bad*, N.Y. TIMES, Oct. 31, 2015, http://www.nytimes.com.proxy4.athensams.net/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html).

Lehner, Mont, and Heiskanen claim that nudging techniques can be grouped into four different clusters:[30] (1) simplification and framing of information,[31] (2) physical choice architecture (changing the physical environment),[32] (3) default policy,[33] and (4) social norm nudges. Though nudges can be used in a plethora of ways, this article focuses on the privacy concerns implicated in the last group of social norm nudges. However, before examining these concerns in depth, it is worth providing an overview of social norm nudges.

Social norms nudges hinge on the human tendency to follow the herd.[34] In other words, individuals' choices and preferences can be influenced by the perceived behavior of others.[35] This influence is either due to peer pressure or because the individual assumes that others have better information.[36] For instance, if an individual observes a group engaged in a particular action or perceives them to be thinking in a particular way, the individual is likely to then use that information as a signal of what is the appropriate way to think or behave.[37] Alternatively, when an individual perceives her own behavior or thoughts to be different from the one which is "socially accepta-

---

30. Matthias Lehner et al., *Nudging – A Promising Tool for Sustainable Consumption Behaviour?*, 134 J. CLEANER PRODUCTION 166, 168–69 (2016).

31. This cluster consists of tools designed mainly to make information more readily available and comprehensive, therefore eliciting a change in the individual's decision-making process. Warnings that are printed on packs of cigarettes are a good example of this type of nudging. It provides people with information regarding the dangers of smoking with the aim of nudging them to stop or at least reduce the consumption of cigarettes. *See id.* at 169.

32. The physical environment in which one makes a decision has an impact on the individual's choice. Nudges that build on this insight utilize changes to the physical environment to steer individuals towards a certain choice. *Id.* The cafeteria example discussed above can serve as a good example of such nudges. *See, e.g.*, Marvin E. Goldberg & Kunter Gunasti, *Creating an Environment in Which Youths are Encouraged to Eat a Healthier Diet*, 26 J. PUB. POL'Y & MARKETING 162 (2007); John Pucher and Ralph Buehler, *Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany*, 28 TRANSP. REV. 495 (2008).

33. Studies have demonstrated that people often prefer the road of least resistance, will not act unless they have to, and tend to maintain the status quo. Thus, people are susceptible to defaults. Consider the case of organ donation. Traditionally, organ donation forms require people to opt-in. However, studies have shown that in jurisdictions where consent to organ donation is the default, the consent rate in much higher than in jurisdictions where the individual must opt in in order to donate organs. By using default rules, policymakers are able to determine the outcome of cases in which people take no action. *See* Johnson & Goldstein, *supra* note 6, at 1338–39; Lehner et al., *supra* note 30, at 169.

34. *See* THALER & SUNSTEIN, NUDGE, *supra* note 2, at 53–74.

35. Kiran Iyer, *Nudging Virtue*, 26 S. CAL. INTERDISC. L.J. 469, 481 (2017); Anne-Lise Sibony & Alberto Alemanno, *The Emergence of Behavioral Policy-Making: A European Perspective*, *in* NUDGE AND THE LAW: A EUROPEAN PERSPECTIVE 4 (Anne-Lise Sibony & Alberto Alemanno eds. 2015).

36. *See, e.g.*, John Beshears et al., *The Effect Of Providing Peer Information On Retirement Savings Decisions,* 70 J. FIN. 1161 (2015).

37. *See* THALER & SUNSTEIN, NUDGE, *supra* note 2, at 54.

ble," she is more likely to alter her behavior "to earn social approval and avoid disapproval."[38] Further, because most people believe that others are paying attention to what they are doing, if an individual cares about what other people think of her, then she is more susceptible to go along with the norm.[39] This is illustrated in Solomon Asch's classic conformity experiment. In this experiment participants were presented with three lines (A, B, or C) and were asked to judge and verbally announce which one of those lines was most like a fourth "target line."[40] The task was fairly easy, and the answer was rather obvious. However, each participant was put in a room with seven other participants that were in fact coconspirators. These seven confederates had agreed in advance on the answer they would provide, but the naïve participant was unaware. Asch found that naïve participants were willing to ignore reality and conformed to the incorrect majority answer.[41] While this might seem anecdotal, studies on exercise, smoking, drugs, teenage pregnancy, obesity, risk-taking, health, and many other areas of life have similarly found that the perceived salient decisions of others have a significant impact on individual choices.[42]

The overall message is clear. As people tend to rely on social norms to "gain an accurate understanding of and effectively respond to social situations,"[43] informing them or emphasizing what most people should do or actually do in certain circumstances (i.e., the social norm) has the potential to change their behaviors.[44] Moreover, it could create a phenomenon of "*com-*

---

38. Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683, 686 (1999); *see also* Hunt Allcott, *Social Norms and Energy Conservation*, 95 J. PUB. ECON. 1082, 1083 (2011); Johnson & Goldstein, *supra* note 6, at 1338.

39. This phenomenon is known as the Spotlight Effect. *See* THALER & SUNSTEIN, NUDGE, *supra* note 2, at 54.

40. *Id.* at 55–57.

41. *Id.*

42. *Id.* at 54–55; Cass R. Sunstein, *Nudges.gov: Behaviorally Informed Regulation*, *in* THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW 719, 723 (Eyal Zamir & Doron Teichman eds., 2014) [hereinafter Sunstein (2014)].

43. Robert B. Cialdini & Noah J. Goldstein, *Social Influence: Compliance and Conformity*, 55 ANN. REV. PSYCHOL. 591, 597 (2004).

44. To that end, it is common to differentiate between two types of norms which policymakers might deploy: injunctive and descriptive norms. The first communicate what people ought or should do. Whereas, the latter communicates "what most people do." Robert Cialdini et al., *A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Spaces*, 58 J. PERSONALITY & SOC. PSYCHOL. 1015, 1015 (1990); Yuval Feldman & Janice Nadler, *The Law and Norms of File Sharing*, 43 SAN DIEGO L. REV. 577, 598 (2006) ("Descriptive norms are how most people would behave in comparable situations. Injunctive norms refer to the extent to which most people would approve of the target behavior."). Both injunctive and descriptive social norms are used as nudges. Elena Kantorowicz-Reznichenko & Jaroslaw Kantorowicz, *To Follow or not to Follow the Herd? Transparency and Social Norm Nudges* 6 (Inst. for Res. in Econ. & Fiscal Issues, Working Paper No.

pliance without enforcement."[45] Social norm nudge techniques make use of these precise insights.

Government and private entities use social norm nudges across various domains, including education,[46] health,[47] and energy and water conservation.[48] For example, in one of largest-scale energy saving experiments using social norms, the U.S. power company Opower implemented a nudge tailored to the relevant consumer. Opower sent Home Energy Reports to 600,000 households. The Home Energy Reports contained information on the individual's energy usage, as well as peer comparison information (i.e., tailored nudges). The letters were customized to inform recipients about their energy consumption levels in comparison to other people in similar situations.[49] Additionally, the personalized reports depicted smiley faces and associated attributes.[50] Overall, these personally tailored messages led to a decrease in energy consumption.[51]

This level of personalization, however, is not always possible. As such, there are instances where the nudging entity must resort to generic messages to elicit a behavioral change (i.e., non-tailored nudges). For instance, scholars have demonstrated that posting a generic sign near the elevators stating that most people took the stairs caused the number of individuals who used the elevator versus the stairs to drop by 46%.[52] This finding demonstrates that even simple "non-tailored" descriptive norm nudges can promote desirable social outcomes.

It is true that there might be some differences in the steps, requirements and nudging elements of a "tailored" versus "non-tailored" nudge. Howev-

---

201901, 2019), https://en.irefeurope.org/Publications/Working-Paper-Series/article/To-Follow-Or-Not-To-Follow-The-Herd-Transparency-And-Social-Norm-Nudges.

45. Rainer Baisch, *Nudging: Information, Choice Architecture and Beyond*, *in* NUDGING – POSSIBILITIES, LIMITATIONS AND APPLICATIONS IN EUROPEAN LAW AND ECONOMICS 217, 228 (Klaus Mathis & Avishalom Tor eds., 2016); Sunstein (2014), *supra* note 42, at 723.

46. *See* Benjamin L. Castleman & Lindsay C. Page, *Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?*, 115 J. ECON. BEHAV. & ORG. 144, 158 (2015).

47. Jerry M. Burger & Martin Shelton, *Changing Everyday Health Behaviors Through Descriptive Norm Manipulations,* SOC. INFLUENCE 1, 1 (2011).

48. Maria Bernedo, Paul J. Ferraro & Michael Price, *The Persistent Impacts of Norm-Based Messaging and Their Implications for Water Conservation*, 37 J. CONSUMER POL'Y 437 (2014).

49. Allcott, *supra* note 38. One problem with this assertion is that the term "similar situation" can have different interpretations. Opower compared house square footage and heating systems. *Id.* at 1083.

50. *Id.*

51. This outcome is undoubtedly beneficial to both the individual who is saving money and society as whole, as it will help to reduce pollution. Sunstein (2014), *supra* note 42, at 740.

52. Burger & Shelton, *supra* note 47, at 1.

er, based on the examples reviewed, it is possible to conclude that employing social norm nudges (whether by the state or a private entity) generally requires the collection and processing of information from which the nudging entity could deduce the applicable social norm. This information is subsequently used to signal to a specific individual or group of individuals what is the desired or socially acceptable behavior, thus opening up the possibility for privacy concerns.

The growing popularity of social norm nudges brought attention to both the advantages and drawbacks of this instrument. However, to the best of our knowledge, research on privacy implications of them remains sparse. These concerns are to be discussed in the next part.

III. Can Social Norm Nudges Pose a Threat to Privacy?

An individual's right to privacy was recognized for the first time by Samuel D. Warren and Louis T. Brandeis in their 1890 seminal article titled "The Right to Privacy."[53] Seventy years later, William Prosser further advanced the legal protection of privacy by identifying four separate torts within the "right of privacy."[54] In the years that followed, the right to privacy has become a key element in the ensemble of constitutional protections enjoyed by individuals and has been institutionalized in various ways. Still, and despite the richness of the academic literature, there is no consensus regarding the theoretical justifications or the scope of the right to privacy.[55]

---

53. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195–96 (1890). In their article, Warren and Brandeis inspired the discourse pertaining to an individual's right to privacy, which they framed as the "right to be let alone." *Id.* "[T]he authors concluded that there is a common law right of privacy which had, in some instances, been protected under the guise of property rights, and that violation of the right itself is actionable." Billings v. Atkinson, 489 S.W.2d 858, 860 (Tex. 1973); *see also* Benjamin E. Bratman, *Brandeis and Warren's The Right to Privacy and the Birth of the Right to Privacy,* 69 TENN. L. REV. 623, 624–26 (2002) (claiming that Brandeis and Warren's article "is almost universally regarded as the origin of the four invasion of privacy torts" and that judges and scholars continue to cite it as "the original source of a privacy right in American law"); Richard C. Turkington, *Legacy of the Warren and Brandeis Article: The Emerging Unencumbered Constitutional Right to Informational Privacy*, 10 N. ILL. U. L. REV. 479, 484 (1990).

54. William L. Prosser, *Privacy*, 48 CAL. L. REV. 383, 389 (1960).

55. Derek E. Bambauer, *Privacy Versus Security*, 103 J. CRIM. L. & CRIMINOLOGY 667, 672-73 (2013) ("[s]cholars and courts disagree about virtually everything" related to privacy); Robert C. Post, *Three Concepts of Privacy*, 89 GEO. L.J. 2087, 2087 (2001) ("Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all."); Daniel J. Solove, *Conceptualizing Privacy*, 90 CALIF. L. REV. 1087, 1088-90 (2002) [hereinafter Solove, *Conceptualizing*]; Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 485 (2006) [hereinafter Solove, *A Taxonomy*]. First, there is disagreement as to whether privacy is a right in and of itself, or whether it is a label used to protect other interests. Second, there is disagreement regarding the values and interests served by privacy.

In fact, the concept of privacy is so complex and the literature concerning it is so vast that it cannot be addressed within the confines of this article.[56] That being said, it is sufficient to note the threats social norm nudges pose for privacy. Building on the work of Daniel Solove,[57] this article contends that to adequately explore the plausible threats to privacy, one need not focus on the attempts to offer an overarching abstract notion of privacy,[58] but concentrate on a variety of activities that could be harmful to privacy.[59]

As noted in the previous section, any attempt to implement social norm nudges will include, at minimum, the following three steps:

The term privacy has been used as an umbrella concept encompassing one's right to be left alone, limited access to the self, secrecy, control over personal information, personhood, and intimacy. *See* Frederick Davis, *What Do We Mean by 'Right to Privacy'*, 4 S.D. L. REV. 1 (1959); Ruth Gavison, *Privacy and the Limits of Law*, 89 YALE L.J. 421, 422 (1980); Bert-Jaap Koops & Ronald Leenes, *'Code' and the Slow Erosion of Privacy*, 12 MICH. TELECOMM. & TECH. L. REV. 115, 135 (2005); *see also* Danielle K. Citron & Leslie M. Henry, *Visionary Pragmatism and the Value of Privacy in the Twenty-First Century*, 108 MICH. L. REV. 1107 (2010); Harry Kalven, Jr., *Privacy in Tort Law – Were Warren and Brandeis Wrong?*, 31 LAW & CONTEMP. PROBS. 326, 327 (1966); Tommaso Zeccherini, *Privacy: Theoretical Analysis of the Value of this Concept for Greater Protection of Personal Data*, SSRN (March 20, 2017) http://dx.doi.org/10.2139/ssrn.2965925.

56. *See* J. THOMAS MCCARTHY, THE RIGHTS OF PUBLICITY AND PRIVACY §5.59 (2d ed. 2009); ARTHUR R. MILLER, THE ASSAULT ON PRIVACY: COMPUTERS, DATA BANKS AND DOSSIERS 25 (1971). This could be attributed, at least partially, to the fact that privacy "is a product of norms, activities, and legal protections. As a result, it is culturally and historically contingent." Daniel Solove, *The Meaning and Value of Privacy*, *in* SOCIAL DIMENSIONS OF PRIVACY: INTERDISCIPLINARY PERSPECTIVES 71, 74 (Dorota Mokrosinska & Beate Roessler eds., 2015) [hereinafter Solove, *The Meaning and Value of Privacy*]. *See generally* Helen Nissenbaum, *Privacy As Contextual Integrity*, 79 WASH. L. REV. 119, 146–55 (2004) (discussing the different values advanced by the right to privacy); Scott Skinner-Thompson, *Outing Privacy*, 110 NW. U. L. REV. 159, 178–83 (2015) (discussing privacy and autonomy); Solove, *Conceptualizing*, *supra* note 55, at 1145–46 (discussing different conceptions of privacy); Lior Jacob Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHI. L. REV. 919, 923–25 (2005) (discussing different theories and values); Omer Tene & Jules Polonetsky, *A Theory of Creepy: Technology, Privacy and Shifting Social Norms*, 16 YALE J.L. & TECH. 59, 72–80 (2013) (discussing privacy values and norms).

57. Daniel Solove claims that many scholars have attempted to conceptualize privacy by pointing to a common denominator that encompasses what we all view as private. However, this method "faces a difficult dilemma. If we choose a common denominator that is broad enough to encompass nearly everything, then the conception risks the danger of being overinclusive or too vague. If we choose a narrower common denominator, then the risk is that the conception is too restrictive." Solove, *The Meaning and Value of Privacy*, *supra* note 56, at 74.

58. Solove, *A Taxonomy*, *supra* note 55, at 480–82; *see also* Citron & Henry, *supra* note 55, at 1109 (discussing Solove's book *Understanding Privacy*).

59. This approach, contends Solove, would allow policymakers to make meaningful choices and balance privacy concerns against other compelling interests in specific contexts. *See* DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 87–89 (2008) [hereinafter SOLOVE, UNDERSTANDING PRIVACY].

Step 1: Information collection;
Step 2: Information analysis; and
Step 3: Information dissemination.

This multi-stage process broadly corresponds to the first three groups of activities included in Solove's taxonomy of privacy harms (information collection, processing, and dissemination).[60] These activities are not all inherently harmful or problematic from a legal standpoint but raise the possibility of privacy threats in each step. Thus, these groups of activities are a good starting point for assessing the threats social norm nudges pose for privacy.

Consider the following example. "A-Water," a city water corporation, is interested in promoting water conservation and eliciting a behavioral change among its customers. To do so, a member of the corporation's managerial board decided to employ social norm nudges in order to reduce water consumption levels.[61] Consequently, the corporation decided to send each household a customized monthly report containing information about the customer's monthly water consumption and a comparison of the aforementioned water usage to the median consumption in the relevant group of his neighbors. The members of the relevant group were determined by the company based on certain characteristics of the household (e.g., address, number of rooms, and number of members in the household). As a practical matter, in order to implement the desired social-norm nudge, the corporation must first *collect information* regarding their customers' water consumptions levels, which will be complemented by other attributes. Subsequently, they must then *process the information,* in order to compute the relevant social norm and determine the relevant group of neighbors. Finally, the corporation will need to *communicate the appropriate information* to each and every household. In each of these stages, the activities performed by the corporation could potentially pose a threat to privacy. While our article emphasizemphasizes the privacy concerns raised primarily during the last stage, it is important to be mindful of all the risks involved in this activity.[62]

A. *Information Collection*

To design a social norm nudge, a corporation must first collect information pertaining to the water consumption of its customers as well as their

60. Solove divided privacy-offensive activities into four principal groups: (1) information collection, (2) information processing, (3) information dissemination, and (4) invasion. Solove, *The Meaning and Value of Privacy*, *supra* note 57, at 77; *see also* SOLOVE, UNDERSTANDING PRIVACY, *supra* note 59, at 106–70.

61. Bernedo, Ferraro & Price, *supra* note 48, at 437.

62. Note that this example is similar to the Opower study and therefore the threats it presents are not merely theoretical. Allcott, *supra* notes 38, 49 and accompanying text.

household's characteristics. Solove identifies two forms of information collection: (1) surveillance and (2) interrogation.[63]

1. Surveillance

Surveillance is intrinsically troublesome from a privacy standpoint when it involves spying. Solove, however, emphasizes that surveillance could encompass a larger set of activities, including watching, listening to, or recording an individual.[64] These activities would be problematic in some contexts, but not in others. For example, when an individual is walking on the street, she expects others to watch her as she passes by, and maybe even listen to her conversation. When the scrutiny and monitoring are continuous, however, these activities are likely to make her feel uncomfortable and anxious.[65] In addition to this, constant monitoring has been shown to lead to self-consciousness and self-censorship,[66] resulting in "a subtle yet fundamental shift in the content of our character."[67] Hence, surveillance could bring about colossal negative consequences.

To illustrate these consequences, consider the famous metaphor of the Panopticon. The most well-known example is Jeremy Bentham's proposed prison which depicts a prison designed as a circular shape, with an watchman siting in a central tower from which he can oversee the activities taking place by the inmates in all of the cells.[68] By creating the illusion of constant surveillance, Bentham's prison design aimed to deter aberrant practices and affect the inmates' behavior.[69] Although seemingly distinct from one another, both the Panopticon and social norm nudges employ surveillance tech-

---

63. Solove, *A Taxonomy*, *supra* note 55, at 491.

64. *Id*. at 491–99.

65. *See* Christopher Slobogin, *Public Privacy: Camera Surveillance of Public Places and the Right to Anonymity*, 72 MISS. L.J. 213, 237–67 (2002).

66. Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193, 1260 (1998); Solove, *A Taxonomy*, *supra* note 55, at 498–99; *see also* Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609, 1656 (1999) ("[P]erfected surveillance of naked thought's digital expression short-circuits the individual's own process of decisionmaking.").

67. Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object,* 52 STAN. L. REV. 1373, 1426 (2000).

68. *The Panopticon*, UCL, https://www.ucl.ac.uk/bentham-project/who-was-jeremy-bentham/panopticon (last visited Sept. 30, 2019).

69. *See also*, DAVID LYON, THE ELECTRIC EYE: THE RISE OF THE SURVEILLANCE SOCIETY 62–63 (1994); Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1264, 1267 (2004). In his book Discipline and Punish, Foucault notes that the architecture that allows surveillance by the state is a form of power that modifies behavior and allows better control. Massimo Sargiacomo, *Michel Foucault, Discipline and Punish: The Birth of the Prison*, 13 J. MGMT. & GOVERNANCE 269 (2009).

niques in an effort to appraise the individual's behavior and control it.[70] In the Panopticon example, the goal was to help identify problematic behaviors among inmates. With social norm nudges, the objective is to provide the nudging entity with the necessary information to determine whether an individual deviates from a certain standard of behavior.[71] In our context, A-Water Corporation must continuously monitor its customers' water consumption in order to deduce the social norm and whether or not any single individual's behavior deviates from it. Awareness of this continuous monitoring might have adverse psychological effects especially if it becomes part of a wider surveillance network.[72] At the same time, however, these inhibitory effects will enhance the desired outcome of the nudge implemented by the corporation.[73] Thus, striking a balance between the potential regulatory utility of a nudge and individuals' privacy concerns constitutes an important challenge for regulatory agencies.

Surely, one could maintain that A-Water's customers gave the corporation an implicit (if not explicit) authorization to constantly monitor their water usage. After all, water companies are authorized to bill consumers for water supplied within their network and must monitor water usage in order to do so. However, it seems problematic to assume that this consent also extends to using the information for the purposes of designing a social norm nudge. It would prove to be even more problematic as the use of social norm nudges increases and spreads to other areas.

2. Interrogation

The second form of information collection is interrogation. Interrogation, according to Solove, is a means of acquiring information through compulsion, or pressuring the individual to divulge information.[74] Solove clarifies that the compulsion need not be direct nor rise to the level of coercion; it could be subtle or indirect. For example, subtle interrogation could be

---

70. *See* James P. Nehf, *Recognizing the Societal Value in Information Privacy,* 78 WASH. L. REV. 1, 12 (2003) ("One difference between modern 'dataveillance' and the pervasive observation of the Orwellian and Panopticon worlds is that we are 'watched' not through a camera or guard tower, but by a computer collecting facts and data").

71. In this context, one could invoke George Orwell's "Big Brother" metaphor. This is by no means the first utilization of this metaphor as it has been invoked by numerous privacy scholars. *See e.g.*, Michael Froomkin, *The Death of Privacy?*, 52 STAN. L. REV. 1461, 1463 (2000); Nehf, *supra* note 70, at 10–11; Daniel J. Solove*, Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1396 (2001) [hereinafter Solove, *Privacy and Power*].

72. Solove, *A Taxonomy*, *supra* note 55, at 498–99; Kang, *supra* note 66, at 1260; Schwartz, *supra* note 66, at 1656.

73. For further discussion of personalized nudges, *see supra* notes 49–52 and accompanying text.

74. Solove, *A Taxonomy*, *supra* note 55, at 499–500.

seen in cases in which an individual is forced to answer probing questions out of fear of losing her job or suffering social criticism.[75] In these scenarios, the mere refusal to answer could create an appearance of having something to hide, and therefore make the subject feel uncomfortable. Interrogation could pose a threat to privacy as it is invasive and often involves the revealing of private information.[76] It is possible to imagine certain instances in which the individual will allow data collection for the purpose of social norm nudging only to avoid being perceived as "having something to hide." However, out of the two alternatives (i.e., surveillance and interrogation), interrogation seems less likely to pose a threat to privacy in the context of social norm nudges.

B. *Information Processing*

The second group of activities identified by Solove involves the "use, storage, and manipulation of data that has been collected."[77] Namely, five forms of information processing: (1) aggregation, (2) identification, (3) insecurity, (4) secondary use, and (5) exclusion.[78]

1. Aggregation

Aggregation is combining bits of information about a specific person in new ways. The idea is that an isolated piece of information might not be very revealing.[79] However, when combined, the different pieces could generate new information about the individual that she did not expect to be revealed about her when the original "isolated" bits of information were gathered.[80] For example, imagine an employee in a hauling company. The company has integrated a GPS system into their fleet of vehicles in order to promote efficiency. The GPS system routinely collects and aggregates information. From this data, the company can learn how long it takes the employee to complete his daily route, how long he needs to spend at each stop, and so forth. However, through the GPS system, the company can also gather information pertaining to the employee's daily routine and lifestyle. The aggregation of information is not, in and of itself, necessarily harmful. It could, in fact, serve a wide array of positive purposes. It does, however, carry broad implications for privacy, especially when the information is incomplete or disconnected from the original context in which it was gath-

75. *See id.* at 500–02.
76. *Id.*
77. *Id.* at 504.
78. *Id.* at 505.
79. *Id.* at 507.
80. *Id.* at 507-08.

ered.[81] In those instances, the information could misrepresent and mischaracterize the individual.

2. Identification

Identification goes an additional step beyond aggregation, by linking the information to a particular individual.[82] As in the case of aggregation, identification could have some benefits for the individual who is the subject of the information collection; for instance, it can assure that only she can access her records. At the same time, identification diminishes the individual's ability to remain anonymous, which is recognized as an important pillar of the right to privacy.[83] For example, using a store loyalty card could reveal that a particular consumer has gone through a certain medical procedure or has a specific medical condition.[84]

To better illustrate how aggregation and identification activities could pose a threat to privacy in the context of social norm nudges, let us return to our water corporation example. Aggregating information concerning customers' water usage is a critical component of A-Water's plan. The company must accumulate and combine different bits of information to form a portrait of each household's water consumption, as well as to deduce the social norm. On the one hand, this portrait could help induce more conscious water consumption usage among the company's customers. On the other hand, however, it raises privacy concerns as it involves combining information in new ways. The company's customers likely have some expectations with regard to the kind of information the company can collect about them and what could be done with the information. They are likely to assume that the company monitors their monthly water consumption and that it may have access to other personal information (e.g., name and billing information). The aggregation could, however, potentially upset these expectations "because it involves the combination of data in new, potentially unanticipated ways to reveal facts about a person that are not readily known."[85]

One might argue that in our scenario, the harm to privacy is so marginal that it is insignificant, given the fact that the information in question pertains to water consumption. However, aggregation and identification may result in losses to privacy even when the featuring information seems unimportant or detached from the data subject.[86] This privacy loss is particularly

---

81. *Id*. at 507-12.

82. *Id.* at 515.

83. *Id.* at 511-13.

84. Adam B. Thimmesch, *Tax Privacy?*, 90 TEMP. L. REV. 375, 426 n.240 (2018) (citing Solove, *A Taxonomy*, *supra* note 55, at 511–13).

85. Solove, *A Taxonomy*, *supra* note 55, at 507.

86. In this respect, it is important to note the importance of personal information and personally identifiable information. For further discussion, see *infra* Part IV(A).

relevant when the information is used to single out a specific person.[87] For example, if you had an extremely high level of water consumption. This harm to privacy will be even more severe if not only your water provider, but also your local health care clinic, were to aggregate information in order to induce behavioral change, seeing as medical information is generally perceived as highly sensitive. The higher the perceived sensitivity of the information, the higher the privacy concerns.

3. Secondary Use

Secondary use of information, which was collected and processed for one purpose and then later used for another, is an additional source of concern, mainly due to the lack of control one has over their personal information.[88] Solove further emphasizes that when put into secondary use, the information is "removed from the original context in which it was collected."[89] Thus, the data subject is likely to be misjudged or misunderstood.[90] Today, the number of possible secondary uses for personal data is almost unlimited. This unlimited potential is what makes the secondary use so threatening to the individual.[91] A clear example of this type of threat is the Cambridge Analytica scandal where the company harvested the personal data of 87 million Facebook users without their consent and used it for other purposes, including political purposes.[92]

It is not entirely clear whether processing water usage information for the purpose of nudging would constitute a secondary use or not. As will be further discussed below, this depends to a large degree on whether or not the A-Water's customers have consented to their information being processed for the purpose of implementing social norm nudges. One might argue that even if the company's usage is indeed a secondary use, it does not merit consideration because it does not cause actual harm; on the contrary, the nudge could help people be more environmentally aware and reduce their expenses by reducing their water consumption. Moreover, it could create an

87. Thimmesch, *supra* note 84, at n.240.
88. *See id*. at 381.
89. Solove, *A Taxonomy*, *supra* note 55, at 522.
90. Nehf, *supra* note 70, at 23–24.
91. *Id.*
92. Known as the Facebook–Cambridge Analytica data scandal. *See* Bill Hutchinson, *87 Million Facebook Users to Find Out if Their Personal Data Was Breached,* ABC NEWS (Apr. 9, 2018), https://abcnews.go.com/US/87-million-facebook-users-find-personal-data-breached/story?id=54334187; *see also* Kathleen Chaykowski, *Mark Zuckerberg Addresses 'Breach Of Trust' In Facebook User Data Crisis,* FORBES, (Mar. 21, 2018), https://www.forbes.com/sites/kathleenchaykowski/2018/03/21/mark-zuckerberg-addresses-breach-of-trust-in-facebook-user-data-crisis/#72f8c3e367a0.

inadequate impression of the individual.[93] Thus, the potential harm to the individual posed by such secondary use, although not financial in nature, includes damages to dignity and reputation. This is particularly true given the limited efficacy of de-anonymization techniques, as will be further elaborated in the next chapter.[94] Further, although people might change their behavior in response to the company's secondary use of the information (i.e., nudge), it would increase their insecurity.

4. Insecurity

Regardless of how much money is invested in data security protocols, there is always a possibility of a data breach.[95] Over the past few years, the number of successful data breaches has continued to rise.[96] In fact, data breaches have been documented even for very large and successful companies such as Facebook, Yahoo, and Home Depot.[97] Even if the information collected and processed is not targeted or abused by thieves, these breaches could still cause considerable harm to the data subjects.[98] Consider, for example, the case of PumpUp, a popular fitness app with over six million users, which left a server exposed without a password, allowing anyone to access users' health data and private messages.[99] This instance did not involve a malicious data breach, but it highlights the prevalence of security issues today. Insecurity is the name Solove uses to describe this group of privacy concerns,[100] namely identity theft, security lapses, abuses and other illicit uses of information.[101] "[I]nsecurity is the injury of being placed in a weak-

93. Kelsey L. Zottnick, *Secondary Data: A Primary Concern*, 18 VAND. J. ENT. & TECH. L. 193, 209 (2015); *see also* Solove, *A Taxonomy*, *supra* note 55, at 520–21.

94. For further discussion, *see infra* Part IV(A).

95. Adam R. Pearlman & Erick S. Lee, *National Security, Narcissism, Voyeurism, and Kyllo: How Intelligence Programs and Social Norms Are Affecting the Fourth Amendment*, 2 TEX. A&M L. REV. 719, 762 (2015).

96. *See* Ponemon Institute, *2018 Cost of a Data Breach Study: Global Overview* 3 (July 2018), https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=55017055USEN.

97. *See* Daniel J. Marcus, *The Data Breach Dilemma: Proactive Solutions for Protecting Consumers' Personal Information*, 68 DUKE L.J. 555, 556 (2018).

98. The Ponemon Institute defines a data breach as "an event in which an individual's name and a medical record and/or a financial record or debit card is potentially put at risk — either in electronic or paper format. In our study, we identified three main causes of a data breach: malicious or criminal attack, system glitch, or human error. The costs of data breach vary according to the cause and the safeguards in place at the time of the data breach." *See* Ponemon Institute, *supra* note 96, at 8.

99. Zack Whittaker, *Fitness App PumpUp Leaked Health Data, Private Messages* (May 31, 2018), https://www.zdnet.com/article/fitness-app-pumpup-leaked-health-data-private-messages/.

100. Solove, *A Taxonomy*, *supra* note 55, at 516–20.

101. *Id*. at 517.

ened state, of being made more vulnerable to a range of future harms."[102] In other words, the threat to privacy posed by insecurity is connected to future injury resulting from inadequate protection of collections of personal data.[103] In our context, insecurity may not be a problem when discussing the information held by the A-Water Corporation, as it hard to imagine thieves aiming to steal one's water consumption data. But if the corporation does not take the necessary steps to protect the collected data, in the case of a data breach, the overall information stored by the company can be accessed and read by third parties. Such information includes sensitive information such as customers' names, home addresses, social security numbers, marital status, and bank account information.[104]

5. Exclusion

The last form of information processing discussed by Solove is exclusion. Exclusion is "the failure to provide individuals with notice and input about their records."[105] Solove claims that exclusion distances the individual from the data collected about her, and therefore creates a sense of defenselessness, similar to secondary use of data and insecurity.[106] This is likely to pose a major problem in instances where the nudging party might not want the target of the nudge to be aware of the information collection and processing, as that knowledge might affect their behavior.

From a legal standpoint, processing personal information is generally prohibited, unless the law expressly allows it. In the EU for example, Article 6 of the General Data Protection Regulation ("GDPR") sets the six lawful bases for processing information.[107] If no lawful basis is applicable, the data controller should seek consent from the data subject. The basic conditions for a valid legal consent are defined in Article 7 and specified further in recital 32 of the GDPR.[108] In particular, any request for consent must be

---

102. *Id*. at 519.

103. *Id*. at 517–18.

104. It could also be the reason for other harms identified by Solove, as will be discussed below (e.g., intrusion, breach of confidentiality, disclosure, exposure, identification, and blackmail). *See infra* Part III(C).

105. Solove, *A Taxonomy*, *supra* note 55, at 523.

106. *Id.* at 523–24.

107. Among these are: if the processing is necessary to carry out the data controller's contractual or legal obligations, if the processing is necessary for the performance of a task which is in the public interest, and if the processing is necessary for the data controller's legitimate interests or the legitimate interests of a third party. *See* Commission Regulation 2016/679, art. 6, 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 on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1, 7–10 [hereinafter Commission Regulation 2016/679].

108. *Id.* at 6.

clear and in plain language. One of the key requirements is that consent be informed and specific. Consequently, data subjects must, at least, be informed about the data controller's identity, what kind of data is being processed, the purpose of processing, and how the information will be used.[109]

In contrast, in the United States there is no comprehensive, all-embracing regulation which deals with information privacy or data processing and sharing.[110] Instead, there is a combination of federal and state laws, many of which are sector-specific, and therefore apply only to a particular category of data, such as health or financial information.[111] Other laws apply to a specific activity, such as telemarketing and commercial e-mails.[112] Given the wide diversity of data protection and privacy statutes, it is hard to argue that consent is always required in the United States. Nevertheless, it is a key factor in many of the statutes.[113]

Under these circumstances, a reasonable approach is to require the nudging entity to get the consent of its target population for the data processing. While a thorough review of the literature addressing this issue is well beyond the scope of this article, it is worth mentioning that regardless of the standard used, the idea of informed consent as a way to circumvent

---

109. *Id.* at 8. In addition, they need to be informed of their right of withdrawal. Therefore, silence or inactivity would not be considered sufficient to indicate consent for information processing. *Id.* at 37.

110. *See* Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1129–30 (2015).

111. *See, e.g.*, Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996); Fair Credit Reporting Act, Pub. L. No. 91-508, 84 Stat. 1114 (1970) (codified at 15 U.S.C. § 1681); Fair and Accurate Credit Transactions Act of 2003, Pub. L. No. 108-159, 117 Stat. 1952 (2003); Children's Online Privacy Protection Act of 1998, Pub. L. No. 105-277, 112 Stat. 2681-728 (1998); Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999); *see also* Julia Palermo, *You Say "Tomato," I Say "Tomahto": Getting Past the Opt-in v. Opt-out Consent Debate between the European Union and United States*, 9 GEO. MASON J. INT'L COM. L. 121, 133 (2017). In addition, there are two federal informational privacy law that have a more general scope, "[t]he Privacy Act of 1974, 5 U.S.C. §552a (2006), and the Children's Online Privacy Protection Act of 1998 (COPPA), Pub. L. No. 105-277, 112 Stat. 2581 (1998). The scope of the former is determined according to the context and the identity of the parties: the government and citizens. The scope of the latter is determined according to the age (under 13) of the data subjects." Michael Birnhack, *Reverse Engineering Informational Privacy Law*, 15 YALE J.L. & TECH. 24, 56 (2012).

112. *See* 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, 1831–36 (2011) [hereinafter Schwartz & Solove (2011)]; Paul M. Schwartz & Daniel J. Solove, *Reconciling Personal Information in the United States and European Union*, 102 CAL. L. REV. 877, 889 (2014) [hereinafter Schwartz & Solove (2014)].

113. *See* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE 55–60 (2012); Daniel L. Macioce, Jr., *PII in Context: Video Privacy and a Factor-Based Test for Assessing Personal Information*, 45 PEPP. L. REV. 331, 341 (2018); Schwartz & Solove (2011), *supra* note 112, at 1825, 1829.

the protection of information is not without criticism.[114] In fact, in certain circumstances, such as when an individual is seeking medical care, the so-called informed consent is virtually meaningless because of the underlying power disparities.[115] Furthermore, most people rarely read standard data contracts.[116]

Suppose that in our previous example, the A-Water's customers have consented to its information processing practices when signing the company's service contract. However, if a customer did not negotiate the terms of service and, in effect, has no other choice but to accept the terms if she wants to be connected to the city's water-supply system, it is questionable whether her consent was indeed "informed consent." Moreover, although consent is an important feature of privacy and data protection laws, it is unclear whether the fact that the customer consented to the collection of information for the purposes of water supply, or other services granted by a corporation, means that she consents to the processing and communicating of this information for the purpose of nudging.

In sum, the processing of information required for the implementation of the social norm nudge could pose a threat to privacy, particularly if A-Water Corporation does not take adequate steps to protect the data. Privacy failures have happened in the past and, in all probability, will happen again in the future. Personal information has been commodified into a tradable asset in the last few decades. Even though the data collection and processing in our example was intended to help customers by inducing greater awareness of their water consumption, the information collected, aggregated, and created by our hypothetical water company could have unintended negative consequences. For example, the information could end up in the hands of hackers or thieves. Further, the compilation of information could be used in other settings, not all of which are related to the one for which the data subject agreed to share his or her information. Hence, information processing can become what Solove describes as an "architecture[] of vulnerability," which places the individual in a position of weakness.[117] Overall, because the data subjects have no means to check or validate the accuracy of the information, they are likely to feel vulnerable. This vulnerability would be in-

---

114. *See* Mark MacCarthy, *New Directions In Privacy: Disclosure, Unfairness and Externalities*, 6 I/S: J.L. & POL'Y FOR INFO. SOC'Y 425, 427 (2011); Aleecia McDonald & Laurie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J.L. & POL'Y FOR INFO. SOC'Y 543, 544 (2008); Joel R. Reidenberg, *Privacy in the Information Economy: A Fortress or Frontier for Individual Rights?*, 44 FED. COMM. L.J. 195, 212 n.87 (1992); Solove, *Privacy and Power*, *supra* note 71, at 1454.

115. "When people seek medical care, among the forms they sign are general consent forms which permit the disclosure of one's medical records to anyone with a need to see them." Solove, *Privacy and Power*, *supra* note 71, at 1454.

116. Solove, *A Taxonomy*, *supra* note 55, at 520.

117. SOLOVE, UNDERSTANDING PRIVACY, *supra* note 60, at 178.

creased, and even more evident, if our social norm nudge was to be implemented by an entity operating in areas such as education, health or financial markets.

However, collecting and processing information alone will not, and cannot, help A-Water "push" its customers to consume less water. For the company to induce a change in behavior it must influence what the customers perceive to be the prevailing social norm or social custom. In other words, the company ought to communicate information concerning the particular household's individual water consumption and the median water consumption of the neighborhood. This comparison is a major element in any social norm nudge, and one that is at the heart of our discussion.

C. *Information Dissemination*

The third group of harmful activities refers to activities which "involve the spreading or transfer of personal data or the threat to do so."[118] These activities include breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation, and distortion. This is where a rather interesting dynamic that is not often talked about can amplify the threat to privacy posed by social norms nudges.

1. Disclosure and Breach of Confidentiality

A disclosure consists of two elements: (1) the release of true information about the individual to third parties and (2) the influence of such disclosure on people's evaluation of the targeted individual.[119] When information about an individual is made public and she is no longer able to keep it secret, there is a chance it will humiliate her, damage the trust she has for authority, harm her reputation, or deter others from associating with her.[120] Breach of confidentiality involves not only dissemination of information, but betrayal. When individuals entrust their information with a specific entity, be it the government, the city's water corporation or the bank, they are doing so with an understanding (either explicit or implicit) that their information will be kept confidential.[121] When there is a breach of confidentiality, this promise is broken, and the individual may consequently feel deceived.[122]

When implementing a social norm nudge with the aim to increase or decrease the likelihood of engaging (or not) in a particular behavior, the nudging entity could place the individual under scrutiny by her peers for

118. Solove, *The Meaning and Value of Privacy*, *supra* note 56, at 78.
119. *Id.* at 78–79.
120. Solove, *A Taxonomy*, *supra* note 55, at 546.
121. *Id.* at 527.
122. *See* Solove, *The Meaning and Value of Privacy*, *supra* note 56, at 78–79.

noncompliance with the social norm. Thus, the nudging entity may be indirectly responsible for those negative social consequences that might follow the communication of information about the individual.[123] This article argues that these harms are particularly prevalent when it comes to social norm nudges. This is because the nudging entity, in the process of designing the nudge, not only disseminates information but also redraws the line between a normal, or acceptable, social behavior and an abnormal behavior.

To illustrate, take David and Rose. They are clients of the city water corporation and have been living in the same house for the last five years with their daughter. The couple has always considered themselves to be an environmentally-conscious household. They recycle, purchase energy-saving appliances and use environmentally-friendly household and personal hygiene products. David and Rose also publicly advocate for more environmentally-aware consumption and are actively involved in protests and demonstrations to raise awareness of the environment. Last month, they received the first monthly bill following the implementation of the social norm nudge by A-Water Corporation. According to the information presented in the monthly bill, the couple learned that their household is consuming a substantially higher level of water in comparison to what is considered to be the social norm (i.e., median consumption).

Revealing the couple's water consumption in comparison to the median water consumption – what is considered to be socially acceptable – especially in times of drought, could have a negative impact on their reputation and the way they are perceived by their community.[124] In other words, A-Water's dissemination of information could place the couple under scrutiny by their peers for noncompliance with the social norm. As a consequence, it could profoundly affect their self-image and self-esteem. Even the disclosure of the median consumption level, without revealing individual water consumption, could pose a threat to David and Rose's privacy. Suppose that following the implementation of A-Water's nudging program, it emerged that the water consumption of the residents of David and Rose's neighborhood was significantly higher than other neighborhoods. People who be-

123. "The more difficult issues arise not with informational nudges designed to engage the audience on a conscious, cognitive level, but with nudges that work on a sub-conscious or emotional level." Mark Schweizer, *Nudging and the Principle of Proportionality*, *in* NUDGING-POSSIBILITIES, LIMITATIONS AND APPLICATIONS IN EUROPEAN LAW AND ECONOMICS 93, 99 (2016). Further, as Solove's taxonomy demonstrates, there are other dignitary harms beyond reputational injury, namely lack of respect and emotional anxiety. Solove, *A Taxonomy*, *supra* note 55, at 486.

124. Commonly, social shaming and ostracism are not directly attributable to the state. However, when implementing a social norm nudge with the aim to induce a certain behavior, the state could be indirectly responsible for those negative social consequences. Mark Schweizer, *Nudging and the Principle of Proportionality*, *in* NUDGING-POSSIBILITIES, LIMITATIONS AND APPLICATIONS IN EUROPEAN LAW AND ECONOMICS 93, 99 (2016).

come aware of this comparison may – rightfully or wrongfully – use this information to draw conclusions about the couple. As explained by James Nehf, "[t]he resulting judgments might be proved correct in the aggregate . . . but they can be unfair in individual cases."[125]

Implementing this type of nudge, without proper consent, could be considered a breach of confidence. To demonstrate this, let us return to Rose and David. At first, Rose and David decline to believe this information. They even called the water company's customer service to check whether there were leaks or any other technical problems. When no issues were found, they realized that they were not as environmentally aware as they thought themselves to be, which consequently affected their self-image. At the same time, they began wondering what else the company might be using their information for and who else might have access to it. When initially contracting with A-Water Corporation, they were under the impression that the company would utilize the information they provided primarily to monitor their monthly water consumption and generate a monthly bill. But more than that, they implicitly (or explicitly) trusted the company to protect their information. Now that the company has begun to send personalized messages, they feel that the company has breached their confidentiality. Not only that, but they started dreading the possibility that one of their neighbors, or members of their activist group, would find out about their high-water consumption levels.[126]

In marketing, the use of personalized or targeted advertising has been found to raise consumers' privacy concerns; however, firms that were able to provide consumers with some control over their personally identifiable information were more likely to attract consumers through targeted ads.[127] This suggests that if people realize that their personal usage information is being used for creating the social-norm nudge, they might react against the firm in a way which could jeopardize the effects of social-norm nudge.[128] If individuals were to lose trust in governments or other public institutions which deploy social norm nudges, this would have a negative impact on the efficacy of social norm nudges, especially in times where the proliferation of "fake news" and misinformation lead to increasingly public skepticism.[129]

---

125. Nehf, *supra* note 70, at 25.

126. Solove classifies wrongful disclosure as "spreading information beyond expected boundaries" that can potentially risk the reputation of a person. *See* Solove, *A Taxonomy*, *supra* note 55, at 532.

127. *See* Catherine E. Tucker *Social Networks, Personalized Advertising, and Privacy Controls*, 51 5 J. MARKETING RES. 546, 546 (2014).

128. *But see infra* Part V, where we suggest using differential privacy as a way to mitigate some of these concerns.

129. *See* Nabiha Syed, *Real Talk About Fake News: Towards a Better Theory for Platform Governance*, 127 YALE L.J.F. 337, 343 (2017).

The idea that David and Rose will feel ashamed and humiliated if their neighbors will learn about their excessive water usage might be dismissed by some. However, to the extent to which this hypothetical couple includes being environmentally-conscious in their self-identity, it is easy to imagine how such information will negatively affect their public and self-image, particularly in the context of online shaming.[130] Even further, suppose David or Rose decide to run for some governmental office in their city, like city mayor. Then, this idea about shaming tactics in the age of social media would be even more pertinent.

2. Exposure, Increased Accessibility, Blackmail, Appropriation, and Distortion

The remaining subcategories of dissemination of information identified by Solove are exposure, increased accessibility, blackmail, appropriation, and distortion.[131] These activities, however, are less applicable to the context of social norm nudges and so they will not be discussed further.

---

130. *See, e.g.*, Emily Laidlaw*, Online Shaming and the Right to Privacy*, 6 LAWS 3, 3 (2017).

131. Exposure is somewhat similar to disclosure, as it involves revealing information about an individual; but, the two subcategories diverge in that exposure does not require revealing information that others will typically use to judge the data subject's character. Instead, it involves the revealing of physical or mental situations that people tend to consider highly private (e.g., grief, suffering, trauma, injury, nudity, sex, urination, defecation). Because human beings have developed social norms concerning these situations, revealing them to others often results in embarrassment, humiliation, and a loss of self-esteem. The next subcategory, increased accessibility, does not necessary entail direct disclosure of secret information; rather, it refers to the increased ease in which information already available to the public can be accessed by others. Thus, the possible harm is not the result of revealing the information, but a matter of it increasing the chances of experiencing other threats to privacy, such as disclosure, and the corresponding repercussions. Blackmail refers to the threat of disclosure of personal information. It allows the person or persons controlling the information to have power over the individual who is the subject of information. Thus, the actual harm is a loss of power and is a result of the control held by the person making the threat and does not require the actual disclosure or exposure of information. Appropriation involves the use of an individual's – the data subject's – identity to promote the interests of another person. Like other activities discussed previously, it raises privacy concerns because it affects the way the individual is able to present herself to others. Thus far, we have discussed instances in which privacy concerns result from the dissemination and communication of truthful information about individuals. Distortion, however, pertains to the dissemination of false or misleading information. It involves purposeful manipulation of the way an individual is presented and perceived by others. Solove emphasizes that distortion, like disclosure, involves activities that affect the way the individual is perceived by society, and therefore could result in embarrassment, humiliation, and reputational harm. Solove, *A Taxonomy*, *supra* note 55, at 533–52.

D. *Invasion*

Solove's fourth and final group of activities involves invasions into one's private affairs. Solove refers to two forms of invasion: (1) intrusion and (2) decision interference.[132] Intrusion is an invasion of an individual's private life, whereas decision interference pertains to instances in which the state or a third party interferes with an individual's decisions.[133] Under Solove's definition, the utilization of nudging techniques would be categorized as an invasion activity. Although interferences with an individual's decision-making process are commonly criticized as posing a risk to autonomy and liberty (not privacy), Solove claims that such interference is deeply connected to privacy, emphasizing that the individual's protected "zone of privacy" spreads to the "interest in independence in making certain kinds of important decisions."[134]

Referring back to our water consumption example, the corporation's collection and dissemination of household and neighborhood water consumption levels, although not coercive in nature, represents both direct and indirect attempts to interfere with individuals' activities (i.e., water usage) and decision-making processes (e.g., how often to use water).

In sum, although social norm nudges would not directly make private information public, there could be an unexpected and undesirable impact on privacy. Whereas the prevailing analysis in this article focuses mainly on pro-environmental behaviors, the principles discussed throughout this article are applicable to other social norm nudges that share similar characteristics to the ones identified in this article. However, it is important to emphasize that the level and type of privacy harm may vary depending on the specific nudge employed. It is useful to distinguish in this context between two scenarios.

In the first scenario, the social-norm nudge employed is broad, "non-tailored," and descriptive. It includes a general account of the relevant social norm and refers to a very large population. For instance, an online advertisement informing the public of the percentage of youth smoking in Texas,[135] the maximum number of drinks most college students drink at a party,[136] or the percentage of London youth involved in violent incidents.[137]

---

132. *Id.* at 552.

133. Solove, *The Meaning and Value of Privacy*, *supra* note 56, at 78 (claiming that invasion, "unlike the other groupings, need not involve personal information (although in numerous instances, it does)").

134. Solove, *A Taxonomy*, *supra* note 55, at 558 (quoting Whalen v. Roe, 429 U.S. 589, 599–600 (1977)).

135. The nudge used was: "8 out of 10 Texas teens don't smoke."
Social Norms + Foot Humor = Fewer Teen Smokers in Texas? NUDGE (June 14, 2011), http://nudges.org/2011/06/14/new-social-norm-campaign-on-teen-smoking-in-texas/.

136. The nudge used was: "most hoyas have 0-4 drinks when they party."

These nudges, although drawing on social norms, will generally pose a relatively small threat to privacy, if any at all, because it would be almost impossible to attribute the information to a specific individual.

On the other side of the spectrum, there are individually tailored nudges, similar to the one used in the example of A-Water Corporation. As demonstrated throughout this chapter, these nudges pose a threat to privacy at each and every step of the nudging process. Nevertheless, to date, scholars and policymakers have praised the value of social norm nudges without giving the necessary attention to the privacy threats that they pose nor to the task of balancing privacy protection with facilitating social norm nudges.[138]

Although privacy concerns may arise at any stage of the nudging process, the next two chapters center mainly on privacy threats raised during, and as a result of, the communication of information. Of course, privacy concerns that are raised in other stages of the process should not be dismissed. Social norm nudges derive the majority of their power from individuals' need to belong, their desire to fit in with others, and their fear of peers' negative reactions.[139] It is the stage in which the information is no longer under the control of the state or a particular authority, rather it is publicly available. Focusing on this stage allows us to examine the use of differential privacy to address privacy concerns that arise from the use of social norm nudges. In particular, it shows how using algorithms can allow a regulatory agency to utilize social norm nudges while maintaining individual privacy.[140]

1. Using Social Norm Nudges – the Perspective of Privacy-Regulation

Part III argued that there are certain risks to privacy that are inherent in the use of social norm nudges. The degree and scope of these privacy threats may vary depending on the type of the nudge implemented, the scope, form, and type of information communicated, as well as the number of people in-

---

We'll Say It Again. If You're Trying to Curb Binge Drinking, Use Social Norms NUDGE (Mar. 4, 2010), http://nudges.org/2010/03/04/well-say-it-again-if-youre-trying-to-curb-binge-drinking-use-social-norms/.

137. The nudge used was: "99% of young Londoners do not commit serious youth violence." *U.K. Social Norms Campaign to Reduce Youth Violence*, NUDGE (Oct. 24, 2010), http://nudges.org/2010/10/24/social-norms-campaign-to-reduce-youth-violence-in-london.

138. *See, e.g.*, Magali A. Delmas & Neil Lessem, *Saving Power to Conserve Your Reputation? The Effectiveness of Private Versus Public Information,* 67 J. ENVTL. ECON. & MGMT. 353 (2014).

139. Some of these concerns have been covered extensively in the literature addressing information and data security.

140. It is important to note that although this paper focuses on potential harms to privacy that are entangled in water consumption-related nudging, the presented ideas are relevant to broader areas of policy interventions.

cluded in the relevant group. These risks are likely to have far-reaching implications in the big-data era.[141] In particular, this is so because once information is disseminated, it is almost impossible to take it back. However, given the many advantages of social norm nudges, the challenge is to develop a framework which could allow regulators to use this instrument while providing a reasonable level of privacy protection.[142]

Currently, the prominent approach for protecting people's privacy has been based on the notion of *de-identification*. De-identification focuses on reducing identifiability: de-identification policies consider information as "anonymous" once certain types of personally indefinable information ("PII") has been removed from it. The term "anonymous" suggests that this information cannot then be traced back to concrete individuals ("re-identification"). However, in recent years, scholars have demonstrated that de-identification does not guarantee anonymity in various scenarios.[143] These limitations of de-identification techniques are particularly relevant to social norm nudges, which are inherently based on sharing and communicating information.

In this context, our main argument is that regulators should adopt a *differential privacy* framework that should serve as the basis for *privacy-preserving social norm nudges*. The appeal of this framework is that it should enable the regulator to minimize privacy risks and at the same time continue using social norm nudges, estimating, ex-ante, the tradeoff between privacy protection and the efficacy or accuracy of the nudge.

The remainder of this part identifies the challenges that are involved in implementing and relying on de-identification techniques as a means of balancing the utility of data usage with individual privacy interests. The chapter concludes with demonstration of the advantages of differential privacy strategy.

A. *De-Identification Techniques*

Privacy laws and policies are driven by the desire to protect individuals from harm related to collecting, processing, and disseminating information about them. Yet, in various instances policymakers seek to design a way to balance the social benefit associated with using private data against the potential harm to the individual's privacy interests. De-identification is one

---

141. *See* Micah Altman et al., *Practical Approaches to Big Data Privacy Over Time*, 8 INT'L DATA PRIVACY L. 29, 31 (2018).

142. *See* Nehf, *supra* note 70, at 7 (arguing that "in the modern digital world, information privacy should be viewed as a societal value justifying a resolution in the public interest, much like environmental policy and other societal concerns, with less emphasis on individual self-policing and market-based mechanisms").

143. *See, e.g.*, Andrew G. Ferguson, *Big Data Distortions: Exploring the Limits of the ABA LEATPR Standards,* 66 OKLA. L. REV. 831, 871 (2014).

way to strike such a balance.[144] De-identification is a term used to describe a number of techniques "that organizations can use to remove personal information from data that they collect, use, archive and share with other organizations."[145] The aim of these techniques is to maintain the benefits of utilizing the information while reducing the privacy risks associated with its dissemination.

Privacy and data protection laws around the world vary significantly both in scope and in the means of protection that they render. Despite these differences, most privacy laws converge around the principle of de-identification. De-identification is based on one basic assumption: that, in the absence of personally identifiable information, there will be no privacy harm. In other words, the vast majority of data protection and privacy laws determine that private information that has been de-identified falls outside the scope of the law, thereby allowing it to be used or disseminated in ways that would otherwise be considered a violation of the individual's right to privacy. De-identified information can therefore be circulated legally. Consequently, data controllers, such as our A-Water Corporation example, typically take measures to modify their datasets and remove or encrypt PII. By implementing such de-identification techniques, the data controller supposedly protects the data subjects' privacy and, at the same time, maintains the utility of the data.[146]

The initial step in any de-identification technique is to delete or replace (with pseudonyms or arbitrary data) certain personal identifiers such as

144. *See* Chris Achatz & Susan Hubbard, *US vs. EU Guidelines for De-Identification, Anonymization, and Pseudonymization*, 20 J. INTERNET L. 11 (2017); Nissim et al., *supra* note 20, at 699–700; Ira S. Rubinstein et al., *Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches*, 75 U. CHI. L. REV. 261, 266, 268 (2008).

145. Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 WASH. L. REV. 703, 710 (2016) (quoting Simson L. Garfinkel, *De-Identification of Personal Information* 1 (2015), http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf).

146. It is important to note that what constitutes personally identifiable information will depend on the definition adopted by the statute governing the data at issue. For example, the FTC defines personally identifiable information as data which is "reasonably linked to a specific consumer, computer, or other device." *See* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS vii (2012). In contrast, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) defines individually identifiable health information as "information that is a subset of health information, including demographic information collected from an individual." 45 C.F.R. § 160.103 (2017). However, it allows health care providers to share the information if it satisfies the Privacy Rule's de-identification standard by removing eighteen specified individual identifiers. *See* 45 C.F.R. § 164.502(d)(2) (2017); DEP'T OF HEALTH & HUM. SERVS., GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA) PRIVACY RULE 7–8 (2012) .

names, addresses, and social security numbers from the dataset.[147] Hence, a key aspect of every de-identification process is to denote personally identifiable information (PII).[148] Put simply, if a certain piece of information is considered to be personally identifiable information by a specific statute, then it is protected by law and generally cannot be revealed, shared, or disseminated. However, if it is not personally identifiable, then it is not protected by law and can be disseminated without restrictions.[149]

In the United States, for example, the legislative framework for protection of PII is comprised of an amalgam of sectorial or context-specific statutes at the federal and state levels.[150] These laws and regulations are not based on a uniform definition of PII or on a general rule for classifying PII.[151] Thus, the definition of PII varies depending on the underlying law or regulation. For instance, the Health Insurance Portability and Accountability Act ("HIPAA") protects a person's health information. The Act allows health care provider to satisfy the Privacy Rule's de-identification standard by removing eighteen specified individual identifiers.[152] The Family Educational Rights and Privacy Act ("FERPA"), on the other hand, protects personally identifiable information in educational records.[153] However, FERPA

147. *See* Mike Hintze & Khaled El-Emam, *Comparing the Benefits of Pseudonymization and Anonymization under the GDPR*, 2 J. DATA PROTECTION & PRIVACY 146, 146 (2018).

148. Schwartz & Solove (2011), *supra* note 112, at 1814.

149. *Id.* at 1816, 1819–26.

150. *See, e.g.*, Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996); Fair Credit Reporting Act, Pub. L. No. 91-508, 84 Stat. 1114 (1970) (codified at 15 U.S.C. § 1681); Fair and Accurate Credit Transactions Act of 2003, Pub. L. No. 108-159, 117 Stat. 1952 (2003); Children's Online Privacy Protection Act of 1998, Pub. L. No. 105-277, 112 Stat. 2681-728 (1998); Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999). *See also* Palermo, *supra* note 111, at 133. In addition, there are two federal informational privacy law that have a more general scope. "[T]he Privacy Act of 1974, 5 U.S.C. §552a (2006), and the Children's Online Privacy Protection Act of 1998 (COPPA), Pub. L. No. 105-277, 112 Stat. 2581 (1998). The scope of the former is determined according to the context and the identity of the parties: the government and citizens. The scope of the latter is determined according to the age (under 13) of the data subjects." Birnhack, *supra* note 111, at 56.

151. *See* Yuen Yi Chung, *Goodbye PII: Contextual Regulations for Online Behavioral Targeting,* 14 J. HIGH TECH. L. 413, 420–21 (2014); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1733–34 (2010) [hereinafter Ohm, *Broken Promises of Privacy*]; Schwartz & Solove (2011), *supra* note 112, at 1836; Omer Tene, *Privacy Law's Midlife Crisis: A Critical Assessment of the Second Wave of Global Privacy Laws*, 74 OHIO ST. L.J. 1217, 1219 (2013).

152. *See* 45 C.F.R § 164.514(b)(2)(i) (2010). It is important to note that the Privacy Rule Safe Harbor provides two de-identification methods: (1) formal determination by a qualified expert; or (2) removal of eighteen types of identifiers as well as no actual knowledge by the covered entity that the remaining data could be used to identify individuals. *See* DEP'T OF HEALTH & HUM. SERVS., *supra* note 146, at 6.

153. 34 C.F.R. § 99.30 (2019).

authorizes the release of information "after the removal of all personally identifiable information."[154]

Schwartz and Solove have identified three approaches for defining PII in existing statutes: (1) the tautological approach, (2) the non-public approach, and (3) the specific-types approach.[155] Under the tautological approach, PII is generally defined as any information that identifies a person.[156] The advantage of the tautological approach lies in its high flexibility, which allows it to be applied to new technological advancements;[157] however, its tautological structure means that it provides no guidance as to the meaning of PII because it merely states that "PII is PII".[158] The second approach, the non-public approach, attempts to define PII by focusing on what it is not. Schwartz and Solove argue that it is merely a variant of the tautological approach.[159] They emphasize that the non-public approach is problematic since it focuses on whether information is public or not, and thus fails to take into account whether the information in question is in fact identifiable.[160] The third and last approach is based on a list of specific types of information. If the information in question falls into a specified category, then it automatically is considered PII and triggers the statute.[161] Though this approach establishes a clear line with regards to PII, Schwartz and Solove emphasize that it can be rather vague and under-inclusive.[162] In summary, all three approaches are flawed and are likely to lead to ambiguity, uncertainty and/or over- or under-protection of information. This lack of clarity may make it very difficult for policymakers to devise a credible, effective, and practical set of principles for designing privacy-preserving social norm nudges.

Under the European Union's GDPR, on the other hand, information is considered "personal" when it is "relating to an identified or identifiable natural person ('data subject')."[163] Specifically, an identifiable natural person is defined as "one who can be identified, directly or indirectly, in par-

---

154. 34 C.F.R. § 99.31(b)(1) (2019).

155. *See* Schwartz & Solove (2011), *supra* note 112, at 1828.

156. *Id*. at 1829 (using The Video Privacy Protection Act (VPPA) as an example for this model).

157. *Id*.

158. *Id*.

159. *Id*. at 1830 (arguing that "[t]he Gramm-Leach-Bliley Act (GLBA) epitomizes one aspect of this approach by defining 'personally identifiable financial information' as 'nonpublic personal information').

160. *Id*.

161. *Id*. at 1831 (illustrating different variations of this approach by discussing Massachusetts's breach notification statute of 2007, California's Song-Beverly Credit Card Act of 1971 and the federal Children's Online Privacy Protection Act (COPPA) of 1998).

162. *Id.* at 1832. *See also* Ohm, *Broken Promises of Privacy*, *supra* note 151, at 1742.

163. Commission Regulation 2016/679, *supra* note 107, at art. 4(1).

ticular 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."[164] Recognizing both direct and indirect means of identification enables the GDPR to accommodate technological advancements and takes into account the expanding abilities of re-identification techniques.[165] This is something that the earlier approaches are not quite equipped to do. Another interesting feature of the GDPR is its continuum of de-identification paradigms. On the one hand, truly anonymized information is considered non-identifiable, and therefore falls outside the scope of the GDPR.[166] Pseudonymized information, on the other hand, may be re-linked to an individual using additional information or a key (e.g., code or algorithm).[167] Therefore, it remains personal data and thus, is subject to the GDPR.[168]

However, in accordance with the GDPR, the obligations to which the data controller must adhere are relaxed when it comes to the use of "pseu-

---

164. *Id.*

165. In that respect, Recital 26 clearly states that "[t]o 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 of 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." *Id.* at recital 26.

166. "The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes." *Id.*; *see also* Data Protection Working Party, *Opinion 05/2014 on Anonymisation Techniques*, at 5 (Apr. 10, 2014), https://www.pdpjournals.com/docs/88197.pdf. Although, as will be further discussed, numerous scholars have argued that re-identification of anonymized information is possible.

167. The GDPR defines pseudonymization as "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person." Commission Regulation 2016/679, *supra* note 107, at art. 4(5); *see also* Hintze & El-Emam, *supra* note 147, at 146.

168. Commission Regulation 2016/679, *supra* note 107, at recital 26; *see also* Hintze & El-Emam, supra note 147, at 146; Kimberly A. Houser & W. Gregory Voss, *GDPR: The End of Google and Facebook or a New Paradigm in Data Privacy?*, 25 RICH. J.L. & TECH. 1, 62–63 (2018); Sophie Stalla-Bourdillon & Alison Knight, *Anonymous Data v. Personal Data – A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data*, 34 WIS. INT'L L.J. 284, 287 (2016); Laura Tarhonen, *Pseudonymisation of Personal Data According to the General Data Protection Regulation*, REFEREE-ARTIKKELI 10, 11–16, https://www.edilex.fi/artikkelit/18073.pdf (last visited Oct. 10, 2019).

donymization of personal data."[169] In sum, under the GDPR the term "personal information" encompasses a broad range of information types. Consequently, some types of information could constitute personal data under European law but not under the United States statutes. Still, it is not possible to draw a firm line between personally identifiable information and non-identifiable information, particularly because the distinction between personally identifiable information and non-identifiable information is, in fact, highly context-dependent. [170] The same is true for information that has gone through a de-identification process.

For many years, de-identification techniques were considered the "silver bullet," which allowed organizations to collect, process, reuse, and disseminate information while, at the same time, preserving the individual's privacy.[171] There is however, ample evidence showing that de-identification via anonymization or pseudonymization of personal data is not always able to prevent the exposure of individualized data. Consider, for example, Netflix's failed attempt to use anonymization or de-identification techniques as a means to protect subscribers' data, which eventually led to a lawsuit against the company for violation of the Federal Video Privacy Protection Act along with several other California laws.[172] Netflix launched a contest that invited the public to compete in developing a better movie-recommendation algorithm than the one the company was using.[173] In order to support the developers' and researchers' efforts, the company released 10 million movie rankings by 500,000 customers. The data was anonymized by removing personal details and replacing names with random numbers to protect the privacy of the recommenders. Just a few weeks after the contest had begun, two researchers from the University of Texas announced that they were able to identify several Netflix users by comparing their so-called "anonymous" reviews in the Netflix dataset to reviews posted on the Internet Movie Database website.[174]

---

169. Elizabeth A. Brasher, *Addressing the Failure of Anonymization: Guidance from the European Union's General Data Protection Regulation*, 2018 COLUM. BUS. L. REV. 209, 214 (2018).

170. *See* supra note 151.

171. Ohm, *Broken Promises of Privacy*, *supra* note 151, at 1736 ("Legislatures have deployed a perfect, silver bullet solution—anonymization—that has absolved them of the need to engage in overt balancing. Anonymization liberated lawmakers by letting them gloss over the measuring and weighing of countervailing values like security, innovation, and the free flow of information"); *see also* OMER TENE & CHRISTOPHER WOLF, THE DEFINITION OF PERSONA DATA: SEEING THE COMPLETE SPECTRUM 4 (2013).

172. Complaint at 1, Doe v. Netflix, Inc., No. 5:09-cv-05903 (N.D. Cal. Filed Dec. 17, 2009); *see also* Nissim et al., *supra* note 20, at 700–02 (discussing several other examples, exposing the vulnerabilities of anonymization techniques across many types of data).

173. NETFLIX PRIZE, https://www.netflixprize.com/ (last visited Feb. 5, 2019).

174. Bruce Schneier, *Why 'Anonymous' Data Sometimes Isn't*, WIRED (Dec. 12, 2007, 09:00 PM), https://www.wired.com/2007/12/why-anonymous-data-sometimes-isnt/; Ryan

This Netflix example is by no means an isolated incident. A recent study conducted by a group of researchers at Carnegie Mellon University demonstrated that an individual's social security number can be predicted based on various datasets, including public datasets and social networks.[175] Others have argued that one could re-identify 90% of consumers based on the credit transactions of 1.1 million users.[176] Several studies published in the past decade have pointed out the inherent difficulties of relying on anonymization techniques as a way to protect people's privacy.[177] Paul Ohm argues that anonymization techniques have become ineffective in balancing the protection of individuals' privacy with the utility of data.[178] According to Ohm, policymakers and technologists adopted the idea that de-identification techniques "could robustly protect people's privacy by making small changes to their data"; Ohm argues that this assumption is flawed.[179] Ohm contends that existing technology facilitates the association of anonymized data with identifiable information to the point that it is possible to override traditional anonymization methods.[180] He further argues that due to the increased accessibility of public, commercial, and other datasets, re-identification has become highly feasible.[181] Others have noted that, as the technological capabilities for inferring information about individuals from aggregated datasets is advancing, the risk to privacy is expected to rise.[182]

In conclusion, re-anonymization techniques enable third parties to link data from other sources to identity the data subject, even when the underly-

---

Single, *Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims*, WIRED (Dec. 17, 2009, 04:29 PM), https://www.wired.com/2009/12/netflix-privacy-lawsuit/.

175. Alessandro Acquisti & Ralph Gross, *Predicting Social Security Numbers from Public Data*, 106 PROC. NAT'L ACAD. SCI. 10975, 10975 (2009).

176. Yves-Alexandre de Montjoye et al., *Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata*, 347 SCIENCE 536, 537 (2015); *see also* Yves-Alexandre de Montjoye & Alex "Sandy" Petland, *Response to Comment on "Unique in the Shopping Mall: On the Reidentifiability of Credit Card Data"*, 351 SCIENCE 1274-b (2016). *But see* David Sanchez et al., *Comment on "Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata"*, 351 SCIENCE 1274-a (2016) (claiming that "data owners and subjects can be reassured that sound anonymization methodologies exist to produce useful anonymized data that can be safely shared for research").

177. *See, e.g.,* Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 716 (2010); Chin & Klinefelter, *supra* note 20, at 1426–27; Khaled El Emam et al., *A Systematic Review of Re-Identification Attacks on Health Data*, PLOS ONE (Dec. 2011). *But see* Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J.L. & TECH. 1, 8–10, 36–42, 48–50 (2011) (arguing that the re-identification risk has been extremely overstated by legal scholars and that imposing "k-anonymity" will sufficiently protect the privacy interest in datasets).

178. Ohm, *Broken Promises of Privacy*, *supra* note 151, at 1704–44.

179. *Id*. at 1706–07.

180. *Id*. at 1716–30.

181. *Id*. at 1723–30.

182. Nissim and el., *supra* note 20, at 722–23.

ing data would not be classified as "personally identifiable" under existing privacy laws.[183] Furthermore, even if the underlying information cannot, theoretically, directly identify the individual, when multiple data sources are combined, re-identification may be possible. These limitations of de-identification techniques are particularly relevant to social norm nudges, which are inherently based on sharing and communicating information.[184] Even in situations where the information has supposedly been anonymized, it would be nearly impossible for the nudging entity to guarantee that re-identification is not possible.

To illustrate the problematic nature of re-identification, suppose that recently, due to reports demonstrating that the level of measles vaccination for children has declined for the third successive year,[185] the government decided to use social norm nudges in order to increase vaccination rates. For example, suppose that every month parents would be given the latest data on vaccination rates, both at the country level (80-90% of US children are fully vaccinated by age five) and at the level of their own local community (which could be further reduced to school level).[186] At first, the disclosure of such aggregated information may seem harmless in terms of the risk it poses to individuals' privacy. But suppose that a few months before the last disclosure, David and Rose decide to move from one neighborhood to another. Imagine that following their move into the new neighborhood, there is a decline in the vaccination rates at the local school. The parents who are informed of this decline may start to worry about the way it could affect their children and try to identify the culprit. One way to do so would be to investigate who, in the community, has recently enrolled their children in the lo-

183. *See* Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, 2008 IEEE SYMP. ON SECURITY & PRIVACY 111, 119, 121.

184. Nevertheless, it is interesting to note that up until a few years ago, the California Public Record Act ("CPRA"), which generally mandates that public records be open to inspection by any person, recognizes that the government possesses various types of sensitive information about people. And so, this act specifically exempts numerous records from disclosure, including "the names, home addresses, telephone numbers, credit histories, and usage data of local utility agency customers are also exempt from general disclosure under the CPRA." Scott A. Baxter, *Informational Privacy and the California Public Records Act*, 30 MCGEORGE L. REV. 778, 787 (1999); *see also* CAL. GOV'T CODE § 6254 (West 1998); Rubinstein & Hartzog, *supra* note 145, at 754.

185. For general discussion and reports regarding measles vaccine rates, *see* Einav Keet, *Several US Cities Vulnerable to Measles Outbreaks Due to Declining Vaccination Rates*, CONTAGION LIVE (June 19, 2018), https://www.contagionlive.com/news/several-us-cities-vulnerable-to-measles-outbreaks-due-to-declining-vaccination-rates; Jacqui Wise, *Child Vaccination Rates Drop in England as MMR Uptake Falls for Fourth Year*, BMJ (Sept. 18, 2018), https://www.bmj.com/content/362/bmj.k3967;.

186. *See generally* Meng Li & Gretchen Chapman, *Nudge to Health: Harnessing Decision Research to Promote Health Behaviour*, 7 SOC. & PERSONALITY PSYCHOL. COMPASS 187 (2013); Mark C. Navin, *The Ethics of Vaccination Nudges in Pediatric Practice*, 29 HEC F. 43, 45 (2017).

cal school. More than that, assuming that the majority of parents in the relevant neighborhood know each other, it would be fairly simple for any interested neighbor to deduce who, among her neighbors, did not vaccinate their children. The social ramifications of such identification could be severe, subjecting the family who has not adhered to the "social norm" to various social sanctions.[187] As this hypothetical example illustrates, the anonymization of data does not necessarily preclude a third party from combining external sources of information to infer a particular data subject's identity.[188]

The key point is that beyond obtaining explicit consent, policymakers should adopt a more effective framework to balance the utility of social norm nudges against the potential threat to privacy. This article suggests a framework which is not as dependent upon the idea of identifiable information and is also less vulnerable to privacy risks emanating from combining varied external resources. Particularly, the use of "differential privacy" algorithms is an effective method for mitigating the risks to privacy via re-identification.[189]

B. *Differential Privacy as a Tool for Social Norm Nudges Design*

Differential privacy is a mathematical framework for guaranteeing privacy protection when analyzing data. To illustrate the concept of differential privacy, consider again the A-Water Corporation example. Assume that the corporation would like to develop a social-norm nudge based on data on household water consumption for a specific community consisting of sixteen households during the month of March 2018. The company first collects the following data:

---

187. For a discussion of vaccinations as a social dilemma and the effect of social nudge interventions that aim to increase individuals' motivation to act in the group's interest, *see* Lars Korn et al., *Social Nudging: The Effect of Social Feedback Interventions on Vaccine Uptake*, 37 HEALTH PSYCHOL. 1045 (2018).

188. The vulnerability of de-identification techniques has been subject to discussion for the past few years. And so, a growing number of legal scholars have called to amend privacy and data protection regulations and relax our dependence on de-identification techniques such as anonymization and pseudonymization. For example, Ohm, who is one of the biggest critics of current de-identification techniques, has recommended several approaches to mitigate privacy concerns, including disclosing de-identified data only to trusted researchers while, at the same time, establishing contractual and/or regulatory restrictions on the uses of such data. *See* Ohm *Broken Promises of Privacy*, *supra* note 151, at 1764–69. Paul Schwartz and Daniel Solove oppose suggestions that would require reforming the law and the practice. They do, however, advocate for modifying and refining the concept of personally identifying information. *See* Schwartz & Solove (2011), *supra* note 112, at 1883–86; Schwartz & Solove (2014), *supra* note 112, at 877. Others have recently proposed harnessing the advantages of differential privacy as a means to lessen the threat of re-identification. *See, e.g*., Chin & Klinefelter, *supra* note 20, at 1427–28.

189. Arvind Narayanan, Joanna Huey & Edward W. Felten, *A Precautionary Approach to Big Data Privacy* 11–12 (2015), http://randomwalker.info/publications/precautionary.pdf.

TABLE 1: WATER CONSUMPTION IN MARCH 2018 (HYPOTHETICAL)

Household No.	Water Consumption
1	1
2	3.4
3	5.3
4	6.1
5	6.1
6	6.3
7	6.6
8	7
9	7.2
10	7.3
11	7.8
12	8
13	8
14	8.9
15	9.1
16	11

As summarized in Table 1, household number one has consumed 1 cubic meter of water in March 2018, household two consumed 3.4 cubic meters, and so on. After gathering the data, the company is able to process it and determine, e.g., that the social norm is the median consumption (which equals 7.1 [(7+7.2)/2]). Suppose that customers receive information about their own household water consumption, as well as the value of the "*social norm*." The danger here is that this information (possibly when combined with external data sources) might leak information about other consumers. For example, suppose that all members of the community, except for the residents of household number 9, decide to combine the information they received from A-Water Corporation in order to learn about the water consumption of household number 9. By combining their own water consumption data with the median consumption, the community easily learns that household number 9 consumed 7.2 cubic meters of water, because this is the only possible value that would be consistent with a median of 7.1.

The idea of differential privacy is to add uncertainty to the computation of the social norm, so that even when combining it with the data of all other households, every value for the consumption of household number 9 is still possible. To achieve this goal, instead of computing the social norm as the exact median, the A-Water Corporation computes the social norm by *randomly picking an "approximate median*." The requirement of differential privacy is that any value of the approximate median (say 6.9) is roughly as equally likely to result from the original data (specified in Table 1) or from an alternative data set in which the information of household number 9 is

replaced with an arbitrarily different value (or removed altogether). This means that even if all other households combine the information they received from the A-Water Corporation (their own consumption and the approximate median, say 6.9), then they still cannot deduce essentially anything about the consumption of household number 9. This is because an approximate median of 6.9 would be almost as equally likely, no matter the water consumption of household number 9.

It is important to point out here that differential privacy is not a property of the approximate median itself, but rather a property of the *randomized process* (or *the algorithm*) that computed the approximate median.

In more general terms, the differentially private algorithm guarantees that a third party cannot learn any more from the information disseminated by A-Water than he could have learned had the computation been done without the data of any single individual.[190] In our scenario, a third party might be able to learn that A-Waters' customers are residents of a specific city, say Nashville, but this piece of information could have been deduced even if a *specific* household's members (i.e., David and Rose) were to opt-out from the database entirely.

What the third party will not be able to determine, however, are properties pertaining to individuals, for example that David and Rose are costumers of A-Water Corporation.

Certainly, the differential privacy framework will not solve all privacy concerns that may arise with regard to social norm nudges, nor will it protect the individual against unauthorized collection or processing of information done by the data controller (i.e., the nudging entity). Additionally, it cannot prevent security breaches. The differential privacy framework, however, will provide an increased level of privacy preservation in comparison to leading methods of de-identification.[191] It can thus provide regulators – whether private or public – with a workable strategy for using social norm nudges while protecting privacy.[192]

Moreover, it is important to note that differential privacy comes at a cost to accuracy.[193] As previously illustrated, for the company to provide David, Rose, and all of the other A-Water customers with protection, it will

---

190. Kobbi Nissim et al. define computation as the "mechanizable procedure for producing an output given some input data." Nissim et al., *supra* note 20, at 696; *see also* Wood et al., *supra* note 14, at 223.

191. TENE & WOLF, *supra* note 171, at 6.

192. It is important to note that differential privacy is a technical definition or a criterion. This means that a variety of algorithms could satisfy it. Nevertheless, for a proposed solution to satisfy the standard of differential privacy, it must adhere to its mathematical definition. *See* Nissim et al., *supra* note 20, at 714.

193. SIMON L. GARFINKEL, NAT'L INST. OF STANDARDS & TECH., U.S. DEP'T OF COMMERCE, NISTR 8035, DE-IDENTIFICATION OF PERSONAL INFORMATION 8 (2015).

need to add statistical noise to obscure personal information.[194] In our case, if A-Water Corporation is interested in implementing a privacy-preserving social norm nudge, it would not be able to use the *exact* median of consumption, but only a randomized approximation of it. The approximation would vary depending on the level of privacy the company wishes to achieve and the level of noise added. More privacy means adding more noise which, in turn, translates to less accurate computations.[195] Stated differently, the more privacy the company will attempt to attain, the bigger the error the approximation will display.

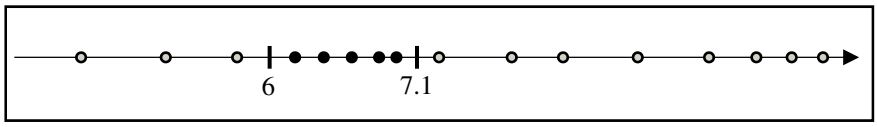
This tradeoff between privacy and utility can be adjusted using a privacy-loss parameter, traditionally denoted as ϵ. The lower the value chosen for •, the more privacy the company can accomplish. This will also create a bigger error with regards to the approximate median computed.[196] For example, in the case of the abovementioned households, if the A-Water Corporation will use 6 m^3 as the median (instead of the true 7.1 m^3), then five households within the relevant group would receive an inaccurate social norm nudge. In other words, five households would incorrectly determine whether their water level usage was above or below the social-norm (i.e., true median). Therefore, we define the *error* of an approximate median (as compared to the true median) as the number of households that would receive an inaccurate social norm nudge.

194. Henry Kenyon, *DARPA Research Seeks New Ways to Keep Private Data Anonymous*, CQ ROLL CALL (Nov. 20, 2018), 2018 WL 6057838; Scripa Els, *supra* note 20, at 220; Wood et al., *supra* note 14, at 232.

195. *See* Cynthia Dwork & Moni Naor, *On the Difficulties of Disclosure Prevention in Statistical Databases or the Case for Differential Privacy*, 2 J. PRIVACY & CONFIDENTIALITY 93, 103 (2010).

196. "A parameter quantified and limits the extent of the deviation between the privacy ideal and real world scenarios. . . . The parameter ϵ measures the effect of each individual's information on the output of the analysis. It can also be viewed as a measure of the additional privacy risk an individual would incur beyond the risk incurred in the privacy-ideal scenario." *See* Nissim et al., *supra* note 20, at 715.

FIGURE 1: AN ILLUSTRATION FOR THE ERROR OF AN APPROXIMATE MEDIAN.



The sixteen dots represent the water consumption of sixteen households, and their (exact) median is 7.1. An approximate median of 6 has error 5 since there are 5 points (the black points) that are bigger than this approximate median even though they are smaller than the true median.

Social norm nudges create something similar to "peer pressure" which is meant to induce "desirable" behavior. If a large number of people will be inaccurately "nudged," it will undoubtedly affect the overall efficacy of the nudge – as a means to induce water conservation.[197] This effect, however, is less pronounced in larger groups, or groups where the distribution is less dense around the median, because noise wouldn't affect the median approximation as much with more data points available.

As an example, consider real data obtained from a city with 10,013 households.[198] The "real" median based on the water consumption in the city is 22.3 m^3 (cubic meter).

If the corporation chooses to implement a differentially private algorithm, it will need to determine ex-ante what level of privacy it aims to achieve and, consequently, what privacy-loss parameter to deploy. Recall that the error will depend, to a large degree, on the company's choice for the privacy-loss parameter ϵ.[199] This example demonstrates the effect of privacy-loss parameters on the accuracy of nudging in a dataset containing 10,013 households, using two measures of the privacy-loss parameter epsilon: ϵ_1=0.05; ϵ_2=1, where ϵ_1=0.05 represents a high level of privacy and ϵ_2=1 denotes a lower level of privacy.

197. *See* Wood et al., *supra* note 14, at 204.

198. We received this data from an Israeli water corporation for the purpose of this paper.

199. Wood et al. recommend a range of 0.01-1 for •. Wood et al., *supra* note 14, at 252.

TABLE 2: EXAMPLES USING TWO VARIANTS OF PRIVACY-LOSS PARAMETERS:[200]

$\epsilon_2 = 1$	$\epsilon_1 = 0.05$	
0	31	Households receiving false nudge
0 m^3	0.03 m^3	Gap between false median and real one
0%	0.31%	Percentage of households with false nudge

As displayed in Table 2, even when using a relatively high level of privacy protection (i.e., ϵ=0.05), the occurrence of a false nudge arises only for 0.31% of the customers. Meaning that out of the 10,013 households, only 31 received a false nudge.

Notwithstanding the above, it is important to note that two other parameters could also affect the aforementioned error. First, due to the fact that the computation is randomized, different execution may result in different outcomes, and it could be, for example, that with some very small probability the error is large. This un-confidence level, referred to as the failure probability, is traditionally denoted as β.[201] To better understand this concept, consider the following thought experiment.

When we flip a coin, we have a 0% chance of flipping nothing, a 50% chance of flipping heads, a 50% chance of flipping tails, and a 100% chance of flipping something, heads or tails. This is all very intuitive.

Let us suppose that we flip a coin 1000 times. Even though the expected number of heads is 500, the probability of actually getting exactly 500 heads is quite small. Nevertheless, we can be "pretty sure" that the number of heads will be between 450 and 550. Actually, this fails to be the case with probability at most 0.02. We can be even more confident that the number of heads will be between 440 and 560, as this fails to be the case with probability at most 0.0015. The point here is that by making a "looser" statement, we can be more confident that it holds.

---

200. Due to the fact that the computation is noisy, different executions may result in different outcomes, and the values in the table reflect the outcome we obtained after a single execution (with the appropriate parameters). As will be made clear later, our algorithms are accompanied by a theoretical analysis providing worst case bounds on the error. The values in the table are well within these guaranteed bounds.

201. It is important to emphasize that in the computer science literature, this parameter is called the "confidence parameter" even though bigger values of it correspond to less confidence. To accentuate this, we refer to β as the "un-confidence parameter." Similarly, the parameter ϵ is commonly referred to as the "privacy parameter." However, throughout this paper it is titled the "privacy-loss parameter" to emphasize the fact that smaller values of it correspond to more privacy.

Returning to our example, smaller values for β mean that our bounds hold with higher probability, but result in looser error bounds. Intuitively, we can think of β as 1/1000000. Again, this means that with small probability (at most 1/1000000) the actual error incurred might be bigger than the bounds stated next. Such a small probability of "failure" is standard. In addition to the privacy-loss parameter ϵ and the failure probability β, at least in some cases the error also depends on the way in which we discretize the resulting approximation of the median. To better illustrate this concept, recall that, in the above example, the true median for the relevant group of customers is 7.1. Now consider the following two scenarios. In the first scenario, the city water company wants an approximate median which is a natural number between 1 and 15. In the second scenario, they want an approximate median in that range which is a multiplication of 0.1 (i.e., a number from 0.1, 0.2, 0.3, 1, 1.1, 1.2, . . ., 14.9, 15). That is, in the first scenario they aim to choose the approximate median from one of fifteen possible values and, in the second scenario, from 150 possible values. On the one hand, our discretization itself might incorporate errors into our estimation. To illustrate this idea, in the first scenario we can never identify a median with an error of zero (since a median with an error of zero cannot be an integer). On the other hand, it turns out that privately choosing an approximate median becomes harder (i.e., requires more noise) when our discretization contains more points. The reason for this increased difficulty is that an approximate median with higher precision contains (potentially, at least) more information about the data; hence, it requires more noise to be released privately. For example, deciding which of two numbers (say, 3 or 6) is a better choice for the median reveals very little information about the data (this is basically a single yes/no question), while identifying an approximate median with high precision necessitates much more information.[202]

The lesson from this example is that by discretizing the set of possible outputs (approximations for the median), the water company will inevitably incur one of these two types of errors. The first type, which we will call the discretization error, simply follows from the fact that A-Water's discretization might fail to contain a good choice for the median. This type of error becomes smaller if the company decides to use more points in the discretization. The second type of error follows from the fact that, to ensure privacy, the company ought to add more noise if the discretization contains more points. This type of error becomes larger if A-Water uses more points in the discretization. To further illustrate this concept, we will provide an example in which the number of points in the discretization will be denoted as X and X will equal 1000. Experiments with real data show that this number is a

202. *See* Mark Bun et al., *Differentially Private Release and Learning of Threshold Functions*, IEEE 56TH ANN. SYMP. ON FOUNDS. COMPUTER SCI. 634, 634 (2015).

reasonable choice in order to ensure that the discretization error remains small.

The exponential mechanism provides a good baseline for the problem of privately estimating the median.[203] This mechanism is capable of identifying an approximate median with an error of, at most, $(2/\epsilon)\ln(X/\beta)$ where ϵ is our privacy parameter, X is the number of elements in our discretization, and β is the failure probability. To put things in context, assume that we choose the failure probability to be β = 1/1000000, and the privacy parameter to be 1 (comparable, and typically much better than current industry implementations of differential privacy). In addition, assume that our discretization satisfies the condition X = 1000. With these choices, the exponential mechanism guarantees an error of no more than 42. Again, this means that, except with probability at most 1/1000000, at most 42 consumers will "suffer" from an inaccurate comparison to the approximate median. Assuming that the city population is much bigger than 42, say ≈ 100,000, then this error rate seems reasonable.[204]

There are several advanced constructions for differentially private medians that, under the above conditions, allow for privately identifying an approximate median with an error of at most $(C/\epsilon)\cdot 2^{\log^* X}\cdot\log(1/\beta)$ for some large constant C. This means that (asymptotically) the error needs to grow much slower with the size of the discretization X. However, in the currently known constructions, the constant C is relatively big, and the gains of having a weaker dependency of the error in X only "kicks in" when X is huge. For our setting, in which X ≤ 1000 suffices, these advanced constructions do not achieve better guarantees than the exponential mechanism.

The discussion thus far assumes that A-Water Corporation uses all of its users' data in order to identify an estimation for the median. Moreover, it assumes that the computed estimation would be reported to all of the customers in their monthly bills. However, for nudging purposes, it might be sufficient to release much less information which, consequently, increases accuracy. Specifically, in our example, the value of the (exact or approximate) median water consumption might not be important for the customer and thus, not a determinant factor for the effectiveness of the nudge. It is possible that simply indicating to Rose and David what their water consumption was and whether it was below or above the median (without disclosing the actual median value/estimation) will suffice. In this scenario, the error does not need to have any dependency on X, using a framework called joint differential privacy. Specifically, for privacy-loss and failure parameters $\epsilon = 1$ and β = 1/1000000, the corporation will be able to guarantee an

203. *See* Wood et al., *supra* note 14, at 244–46.

204. *See generally* Amos Beimel et al., *Private Learning and Sanitization: Pure vs. Approximate Differential Privacy,* 12 THEORY OF COMPUTING 1 (2016).

error of at most 14, a significant improvement over the error of the exponential mechanism.

TABLE 3: THREE EXAMPLES FOR DISCRETIZATION:

X-whole numbers	X-two points after decimal	No median published	
90	0	0	$\epsilon = 1$
90	31	10	$\epsilon = 0.05$

The errors depicted in Table 3 are for a single median calculation. Ideally, we would want the corporation to protect customers' privacy not only during each billing cycle, but for a much longer time-span (e.g., 10 years, during which there are 60 billing cycles). Repeated calculations could potentially reveal more information about the data subjects over time and therefore, more noise ought to be added to each of the computations to maintain the same level of overall privacy.[205] However, this will unavoidably cause the corporation's errors in median approximations to increase. Thus, depending on the specific context, policy makers might require the nudging entity – A-Water Corporation – to protect against several billing cycles (referred to as user-level differential privacy in the literature)[206] or only against a single billing cycle (referred to as event-level differential privacy).[207]

As previously stated, privacy guarantees of differential privacy are controlled by the parameter ϵ. Smaller ϵ denotes more privacy and, in turn, more error. Traditionally, researchers in the theoretical literature had thought about the privacy parameter ϵ as a small constant, say $\epsilon = 0.1$. However, current industrial applications of differential privacy (e.g., Apple) are using values such as $\epsilon = 1$, or 2, or even 8. In addition, these industrial applications typically use event-level differential privacy (i.e., protect against a single computation, which in our case corresponds to protecting against a single billing cycle).[208] Thus, when urging policymakers to make differential privacy the required standard for achieving privacy-preserving social norm nudges, it would be reasonable to follow the industry-level guarantees and deploy event-level differential privacy of $\epsilon = 1$, at least in the first step.

---

205. *See* Wu, *supra* note 20, at 1137–40.

206. This is also known as user-level privacy.

207. This method is known as event-level privacy.

208. *E.g.*, Differential Privacy Team, *Learning with Privacy at Scale,* APPLE MACHINE LEARNING J. (Dec. 6, 2017), https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

As Table 4 indicates, using $\epsilon = 1$ would allow the corporation to use the social norm nudge for a longer time-span (e.g., one year). Alternatively, if the corporation will choose not to disclose the exact median (or even an approximation of it), but to say that "*at least half of the households consume less than you*" it would even allow our hypothetical corporation to use it for 10 years. In the long-run, the tradeoff between privacy and social norm nudges is clear.

TABLE 4: NUMBER OF HOUSEHOLDS WITH FALSE NUDGE IN THE LONG RUN:

	One bill	One year	10 years
$\epsilon = 1$; no median published	0	0	51
$\epsilon = 1$; median published with two points after decimal	0	0	230
$\epsilon = 0.05$; no median published	0	52	1908
$\epsilon = 0.05$; median published with two points after decimal	31	60	4541

Again, the errors mentioned above are for a single median calculation. Adopting this approach to preserving privacy in the context of social norm nudges yields a number of practical benefits. First, adhering to this mathematical definition of privacy assures stronger protection against a wider range of potential privacy threats, in comparison to existing standards, even in circumstances in which it is difficult for the nudging entity to anticipate the type of attack on privacy.[209] Furthermore, using differential privacy provides guarantees even with respect to future information attacks not known or anticipated at the time of dissemination of the social norm nudge.[210] This increased privacy protection enables a broader utilization of social norm nudges that could substantially increase the effectiveness of the nudges. Last, but not least, differential privacy enables the nudging entity to control the tradeoff between data accuracy and the level of privacy preserved and, therefore, maintain the flexibility to accommodate different forms of nudges.

Thus far, this article highlighted the advantages of differential privacy framework over the widely-used de-identification techniques. However, for policymakers to require all nudging entities to adopt a *differential privacy* framework, one must also consider the existing legislation. As pointed out by Kobbi Nissim and colleagues, there seems to be a gap between the legal and mathematical approach to privacy.[211] This is mainly because data pro-

209. Wood et al., *supra* note 14, at 235.
210. *Id.*
211. Nissim et al., *supra* note 20, at 691.

tection and privacy laws are generally context-specific, flexible, and open to interpretation. By contrast, differential privacy utilizes a mathematically rigorous definition. In this sense, it is plausible that this method of preserving privacy will pose a challenge to policymakers aiming to implement it as a binding standard. This challenge, or gap, between the two disciplines should not discourage policymakers from utilizing the advantages of differential privacy.[212] In fact, in a recent article, Nissim and colleagues were able to demonstrate that the differential privacy framework could satisfy relevant legal requirements for privacy protection.[213] By extracting formal mathematical requirements based on the legal standard set by the Family Educational Rights and Privacy Act ("FERPA"),[214] they were able to prove that a differential privacy framework satisfies the mathematical solution derived by FERPA.[215]

This is just one example of how the legal and mathematical approaches can coincide. Of course, the variety of data protection and privacy laws, as well as the wide range of areas in which social norms nudges could be implemented, make it difficult to produce a conclusive argument regarding the ability of differential privacy to satisfy the relevant legal requirements for privacy protection. Nevertheless, following Nissim and colleagues, we can assert that the differential privacy framework could protect privacy while satisfying the legal requirement set by legislatures in a wide range of scenarios. Differential privacy can satisfy the legal requirement that disseminated information will not reveal any specific PII or enable a third party to infer any attributes that are specific to a concrete individual.[216]

V. CONCLUSION

Social norm nudges are an important part of the regulatory toolbox. They are non-coercive, inexpensive, relatively easy to implement, and effective. Consequently, the advent of social nudges represents a new way for policymakers to promote a range of welfare-enhancing behaviors. At the same time, the use of social norm nudges presents a serious risk to privacy.

---

212. It is important to note that the differential privacy framework has already been implemented in various real-world scenarios. *See, e.g.*, U.S. Census Bureau, Ctr. For Econ. Studies, ONTHEMAP, https://onthemap.ces.census.gov (last visited Sept. 14, 2019); Machanavajjhala et al., *Privacy: Theory Meets Practice on the Map*, IEEE 24TH ANN. INT'L CONF. ON DATA ENGINEERING 277, 283–85 (2008).

213. Nissim et al., *supra* note 20, at 696.

214. Family Education Rights and Privacy, 73 Fed. Reg. 74,853 (Dec. 9, 2008).

215. Nissim et al., *supra* note 20, at 734–763.

216. *See* Alexandra Wood, *Bridging Privacy Definitions: Differential Privacy and Concepts from Privacy Law & Policy* (Oct. 23, 2017), http://archive.dimacs.rutgers.edu/Workshops/Barriers/Slides/Wood.pdf.

Despite the importance of this threat both the literature on nudges and the regulatory establishment have given little attention to this problem. We have sought to fill this void by making two novel contributions. The first contribution is analytic: demonstrating that the very nature of social norm nudges can pose serious threats to individuals' privacy. The second contribution is policy-oriented: arguing that the strategy of *differential privacy* can be used to balance between the utility of social norms nudges and the risk of harming individuals' privacy.

Although privacy and data protection laws around the world tend to focus on de-identification techniques, scholars have repeatedly shown that these methods are not always able to achieve the desired outcome. Thus, to allay the privacy concerns described throughout this article, we suggest that policymakers should utilize a differential privacy framework to increase the likelihood of privacy protection while, at the same time, taking advantage of the regulatory potential of social norm nudges.

Differential privacy is a mathematical framework for guaranteeing privacy protection when analyzing data. It guarantees that a third party will not be able to learn any more from the information disseminated by analysis than he could have learned had the analysis been done without the data of any single individual, and considered an achievable standard "to replace or supplement fragile anonymization approaches."[217]

Undoubtedly, differential privacy does not solve all privacy concerns that may arise with regard to social norm nudges, nor will it protect the individual against unauthorized collection or processing of information or prevent security breaches. Furthermore, using differential privacy algorithm involves some tradeoffs between accuracy and privacy. Specifically, more "noise" yields better privacy but also less accurate computation (and, hence, poorer social norm nudges).

Nevertheless, throughout this article we have shown that by using an algorithm based on differential privacy, the nudging entity can create a variant of the social norm nudge that is both effective and provides sufficient protection of private data (based on the sensitivity of the data). Hence, the differential privacy framework provides an increased level of privacy preservation in comparison to leading methods of de-identification.

Our choice of focusing on water consumption (where privacy concerns are relatively mild) was motivated, in part, by the lack of literature on the subject. It was also driven by our belief that facilitating water conservation, given the looming risks of climate change, is one of the greatest challenges of global society. Improving our capacity to use social norm nudges can thus have significant value. Further, we believe that our argument and poli-

217. Chin & Klinefelter, *supra* note 20, at 1423.

cy proposals can be applied in other domains, such as health or financial data, where the privacy risks are more salient.

By emphasizing the potential contribution of differential privacy to achieving a better balance between the use of data-sensitive social norm nudges with privacy protection we hope to persuade regulators and other researchers to further experiment with this idea.