WHAT WE DON'T KNOW THEY KNOW: WHAT TO DO ABOUT INFERENCES IN EUROPEAN AND CALIFORNIA DATA PROTECTION LAW

Allan E. Holder†

I. INTRODUCTION

One day in 2003, an irate man walked into a Target store in the State of Minnesota and demanded to speak to a manager.[1] The man's high school-aged daughter had been sent an advertisement with coupons for baby clothes and accessories, and he wanted to know why the store was encouraging his teenaged daughter to become pregnant.[2] A few days later, when the store's manager called to apologize to the man, the man in turn apologized profusely to the manager, revealing that his daughter had admitted to him that she was pregnant and due the following August.[3] The man's pregnant daughter had not disclosed this information to Target, but Target had ascertained it by analyzing the girl's shopping history with a computer program intended to predict the likelihood of a woman being pregnant.[4] The concerned man's daughter had likely purchased several out of a list of "twenty-five different products that, when analyzed together, allowed [Target] to . . . guess what trimester she was in—and estimate her due date—so Target could send her coupons when she was on the brink of making new purchases."[5]

Target's conclusion that the young girl was pregnant was an inference. An inference is, at its simplest, information that can be reasonably predicted from pre-existing data.[6] A person looking out their window to see that the sun is shining brightly and concluding that it is a hot day outside has made an inference about the weather, much like Target made an inference about the young girl's pregnancy. Nowadays, in the rare occasions in which inferences are included in the discourse over data privacy, it is in the context of large

DOI: https://doi.org/10.15779/Z38MP4VP1V

† J.D., 2021, University of California, Berkeley, School of Law.

1. CHARLES DUHIGG, THE POWER OF HABIT: WHY WE DO WHAT WE DO IN LIFE AND BUSINESS 196 (Random House ed., 1st ed. 2012).
2. *Id.*
3. *Id.*
4. *Id.* at 195.
5. *Id.*
6. *See* Michal Kosinski, David Stillwella & Thore Graepel, *Private traits and attributes are predictable from digital records of human behavior*, 110 PROC. NAT'L ACAD. SCI. 5802 (2013).

technology companies whose primary business models depend on using inferences created from data they collect in order to serve their users targeted advertisements.[7] Nonetheless, the collection of data and subsequent creation of inferences is not a phenomenon created by the growth of technology giants like Facebook and Google. Historically, corporate actors like insurance companies[8] and government actors like federal agencies[9] have also availed themselves of the practice of analyzing multiple data sets to infer or predict information about a subject.

In the information technology age, however, technological advancements and the popularization of the internet have enabled the creation of sets of data whose sheer size and level of detail had previously been unimaginable.[10] Companies can now collect information about a user's recent searches, purchases, and current location, as well as a user's preference for books, music, sports, and restaurants (among a host of other kinds of data points).[11] In turn, these organizations can analyze the data they acquire to predict, for instance, a user's age, gender, occupation, and education level.[12] These massive data sets, full of information both provided by and gathered about users (with and without their actual knowledge or consent), lend themselves to being processed by computer programs that find relationships between the data points and infer increasingly more detailed and intimate information about the user, as Target did.[13]

Inferences are often then used by organizations to make decisions about particular users, regarding everything from what ads to serve users based on their demographic profile, to creditworthiness, to suitability for employment, to insurance risk.[14] And even though companies have begun to create

7. *See, e.g.*, SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 78 (PublicAffairs ed., 1st ed. 2019).

8. PAM DIXON & ROBERT GELLMAN, WORLD PRIVACY F., THE SCORING OF AMERICA: HOW SECRET CONSUMER SCORES THREATEN YOUR PRIVACY AND YOUR FUTURE 16 (2014).

9. *See, e.g.*, *id.* at 62 ("The Health and Human Services Department effectively created a score that ultimately measures how sick a person is."); *id.* at 76 (". . .the [Department of Homeland Security's] program collects data about passengers and links the data with other sources of information to establish a risk score for each passenger. The Transportation Security Administration uses the scores to screen passengers.").

10. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 6 (2013).

11. *See* Kosinski, Stillwell & Graepel, *supra* note 6, at 5802.

12. *Id.*

13. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 156.

14. *See* FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY i, 19 (2014) ("The FCRA covers the provision of consumer data by

increasingly sophisticated, sensitive inferences and make decisions based on them, no organization can ever be one hundred percent certain of the veracity of an inference, and sometimes the inferences drawn are flat-out wrong.[15] Given inferences' increased sophistication and sensitivity, lack of certainty regarding their accuracy, and their role in decision-making, inferences have become increasingly important to both data-reliant organizations and privacy advocates.

In recent years, the European Union and several U.S. states have enacted new privacy legislation in order to grant users control over how their data is collected and used. The European General Data Protection Regulation (GDPR) came into effect on May 25, 2018, with the intent of acknowledging the rapidly developing field of consumer technology and its benefits for economic and social relations, while also establishing and protecting the principle that "the processing of personal data should be designed to serve mankind."[16] While the GDPR grants data subjects the rights to know about, rectify, delete, object to the processing of, and transfer personal data that a data controller might hold about them, the regulation significantly curtails subjects' rights when it comes to inferences.[17] In fact, there are gaps in the GDPR's jurisdiction over inferences, some of which have begun to be addressed only recently through jurisprudence from the European Court of Justice, and some of which remain unaddressed.[18]

In 2018, the State of California became the first U.S. state to enact a comprehensive data privacy law, setting an initial standard for what comprehensive data protection could look like in the United States and spurring a number of other states to follow its lead.[19] The California Consumer Privacy Act (CCPA) was passed with the goal of granting consumers similar

consumer reporting agencies where it is used or expected to be used for decisions about credit, employment, insurance, housing, and similar eligibility determinations"), ("In developing their products, the data brokers use not only the raw data that they obtain from their sources, such as a person's name, address, home ownership status, age, income range, or ethnicity ('actual data elements'), but they also derive additional data ('derived data elements').").

15. Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data*, 2019 COLUMBIA BUS. L. REV. 494, 509 (2019).

16. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) 1, Recital 4 [hereinafter GDPR].

17. Wachter & Mittelstadt, *supra* note 15, at 499.

18. *Id.*

19. *See* Sarah Rippy, US State Comprehensive Privacy Law Comparison, IAPP, https://iapp.org/resources/article/state-comparison-table/ (last visited Apr. 18, 2021).

rights to those of the GDPR when consumer data is held by businesses.[20] Enforcement of the CCPA began on January 1, 2020, and the statute called for the California Attorney General's Office to adopt regulations for the CCPA's enforcement.[21] In November 2020, voters in California approved passage of the California Privacy Rights Act, supplementing the CCPA with additional GDPR-like provisions.[22] While the California data privacy regime attempts to grant consumers rights with respect to inferences by classifying them as "personal information," this Note argues that language within the CCPA complicates interpretation of all subsequent regulations and statutes in the California regime by limiting certain rights to information directly collected from a consumer.[23]

While both the European Union and California have acknowledged the importance of inferences, they have only done so to an extent, and inferences' patchy legal status makes for an uncertain and confusing regulatory treatment regarding this type of information. The current gaps in the regulation of inferences left by these two major data privacy regimes are worrisome because the existence of these gaps constitutes suboptimal protection of inferences, going against the most basic justifications for the existence of each data protection regime. This Note argues that comprehensive regulation of inferences should be a top priority for legislators, regulators, and policy makers around the world as they undertake regulation of data privacy. This is because the subject of inferences touches directly on the very concerns over personal autonomy, safety, and dignity that historically justify the recognition of the privacy rights of individuals. The importance of the regulation of inferences is not underscored only by historical precedent. This Note also argues that modern data collection and inference creation practices create unprecedented risks to vulnerable populations that necessitate the strongest protections over personal data available under every data protection regime.

This Note proceeds as follows. Part II provides an overview of contemporary data collection practices and how they fuel the economic model that drives large technology organizations and a significant portion of the digital economy. Part III explains the current legal status of inferences under the GDPR in the European Union and under the California data privacy framework. In Part IV, I argue that inferences are personal data and users ought to have the highest degree of control available over them. I further

20. AB-375, 2017-2018 Assemb., Reg. Sess. (Cal. 2018).

21. CAL. CIV. CODE § 1798.185.

22. Erin Illman, Junaid Odubeko, Steve Snyder & Bradley Aaron Boult Cummings, *Steps for Proactive CPRA Compliance*, BLOOMBERG LAW (2020) (on file with the author).

23. *See infra* Section III.B.

contend that inferences' high level of sophistication and their close relation to identified or identifiable subjects makes them equivalent to the sort of information that the statutes in question cite as the original justification for the regulation of information privacy and protection of personal data. Moreover, inferences leave subjects vulnerable to the same kind of harms—to dignity, reputation, and personality—as other data that have been labeled personal information and over which users have been given rights. In addition to these harms, inferences by their nature increase the likelihood that an invasion of privacy could result in immediate or near-immediate threats to the physical security, autonomy, livelihood, or future prospects of the user. To illustrate the latter point, I draw on some experiences of women, LGBTQ people, and people of color to provide examples of the kind of novel invasions of privacy made possible by inferences that could uniquely affect marginalized groups. In Part V, I present a few anticipated arguments from stakeholders who might argue against strong user rights and protections over inferences. I also attempt to identify the actors who will be involved in resolving these conflicts as well as arguments that they ought to consider in adjudicating them.

II. INFERENCES AND THEIR IMPORTANCE TO THE DIGITAL ECONOMY

In the context of digital data, and for the purposes of this Note, inferences are defined as "information relating to an identified or identifiable natural person created through deduction or reasoning rather than mere observation or collection from the data subject," the data subject being the user to whom the information pertains.[24] In the aforementioned Target example, the company collected information about the teenage girl's purchases and then made the inference that the girl was pregnant. Just as Target sought to benefit from the creation of inferences based on a customer's information, actors of all stripes have come to depend on data collection and subsequent inference creation to better market their products and services, catalyzing the emergence of an entire economy based on the digital collection of user data.

Inferences are used primarily in two contexts: profiling and scoring.[25] Profiling is the practice of "using data from various sources to infer something about an individual, based on qualities of others who appear statistically similar."[26] A business might, for instance, scan a list of customers that

24. Wachter & Mittelstadt, *supra* note 15, at 515.

25. *See* FED. TRADE COMM'N, *supra* note 14, at 19; *see generally* Dixon & Gellman, *supra* note 8.

26. Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, at 7.

purchased camping gear in the last year, identify that they were all men from a certain postal code whose credit cards had a certain range of credit limits, and then use this information to put similar customers in the company's larger database in a category called "Customers Interested in Buying Camping Gear."[27] By way of example, a company called Recorded Future "captures historical data on consumers and companies across the [i]nternet," compiling it into profiles "to predict the future behavior of those consumers and companies."[28]

Scoring is the related practice of using inferences to rank users based on certain traits and behaviors.[29] For instance, some companies turn their analyses of customer data and interests "into marketing scores that . . . rank clients' customers on the basis of how likely they are to respond to particular marketing efforts or to make a purchase, their presence on the web or their influence over others, or other metrics."[30] Political campaigns have ranked television viewers by likelihood of voting for a candidate based on what they watch on cable television and when.[31]

Profiling and scoring can be and have been used by all manner of data-reliant organizations to help make determinations about users, concerning everything from ad targeting to employability, insurance risk, and creditworthiness.[32] However, over the course of the past two decades, technological advancements have helped usher in what scientists have called the era of Big Data, which could be defined (somewhat imperfectly) as a new period in the history of data wherein "the volume of information ha[s] grown so large that the quantity being examined no longer fit[s] into the memory that computers use for processing."[33] The popularization of certain technologies, both digital and physical, has allowed for the collection of massive and unprecedented quantities and types of user data. These large, comprehensive data sets have in turn enabled the creation of more complex and sophisticated inferences about users, increasing the ease and frequency with which scores are assigned, or profiles are created. In the physical realm, the manufacturers of everyday items humans depend on, such as cellphones and laptops (but also clothing, lightbulbs, microwaves, and toothbrushes), have begun augmenting

27. FED. TRADE COMM'N, *supra* note 14, at 19.

28. *Id.* at 9.

29. *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 2 (2014).

30. FED. TRADE COMM'N, *supra* note 14, at iii.

31. *See* Alice E. Marwick, *How Your Data Are Being Deeply Mined*, N.Y. REV. BOOKS (Jan. 9, 2014).

32. Citron & Pasquale, *supra* note 29, at 4.

33. MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 6.

and enhancing these products "through the use of emerging technologies—sensors, actuators, wireless connection, and embedded processing," all of which are able to collect data about users.[34] Through the chips that allow smartphones to make phone calls and connect to other accessories, cellular service providers[35] and even retailers[36] are able to track and log the physical locations of users with great precision. As we move further into the Internet of Things (IoT) era, suddenly our toothbrushes can log what times of day we brush our teeth and for how long,[37] our bathroom scales can keep a record of our weight fluctuations,[38] and our umbrellas can determine where we walked to and from in the rain.[39]

In the digital sphere, large technology services have developed the ability to follow users around the internet and observe their online habits, in order to add this data to the set of data they acquire directly by requesting it from the user. By way of example, Facebook's "Like" button plug-in, which "take[s] the form of a snippet of code to be added to a page,"[40] allows Facebook to recognize users across sites where the button has been encoded regardless of whether the user has "Liked" something or not.[41] Five months after the launch of the Like button in 2010, two million websites throughout the web had added

34. DAVID ROSE, ENCHANTED OBJECTS: DESIGN, HUMAN DESIRE, AND THE INTERNET OF THINGS 47 (Simon & Schuster ed., 1st ed. 2014).

35. *See* Nancy K. Oliver, *Location, Location, Location: Balancing Crime Fighting Needs and Privacy Rights*, 42(3) U. BALTIMORE L. REV. 485, 490 (2013).

36. *See* Michael Kwet, *In Stores, Secret Surveillance Tracks Your Every Move*, N.Y. TIMES (June 14, 2019), https://www.nytimes.com/interactive/2019/06/14/opinion/bluetooth-wireless-tracking-privacy.html.

37. *See Why Switch To A Bluetooth Electric Toothbrush?*, ORAL-B, https://oralb.com/en-us/why-switch/ (last visited Jan. 12, 2021) ("Oral-B's latest electric toothbrushes connects to the Oral-B app on your phone. The results? You'll get real-time feedback on your brushing. You'll know if you're brushing too hard, if you've brushed long enough and even if your brushing habits have improved over time.").

38. *See Body Composition Smart Scales*, WITHINGS, https://www.withings.com/us/en/scales (last visited Jan. 12, 2021) ("Data from every weigh-in automatically syncs to your smartphone via the free Health Mate app, available for iOS and Android.").

39. *See* Harry Hu, *HAZ: The World's 1st Motorized Smart Umbrella*, INDIEGOGO, https://www.indiegogo.com/projects/haz-the-world-s-1st-motorized-smart-umbrella#/ (last visited Feb. 21, 2020) (explaining the ability to perform location tracking with the umbrella and a smartphone app in order to locate the umbrella).

40. Tom Simonite, *Facebook's Like Buttons Will Soon Track Your Web Browsing to Target Ads*, MIT TECH. REV. (Sept. 16, 2015), https://www.technologyreview.com/s/541351/facebooks-like-buttons-will-soon-track-your-web-browsing-to-target-ads/.

41. Cotton Delo, *Facebook To Use Web Browsing History for Ad Targeting*, ADAGE (June 12, 2014), https://adage.com/article/digital/facebook-web-browsing-history-ad-targeting/293656.

the plugin to their pages.[42] Google engages in similar practices by "us[ing] web cookies to track browsing behaviour online by [users'] IP address to deliver targeted ads."[43] In 2007, Google acquired the advertising network DoubleClick, and with it, the massive database of user web-browsing records DoubleClick compiled by relying on "non-personally-identifiable information" to create user profiles.[44] While at first DoubleClick's anonymous browsing records were separate from Google's user profiles (which include personally identifiable information), in 2016 Google changed its policies and indicated that browsing records may in the future be combined with what the company already knows from a user's activities on Gmail and other Google services.[45]

Whereas in the past it was more common for companies to rely primarily on information collected directly from users with their awareness or consent, in the age of Big Data, companies can procure information about users without any action on the user's part. Unburdening the user from having to actively provide data allows data-reliant organizations to collect a significantly greater number of data points that the user creates in the normal course of their daily activities. As humans have become more and more reliant on the internet and connected devices, Big Data has demanded the development of novel approaches to analyzing data due to the sheer quantity now available to be analyzed. It has also allowed analysts and organizations to do more with data than they had been able to before—"to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more."[46] In short, it has allowed for the creation of highly sophisticated inferences.

Inferences have been and continue to be used by organizations to contribute to decisions made about users. In more recent times, however, internet companies like Facebook and Google have based their business

42. Leena Rao, *Five Months In, 2 Million Websites Using Facebook's New Social Plugins*, TECHCRUNCH (Sept. 29, 2010), https://techcrunch.com/2010/09/29/five-months-in-2-million-websites-using-facebooks-new-social-plugins/.

43. Olivia Solon, *Google's ad tracking is as creepy as Facebook's. Here's how to disable it*, THE GUARDIAN (Oct. 21, 2016), https://www.theguardian.com/technology/2016/oct/21/how-to-disable-google-ad-tracking-gmail-youtube-browser-history.

44. *See* Suzanne Monyak, *Google Changed a Major Privacy Policy Four Months Ago, and No One Really Noticed*, SLATE (Oct. 21, 2016), https://slate.com/technology/2016/10/google-changed-a-major-privacy-policy-and-no-one-really-noticed.html.

45. Julia Angwin, *Google Has Quietly Dropped Ban on Personally Identifiable Web Tracking*, PROPUBLICA (Oct. 21, 2016), https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking.

46. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 10, at 6.

models on using inferences to serve users targeted ads.[47] The companies take advantage of massive amounts of data about their billions of users to then score and profile them. They subsequently sell access to their users' attention to advertisers that have products or services that could appeal to certain types of user, or that certain types of user are more likely to purchase.[48] The revenue models of companies like Facebook and Google, which rely on selling users' attention to advertisers, effectively behoove them to collect as much information about users as they can in order to infer increasing amounts of information. With these inferences, they can then provide the most precise access to advertisers and remain competitive in their particular market sectors.

Data points that are usable for inferences are collected not only by large technology companies, but also by government agencies, banks, physical retailers of varying sizes, product manufacturers, and a myriad other kinds of organizations.[49] The Big Data economy has spawned companies called data brokers, whose main purpose is to "collect consumers' personal information and resell or share that information with others."[50] The existence of data brokers indicates that information collected from a user, with or without the user's awareness, is highly likely to end up in the hands of companies or organizations with which the user never intended to share this information.[51] It is foreseeable, then, for a health insurance company to preemptively learn that a person suffers from a chronic health condition, based on the user's shopping history showing that they purchase a certain medical aid every month and the user having searched for home remedies for certain symptoms. Additionally, data brokers themselves make inferences about consumers and use them to sort them into profiles, which they then provide to companies in order to serve consumers targeted ads.[52]

Large technology companies make use of artificial intelligence to process the enormous data sets that they compile from various sources, identify relationships between seemingly unrelated data points, and create inferences about users. Facebook, for instance, makes use of a type of artificial

47. *See generally* Zuboff, *supra* note 7; Caitlin Dewey, *98 Personal Data Points that Facebook Uses to Target Ads to You*, WASH. POST (Aug. 19, 2016), https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/.

48. Kurt Wagner, *This Is How Facebook Uses Your Data for Ad Targeting*, RECODE (Apr. 11, 2018), https://www.vox.com/2018/4/11/17177842/facebook-advertising-ads-explained-mark-zuckerberg.

49. *See generally* Dixon & Gellman, *supra* note 8.

50. FED. TRADE COMMISSION, *supra* note 14, at i.

51. *Id.*

52. *Id.* at 47.

intelligence called "deep neural networks" to process the large amounts of user data it collects in order to improve its ad targeting.[53] YouTube uses the same type of artificial intelligence to fuel its recommendations engine, which makes inferences about what kind of videos users will want to see based on a user's YouTube activity history, past searches, and demographic information.[54]

Deep neural networks are a type of artificial intelligence modeled after human brains so that computers can learn things, adapt the lessons they learn to new information, and make decisions in a human-like manner.[55] Engineers at large data-reliant organizations depend on the massive sets of data available to them to train neural networks. By way of illustration, Facebook could feed neural networks all the user data they have as well as the relationships within the data set, and by providing human validation of the outputs that the machine correctly produces, the model can slowly "learn" the correct relationships between the data points so that it is eventually able to make "predictions" about users based only on input data. Neural networks are so complex that their exact functioning is very difficult to understand, even for the experts who build them, and their complexity also makes it difficult to audit how they create sometimes very accurate relationships among data points.[56]

In this Note, and in most of the literature on machine learning, a "prediction" is meant to refer to an educated guess on the part of the artificial intelligence about unknown data based on known data. When the data set used to train a deep neural network consists of or includes personal data, the unknown facts could be future events that the artificial intelligence deems likely to happen ("User will likely purchase item X in the next seven days"), or current true facts about a user for which the artificial intelligence simply does not have confirmation ("User is a resident of California"). Both types of predictions can be viewed as inferences: the former about a user's propensity to a certain kind of behavior, and the latter about a certain fact about the user being true.

---

53. Cade Metz, *Building AI Is Hard—So Facebook Is Building AI That Builds AI*, WIRED (2016), https://www.wired.com/2016/05/facebook-trying-create-ai-can-create-ai/.

54. Paul Covington, Jay Adams & Emre Sargin, *Deep Neural Networks for YouTube Recommendations*, RECSYS '16: PROC. 10TH ACM CONF. ON RECOMMENDER SYSTEMS 191, 192 (2016).

55. Bernard Marr, *What Are Artificial Human Networks – A Simple Explanation for Absolutely Anyone*, FORBES (Sept. 24, 2018), https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/#23ad3ba12457.

56. Walter A. Mostowy, Note, *Explaining Opaque AI Decisions: How to Satisfy the GDPR's Right to an Ex Post Explanation*, 35 BERKELEY TECH. L.J. 1291 (2020).

As previously mentioned, deep neural networks and other forms of artificial intelligence make educated guesses about facts still uncertain to them, meaning that, barring a user confirming the veracity of an influence, data controllers are in possession of a host of inferences about a user that might range from being slightly inaccurate to flat-out erroneous. This poses dangers particularly when data-reliant organizations use inferences to make decisions that might alter the lives of the users, as will be explained in Part V of this Note.

One would expect a type of information that is essentially the driver of the modern digital economy—and which major data controllers invest so much to develop—to be stringently regulated. However, the widespread use of inferences and the enactment of comprehensive data protection statutes are such recent developments that the regulation of inferences so far remains imperfect, incomplete, and uncertain.

III. THE CURRENT LEGAL STATUS OF INFERENCES

Comprehensive data protection law in the European Union and the State of California both address the existence and especially sensitive nature of inferences. However, both regimes fail to fully protect inferences from misutilization, resulting in uncertainty over their treatment in both regimes.

A. THE EUROPEAN UNION

The European Union began enforcement of the GDPR on May 25, 2018.[57] It was originally adopted to replace the 1995 Data Protection Directive, which provided a framework on which E.U. member states could base their national data protection laws.[58] The GDPR was enacted as a binding regulation in order to create a standardized data protection regime across all E.U. member states, and partly also to "address contemporary privacy challenges, such as those posed by the [i]nternet, social media, mobile apps, cloud computing, 'big data,' and behavioral marketing."[59]

57. Nikhil Kalyanpur & Abraham Newman, *Today, A New E.U. Law Transforms Privacy Rights for Everyone. Without Edward Snowden, It Might Never Have Happened*, WASH. POST (May 25, 2018), https://www.washingtonpost.com/news/monkey-cage/wp/2018/05/25/today-a-new-eu-law-transforms-privacy-rights-for-everyone-without-edward-snowden-it-might-never-have-happened/.

58. *See* GDPR, *supra* note 16, Recitals 9, 10, art. 94.

59. W. Scott Blackmer, *GDPR: Getting Ready for the New EU General Data Protection Regulation*, INFOLAWGROUP (May 5, 2016), https://web.archive.org/web/20180514111300/https://www.infolawgroup.com/2016/05/articles/gdpr/gdpr-getting-ready-for-the-new-eu-general-data-protection-regulation/; *see also* GDPR, *supra* note 16, Recital 6.

The GDPR grants data subjects a number of rights over their "personal data," defined as "any information relating to an identified or identifiable natural person."[60] Under the GDPR, data subjects have the following rights:

- to basic information about the personal data collected regardless of its source (Art. 13–14);
- to access the data that has been collected (Art. 15);
- to rectify inaccurate data (Art. 16);
- to have personal data permanently erased (Art. 17);
- to restrict the processing of personal data (Art. 18);
- to transfer personal data between controllers (Art. 20);
- to object to processing of personal data, including for the purposes of direct marketing (Art. 21); and
- to not have decisions made about the data subject on the basis of automated processing of personal data (Art. 22).[61]

The regulation further earmarks certain special categories of personal data that data subjects have additional rights over, such as data "revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership," as well as "genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation."[62] Processing of these special categories of personal data is prohibited save for a number of exceptions, including when the data subject has granted explicit consent,[63] when they have manifestly made the data public,[64] or when the processing is necessary for reasons of substantial public interest.[65]

While the GDPR itself does not designate inferences as personal data (or even mention inferences at all), the views of the Article 29 Data Protection Working Party (hereinafter, "the Working Party") and recent European Court

60. GDPR, *supra* note 16, art. 4(1).
61. GDPR, *supra* note 16.
62. *Id.* art. 9(1).
63. *Id.* art. 9(2)(a).
64. *Id.* art. 9(2)(e).
65. *Id.* art. 9(2)(g).

of Justice (ECJ) jurisprudence provide a sketch of the current (somewhat convoluted) legal approach to inferences in the European Union.[66]

The Working Party adopted the broadest definition of personal data. The positions of the Working Party and its successor, the European Data Protection Board (EDPB), are recommendations for the practical application of European data protection laws, and have no binding legal effect.[67] The GDPR established the EDPB to replace the Working Party as soon as the regulation came into effect in May 2018, and the EDPB endorsed all of the Working Party's recommendations.[68] According to one of the Working Party's position papers, personal data is created by any data processing whose content, purpose, or result relates to an identifiable person directly or indirectly.[69] The Working Party's test makes it so that data can still be classified as personal data even if it does not describe an identifiable person, or even if it will not be used to make a decision about a person, so long as it has the potential to impact "an identifiable person's rights and interests."[70] Consequently, a data inference that does not include an identified data subject but still has the potential to affect the data subject's life is personal data in the view of the Working Party. While the Working Party's views, as previously mentioned, are merely recommendations, they have been influential in ECJ data protection jurisprudence and are expected to continue to play a role in the court's reasoning in these matters.[71]

In contrast to the Working Party, the ECJ has generally adopted more limited interpretations of the concept of personal data, and its decisions are legally binding across the European Union. In two recent cases, *YS, M and S v. Minister voor Immigratie, Integratie an Asiel*[72] and *Peter Nowak v. Data Protection*

66. Wachter & Mittelstadt, *supra* note 15, at 498.

67. Tim Wybitul, *GDPR Guidance – European Data Protection Board Adopts Art. 29 Working Papers*, HOGAN LOVELLS CHRONICLE OF DATA PROTECTION (May 2018), https://www.hldataprotection.com/2018/05/articles/international-eu-privacy/gdpr-guidance-european-data-protection-board-adopts-art-29-working-papers/.

68. *Id.*

69. Article 29 Data Prot. Working Party, *Guidelines on the Right to Data Portability*, 16/EN, WP242rev.01, at 9–11 (Dec. 13, 2016), https://ec.europa.eu/newsroom/document.cfm?doc_id=44099.

70. *See* Wachter & Mittelstadt, *supra* note 15, at 518. While the Working Party has not given examples, the reading of the text implies that, for instance, information that was inferred, anonymized, and aggregated can be considered personal data if it is likely to become re-identified as a consequence of an inference attack.

71. *See* Wybitul, *supra* note 67; Wachter & Mittelstadt, *supra* note 15, at 498 n.2.

72. Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081 [hereinafter *YS, M and S*].

Commissioner,[73] the ECJ shined some light on its interpretation of what "personal data" encompasses, but the decisions split on some significant points in a way that leaves the legal status of inferences unresolved.

First, the cases split over whether "personal data" includes "opinions, reasoning, and assessments that underlie" final decisions—in other words, the intermediate inferences that lead to a final, inferred result.[74] The ECJ in *Nowak* held that any comments made by an examiner with respect to the exam answers of a data subject—which constitute inferences in that they are assessments of the data subject's knowledge and competence in the exam's field, based on information provided by the data subject[75]—are personal data.[76] In *YS, M and S*, the ECJ held that an explanatory legal analysis included in the file relating to a data subject's application for a residence permit is not personal data even though it may contain personal data.[77] The ECJ in this case characterized the legal analysis at issue as "information about the assessment and application by the competent authority of [the relevant law] to an applicant's situation," therefore designating it as an inference created from subject-provided data.[78] However, the court still concluded that facts about a data subject, "such as the applicant's name, date of birth, nationality, gender, ethnicity, religion and language," constitute personal data in the context of an application for a residence permit.[79]

The court seems to contradict itself, stating in *Nowak* that inferences in the form of assessments, opinions, or reasonings are classified as personal data under the GDPR, while appearing to state the contrary in *YS, M and S*. The key lies partly in the court's implied argument in *Nowak*. The court proposes that if having the right to correct, erase, or block processing of certain information would serve the GDPR's purpose of "guaranteeing the protection of the [data subject]'s right to privacy with regard to the processing of data relating to [them]," then the data subject should have a right to *access* information about them in order to exercise those very rights.[80] If so, then in order for the data subject to have the right to access the information, the information needs to be granted the designation of personal data.[81]

---

73. Case C–434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-582 [hereinafter *Nowak*].

74. Wachter & Mittelstadt, *supra* note 15, at 537–38.

75. *Nowak*, *supra* note 73, ¶ 43.

76. *Id.* ¶ 62.

77. *YS, M and S*, *supra* note 72, ¶ 39.

78. *Id.* ¶ 40.

79. *Id.* ¶ 38.

80. *Nowak*, *supra* note 73, ¶¶ 56–57.

81. *Id.*

In *Nowak*, the ECJ found that the examinee had a legitimate interest in being able to, for instance, correct potential mix-ups where another examinee's answers, or comments by the examiner (i.e., inferences), are mistakenly ascribed to the data subject.[82] The ECJ also found a legitimate interest of the examinee in being able to prevent the sharing or publication of their answers to the examination.[83] Consequently, the examinee deserves the right to protect their right to privacy by accessing the information, making the information personal data.

In contrast, in *YS, M and S,* the court held that the legal analysis pertaining to a decision of residency, considered separately from the facts about the data subject, cannot be checked for accuracy, as it is an application of law to facts.[84] It also cannot be corrected, given that any attempts at correction would constitute an appeal to an administrative decision, a procedure that the GDPR does not provide for.[85] Extending a right to access to the legal analysis would therefore not serve the GDPR's purpose of helping the data subject in protecting their right to privacy by correcting, erasing or blocking processing of information about them. Since the data subject does not have a claim to a right to access the legal analysis, the information is not personal data.

Both cases left unresolved the issue of whether final inferences (the data that results from processing) are personal data. Sandra Wachter and Brent Mittelstadt indicate that it is highly likely that final inferences will be considered personal data, given that that in this regard the court seemed to rely on the views of the Working Party, which include output data in their definition of personal data.[86]

In other respects, the ECJ's position conflicted with that of the Working Party. The ECJ's view is that the rights of data subjects over inferences are limited, and from the two rulings it becomes clear that whether the rights apply, according to Wachter and Mittlestadt, "must be interpreted according to the purposes for which the data was collected."[87] The ECJ rulings in the cases "clarify that the remit of data protection law is not to assess the accuracy of the reasoning behind decisions and assessments, or the accuracy of the decisions and assessments themselves," and the court grounds its holding on the fact that, in the GDPR itself, there are broad exemptions to the right to

82. *Id.* ¶ 54.
83. *Id.* ¶ 50.
84. *YS, M and S*, *supra* note 72, ¶ 45.
85. *Id.* ¶ 46.
86. Wachter & Mittelstadt, *supra* note 15, at 538.
87. *Id.* at 538.

access.[88] The rights to correct, erase, or block processing of inaccurate data seem to exist simply so that data subjects can verify that the initial data that will be processed to create inferences and/or make decisions about the data subjects (like the examination data in *Nowak* and the pre-legal analysis facts in *YS, M and S*) are complete and accurate. Data subjects do not have the same rights over the data resulting from the processing because such data in some cases will not be processed further and, in some cases like in *YS, M and S*, processing does not impinge on the data subject's right to privacy.

As it stands, the legal status of inferences under the GDPR is puzzling, and certainty regarding said status is in its infancy. While the Working Party embraces a broad definition of the concept of personal data that would include inferences, data controllers in the European Union are bound only by ECJ jurisprudence, which seems to imply that inferences deserve personal data status only as long as data subjects have an interest in correcting, erasing, or blocking processing of them.

B. CALIFORNIA

Given the relative recency of the California data privacy framework, there have been no significant public debate or litigation pertaining to the subject of inferences. A reading of the regime's statutes and the attorney general's subsequent regulations, however, reveals that the legal status of inferences under the framework is very uncertain: perhaps more so than in Europe, and in ways that invite future amendments and litigation regarding this unique type of data.

The California Consumer Privacy Act (CCPA) was passed on June 28, 2018, and its enforcement began on January 1, 2020.[89] The legislation designated the California Attorney General as its primary enforcer,[90] and also required them to enact certain regulations to make the legislation effective.[91] On November 3, 2020, the California electorate voted to approve the California Privacy Rights Act (CPRA), which amended key provisions of the CCPA to mirror some of the provisions of the GDPR.[92] The CPRA created a new agency called the California Privacy Protection Agency ("the Agency"), which is slated to take over from the California Attorney General as privacy

88. *Id.* at 539–40.

89. *See* CAL. CIV. CODE § 1798.100.

90. *See id.* § 1798.155(b).

91. *Id.* § 1798.185(a).

92. Illman, Odubeko, Snyder & Boult Cummings, *supra* note 22; *see also*, California Privacy Rights Act of 2020, Proposition 24, 1879 (19-0021A1), Amends Consumer Privacy Laws—Initiative Statute, https://www.oag.ca.gov/system/files/initiatives/pdfs/19-0021A1%20%28Consumer%20Privacy%20-%20Version%203%29_1.pdf.

enforcer "beginning the later of July 1, 2021, or six months after the Agency provides notice to the Attorney General that it is prepared to begin rulemaking" under the CPRA.[93] Taken as a whole, the CCPA, the regulations enacted by the California Attorney General pursuant to the CCPA, and the CPRA constitute the California data privacy framework.

The CCPA grants consumers a number of rights over "personal information," defined in the statute as "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household."[94] The statute goes on to list categories of information that shall be considered personal information under the statute if they are capable of being "reasonably linked, directly or indirectly," with a particular consumer or household.[95] These categories include unique or semi-unique identifiers (like real names and aliases, but also IP addresses and postal addresses), biometric information, internet activity information, geolocation data, employment information, education information, and "[c]haracteristics of protected classifications under California or federal law."[96] Notably, the final category that shall be considered personal information according to the CCPA is "[i]nferences," thereby designating as personal information any information mentioned in the previous categories that has been combined or processed "to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes."[97] Further, the act defines "infer" or "inference" as "the derivation of information, data, assumptions, or conclusions from facts, evidence, or another source of information or data."[98]

While the CCPA's definition of inference and its designation of inferences as personal information are seemingly comprehensive enough to secure robust protection of the sensitive information that inferences can carry, their effectiveness is impeded by the statutory language that actually grants consumers rights over their data. The CCPA grants consumers six rights regarding businesses' use of consumers' personal information, four of which closely mirror rights that the GDPR grants users over their personal data. However, when the CCPA grants consumers the right to access,[99] delete,[100] or

93. CPRA Sec. 21, § 1798.185(d).
94. CAL. CIV. CODE § 1798.140(o)(1).
95. *See id.*
96. *Id.* § 1798.140(o)(1)(A)–(J).
97. *Id.* § 1798.140(o)(1)(K).
98. *Id.* § 1798.140(m).
99. *Id.* § 1798.100.
100. *Id.* § 1798.105.

request disclosure of information collected,[101] the right only applies to information *collected* about a consumer. The right to request disclosure of information sold similarly grants the consumer the right to compel a business to disclose the categories of information "collected" about a consumer, before allowing the consumer to demand a business disclose the categories of personal information sold.[102] The CCPA appears to internally contradict itself, creating uncertainty as to the existence and breadth of rights consumers have over inferences.

The statute's own definition of inference-creation as "derivation" of data from existing information does not fit comfortably within any of those activities enumerated in the statute's definition of "collection." The statute defines "collection" as the "buying, renting, gathering, obtaining, receiving, or accessing [of] any personal information pertaining to a consumer by any means," and it advises that the definition "includes receiving information from the consumer, either actively or passively, or by observing the consumer's behavior."[103] The creation of inferences, however, is most accurately characterized as creating or predicting information about a consumer, activities that could be read to be forms of "gathering" or "obtaining" information about a consumer. Such an interpretation of the statute is likely to be challenged given the imprecise fit of inference-creation into those terms. This uneasy conceptual fit would leave open the matter of whether inference creation falls under "collection" of personal information until courts rule on it, the CCPA is amended, or the enforcer of the California framework provides further clarifying regulations.

The aforementioned right to disclosure of information sold is unique, in that it grants consumers a glimpse into the inferences a business may have made about them. As previously stated, the consumer has the right to obtain certain details about the personal information that has been collected about them, in which case inferences would not be included. However, the statute also grants the consumer the right to know "the categories of personal information that the business sold about the consumer,"[104] which expands the scope of the information subject to disclosure to any personal information about the consumer that has been sold. Since the scope is not limited only to personal information that has been collected, the consumer may have the right to know the categories of information created or derived by the business, which would include inferences.

101. *Id.* § 1798.110.
102. *Id.* § 1798.115.
103. CAL. CIV. CODE § 1798.140(e).
104. *Id.* § 1798.115(a)(2).

The CCPA's right to opt-out allows a consumer "at any time, to direct a business that sells personal information about the consumer to third parties not to sell the consumer's personal information."[105] It therefore seems to wholly cover inferences by virtue of not being limited to personal information that has been collected. However, in its subdivision detailing compliance obligations for businesses, the CCPA specifies that when a consumer exercises their right to opt-out, a business shall "refrain from selling personal information collected by the business about the consumer."[106] Inferences are therefore included in the scope of the right of the consumer, but left out of the scope of the obligation on the business, making it ultimately unclear whether inferences are or are not covered by the CCPA's opt-out procedure.

The CCPA offers confusing treatment of inferences in its right to nondiscrimination as well. The statute's right to nondiscrimination protects consumers from receiving different treatment from a business—including denial of goods or services, differences in price or quality of goods or services, or the suggestion thereof—as a result of having exercised one of the previously mentioned rights.[107] The right also allows businesses to "offer financial incentives, including payments to consumers as compensation, for the collection of personal information, sale of personal information, or deletion of personal information."[108] In effect, the right allows businesses to buy from consumers their rights to have personal information about them be collected, sold, or deleted. It does not, however, explicitly allow businesses to buy the right to derive or create personal information about consumers. The lack of this specific allowance for businesses has two possible implications. At its least damaging, the CCPA once again fails to recognize that businesses come into possession of personal information about consumers by means other than pure collection, and thus businesses are not encouraged to compensate consumers for the creation of sensitive inferences about them that constitute personal information. At its most troublesome, the statute simply does not recognize the rights of consumers over information created or derived about them as a valuable right or as a right that exists at all. In either case, businesses are made to believe that their creation of sensitive inferences is of vastly less importance than their collection and sale of personal information.

It is not only the core rights of consumers that are limited to personal information collected by businesses. Several additional sections of the CCPA are similarly limited, including subsections dedicated to compliance obligations

105. *Id.* § 1798.120(a).
106. CAL. CIV. CODE § 1798.135(a)(4).
107. *Id.* § 1798.125(a).
108. *Id.* § 1798.125(b)(1).

for businesses, applicability of the Act, and exemptions, thereby making them inapplicable to inferences.[109] The CCPA's recurring references to information "collected from the consumer" in defining the rights it confers hints at the possibility that the drafters of the statute failed to consider that derived or inferred information can be among the most sensitive personal information about a consumer, and among the most valuable information for businesses to access.

The California Attorney General's guidance, instead of providing clarification on the uncertainty over collected information, also relies on the limited concept of information "collected from the consumer" by businesses to attempt to explain how businesses are to comply with the statute.[110] For instance, the CCPA explains in detail businesses' obligations with respect to the consumers' right to access, which requires that businesses notify the consumer of their personal information collection practices.[111] The Attorney General's regulations call this the "notice at collection" requirement, which they define as "the notice given by a business to a consumer at or before the time a business collects personal information from the consumer."[112] The language in the regulations still limits the notice requirement to information collected, and it further proscribes behavior of businesses based only on collection, prohibiting them from collecting—but not creating—any types of information not included in the notice at collection.[113] Businesses are also prohibited from collecting information from consumers—but not creating information about them—if the notice at collection is not given.[114] These limitations in the Attorney General's regulations exclude inferences.

In some instances, the regulations might give consumers rights over inferences where the CCPA does not, like in their requirements that businesses' privacy policies "explain that a consumer has the right to request that the business disclose what personal information it collects, uses, discloses, and sells"[115] or that privacy policies explain to the consumer their right to opt out of the sale of their personal information.[116]

109. *See, e.g.*, *id.* § 1798.175 (on Applicability); *id.* § 1798.135 (on Compliance Obligations); *id.* § 1789.145 (on Exemptions).

110. *See generally* CAL. CODE REGS. TIT. 11, DIV. 1, CH. 20 [hereinafter "CCPA Regulations"].

111. CAL. CIV. CODE § 1798.100(b).

112. CCPA Regulations § 999.301(l).

113. *See Id.* § 999.305(a)(5).

114. *Id.* § 999.305(a)(6).

115. *Id.* § 999.308(c)(1)(a).

116. *Id.* § 999.308(c)(3)(a).

The CPRA amended the CCPA to significantly change and expand the rights that the California framework grants to consumers.[117] Notable among the changes is the creation of a new category of data, "sensitive personal information," which consists of any personal information that reveals any of a number of personal data points such as social security numbers, racial or ethnic origin, financial account information, and biometric information.[118] Perhaps drawing from the GDPR's treatment of its special categories of data, the CPRA stipulates that consumers should be able to control the use of their sensitive personal information because its unauthorized use or disclosure "creates a heightened risk of harm to the consumer."[119] The CPRA also carves out special provisions aimed at addressing inferences by stating that the collection or processing of sensitive personal information for inference-creation purposes obligates businesses to comply with additional requests from consumers to limit or disclose use.[120]

The types of data that the drafters of the CPRA chose to designate as sensitive personal data are similar to the sort of information that inferences can reveal about consumers. However, the rights afforded to users over their sensitive personal information, including the aforementioned rights introduced by the CPRA, appear to also still be limited by the fact that they only apply to data "collected," and not created or derived, from the consumer. The newly created California Privacy Protection Agency is tasked with issuing new regulations to clarify the California framework in advance of the CPRA's effective date of January 1, 2023.[121] As of early 2021, however, the text of the recently passed statute seems to leave the legal status of inferences as uncertain as it was in the CCPA and the California Attorney General's regulations.

The fact that the CCPA includes and defines the concept of an inference and inference-making, as well as the fact that the CPRA acknowledges the significance of inference-creation, implies that the drafters of the California framework acknowledge the sensitive nature of inferences and the importance of protecting them by granting consumers rights over them. At some point in

117. Bret Cohen, Tim Tobin & Aaron Lariviere, *Understanding the new California Privacy Rights Act: How businesses can comply with the CPRA*, HOGAN LOVELLS ENGAGE (Nov. 25, 2020), https://www.engage.hoganlovells.com/knowledgeservices/news/understanding-the-new-california-privacy-rights-act-how-businesses-can-comply-with-the-cpra?nav=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQ71hKXzqW2Ec%3D&key=BcJlhLtdCv6%2FJTDZxvL23TQa3JHL2AIGr93BnQjo2SkGJpG9xDX7S2thDpAQsCconWHAwe6cJTn0ksYg%2Fo%2FRPN3OjGmtaEjr&uid=iZAX%2FROFT6Q%3D.

118. *See* CPRA Sec. 14, § 1798.140(ae).

119. CPRA Sec. 3(A)(2).

120. *See* CPRA Sec. 10, §§ 1798.121(a), (d).

121. Cohen, Tobin & Lariviere, *supra* note 117.

the drafting process, that intention was either pushed to the side or forgotten, since key portions of the California framework grant consumers rights using the CCPA's own defined concepts in ways that make the different documents contradict themselves. Ultimately, the legal status of inferences under the framework remains up in the air. The resulting uncertainty means that there is still plenty of room for the California framework to be amended to remedy this uncertain status, as well as a chance that certain contradictory portions of the documents will be litigated in order to address the uncertainties.

The legal status of inferences in California is contradictory and uncertain, leaving this sensitive data without adequate protection. As previously seen, it is not only the case that inferences are inadequately protected in California, which has come to be seen as a floor for privacy protections in the United States since the enactment of the CCPA. E.U. statutes, policy, and jurisprudence have also left significant questions unanswered regarding the legal status of sensitive personal information such as inferences. The patchy and uncertain legal and regulatory framework surrounding inferences poses a number of significant dangers for global and regional data protection regimes generally, and specifically for users and their rights to control their personal information.

IV. USERS DESERVE THE STRONGEST POSSIBLE RIGHTS WITH REGARD TO INFERENCES

Inferences are particular in that they allow data controllers and businesses to know information about users' personal lives without having to ask for such data. The way highly sophisticated inferences are created about a person depends on a phenomenon that is near invisible to the lay user (the collection of massive amounts and types of data) and takes advantage of a technology that sounds inconceivable, like something out of a science fiction film (deep neural networks, or computers that can be trained to think like humans). Because of these reasons, it is understandable how the drafters of recent data protection statutes and regulation could have neglected to attach due importance to inferences.

As previously mentioned, however, inferences can contain or reveal some of the most sensitive information about a person's life, like their pregnancy status, sexual orientation, or race. Were these sorts of sensitive inferences to end up in the hands of bad actors—or even organizations that are not supposed to have access to certain types of information—users could experience tangible, adverse consequences in their everyday lives.

Current and future data protection statutes and regulations should grant users the strongest possible rights over inferences made about them. Under

the GDPR, all inferences made with personal data as their starting point should be afforded the designation of personal data, with special consideration given to the fact that many inferences contain those special categories of personal information, the processing of which is prohibited save for a few exceptions.[122] With regard to the California framework, all consumer rights whose application is limited to information "collected from" the user or "maintained by" the business should be expanded to cover information "inferred" or "created" based on the user's personal data, as well. Further, all language in the framework's statutes should be clarified in such a way that inferences are clearly covered.

These changes should be implemented for two reasons, which I elaborate upon in the following Sections. Firstly, both the GDPR and the CCPA state they were enacted to protect the very kind of sensitive information that makes up inferences. In fact, inferences are among the type of information that first signaled the need for the protection of personal data in general in the early days of the information technology boom. Secondly, the sophisticated nature of modern inferences—and the information that data-reliant organizations gain access to because of inferences—mean that invasions of privacy involving inferences can, unprecedentedly, result in immediate or near-immediate threats to the physical integrity, autonomy, and livelihood of data subjects.

A. DATA PROTECTION STATUTES EXIST PRECISELY TO PROTECT THE SORT OF INFORMATION THAT INFERENCES CONSTITUTE

It could be argued that inferences are such a hyper-specific type of data, or so pivotal to the modern internet economy, that European and California lawmakers can be forgiven for instituting frameworks that treat inferences so haphazardly. However, to dive into the justifications for the existence of these data protection regimes—and even the history of data protection as a field—is to realize that inferences and other similar types of data are the animating issue for the enactment of these laws and regulations.

In its Recital 6, the GPDR acknowledges that "[r]apid technological developments . . . have brought new challenges for the protection of personal data," and that the "scale of the collection and sharing of personal data has increased significantly."[123] It goes on to acknowledge that modern technology "allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities," and that while data should flow freely within the European Union, the GDPR must

122. *See supra* text accompanying notes 61–63.
123. GDPR, *supra* note 16, Recital 6.

ensure a high level of protection of personal data.[124] The unprecedented use and scale of collection of personal information mentioned are undoubtedly a reference to, among other modern data practices, the creation and use of inferences.

The GDPR's preamble also notes that the 1995 Data Protection Directive, while necessary, had several weaknesses that resulted in differing levels of protection of personal data.[125] It also notes that "[e]ffective protection of personal data throughout the [European] Union requires the strengthening . . . of the rights of data subjects and the obligations of those who process and determine the processing of personal data."[126] In Recital 13, the GDPR lays the foundation for the new data protection regime it will establish by stating that, in light of the aforementioned points, the GDPR is necessary to grant all natural persons across the European Union equal levels of legally enforceable rights and ensure consistent monitoring of the processing of personal data.[127]

The creation of highly sophisticated inferences thanks to massive, widespread data collection is exactly the kind of phenomenon the GDPR concerns itself with and mentions as its reason for being in its Recital 6.[128] Such a clear indication of the type of data that necessitated stronger data rights for users implies that inferences, due to their nature, should be afforded the strongest protections under the GDPR.

The California Legislature's language in the CCPA's preamble is striking in that it seems to heavily allude to inferences. In the CCPA bill's preambular § 2, the California Legislature acknowledges that the right to privacy is among the inalienable rights granted by the state's constitution.[129] It goes on to find and declare, "As the role of technology and data in the every daily lives of consumers increases, there is an increase in the amount of personal information shared by consumers with businesses," and that "California law has not kept pace with these developments" and their implications for personal privacy."[130] Remarkably, § 2(e) acknowledges the unprecedented kinds of personal information that businesses may now have access to by stating: "[Businesses] may know where a consumer lives and how many children a consumer has, how fast a consumer drives, a consumer's personality, sleep habits, biometric and health information, financial information, precise

124. *Id.*
125. *See id.* Recital 9.
126. *Id.* Recital 11.
127. *See id.* Recital 13.
128. *See id.* Recital 6.
129. AB-375, 2017-2018 Assemb., Reg. Sess. (Cal. 2018) § 2(a).
130. *Id.* § 2(d).

geolocation information, and social networks, to name a few categories," all information that can be derived or inferred from data that businesses routinely collect.[131] In the following subsection, the CCPA speaks to the potential consequences of the mishandling of sensitive personal information: "The unauthorized disclosure of personal information and the loss of privacy can have devastating effects for individuals, ranging from financial fraud, identity theft, and unnecessary costs to personal time and finances, to destruction of property, harassment, reputational damage, emotional stress, and even potential physical harm."[132] The preamble closes by stating that, in light of the aforementioned facts, it is the intent of the Legislature to "further Californians' right to privacy" by enacting the CCPA.[133]

Mentions within the CCPA of the increased amounts of personal information that consumers share as a result of modern technology are a reference to the widespread collection of personal data that animates not only the modern digital economy, but also all modern data protection. That clause, paired with the specific mention of several potential data points that are often inferred rather than requested (such as a consumer's personality, sleep habits, and social connections) points toward the fact that inferences were one of, if not *the* primary motivation behind efforts to beef up California consumers' rights over their personal data. As previously mentioned, the CCPA ends up contradicting itself in ways that result in a lackluster protection of inferences, which might have been the result of a rushed drafting process and legislative trajectory for the CCPA.[134] However, the fact that California's opening salvo in addressing data protection justifies this effort by name-checking inferences bolsters the idea that the CCPA should be amended to remedy the patchy protection of this sensitive type of data.

But the significance of data inferences was recognized much earlier than these recent data protection efforts in Europe and California. There is precedent for the stronger protection of inferences within the early history of data privacy rights and data protection, when experts and advocates were first noticing the implications of information technology on personal privacy. Technological developments that facilitated the processing of data—and

131. *Id.* § 2(e).

132. *Id.* § 2(f).

133. *Id.* § 2(i).

134. *See generally* Issie Lapowsky, *California Unanimously Passes Historic Privacy Bill*, WIRED (June 28, 2018), https://www.wired.com/story/california-unanimously-passes-historic-privacy-bill/ ("The so-called California Consumer Privacy Act of 2018 (AB 375) was introduced late last week by state assemblymember Ed Chau and state senator Robert Hertzberg, in a rush to defeat a stricter privacy-focused ballot initiative that had garnered more than 600,000 signatures from Californians.").

ultimately facilitated the creation of inferences—are among the factors that first animated the regulation of personal data in the United States. Paul Schwartz and Daniel Solove explain that numerous American privacy laws turn on the concept of personally identifiable information (PII), for which there is no uniform definition.[135] PII can be said to equate to the concept of personal data, and Schwartz and Solove indicate that it "first became an issue in the 1960s with the rise of the computer," which not only allowed entities both public and private to collect more information, but also to process that information in unprecedented ways.[136] Computers removed the limiting factors for how data could be stored and retrieved, and "permitted information to be searched and organized by *multiple attributes* rather than simply through a single index, as, for example, a person's first and last name."[137] This technological development "changed the way information could be linked to an individual,"[138] and facilitated the practice of inference-making, albeit a rudimentary version of the practice conducted by humans instead of artificial intelligence. The new data handling capabilities subsequently "required Congress to confront the issue of the kinds of information that should matter for information privacy law."[139] Following this unprecedented expansion of the level of access to personal information, Congress—albeit with much delay—responded by starting to enact privacy legislation that strongly protected PII by making the presence of PII the trigger for privacy protections. According to Schwartz and Solove, Congress enacted the Cable Communications Policy Act of 1984 ("the Cable Act") and included this new approach to PII in response to how technological advances made it theoretically possible for consumers to send information to broadcasters and television operators via their televisions, utilizing a technology called "videotex."[140] Policymakers were concerned that "by collecting these data, the cable operator would be able to construct detailed profiles about viewing choices" and derive information about viewers' interests from them.[141]

The ease with which PII could be collected and processed triggered a rethinking of the legislative and policymaking approaches to privacy protection precisely because of concerns that technology made PII more vulnerable to exploitation and the creation of new information based on it, in the form of

135. 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, 1816 (2011).

136. *Id.* at 1820.

137. *Id.*

138. *Id.*

139. *Id.* at 1821.

140. *Id.* at 1826.

141. *Id.*

inferences. Lawmakers have once, in the past, acknowledged the threat to privacy posed by technological advances that made it easy to collect and derive more information about individual persons. It follows that it is appropriate and necessary for lawmakers to implement changes to new privacy legislation that are responsive to AI-facilitated inference-making, which, like computers were in the 1960s, is the novel data-handling practice that threatens personal privacy in contemporary times.

Granting users more protection over inferences is a policymaking approach that is supported not only by present justifications for data protection regimes in the very text of the GDPR and CCPA statutes, but also by the history of data protection policymaking itself. These sources show that when technological advancements bring about uncertainty as to the protection of sensitive personal information, legislative and policy action aimed at adapting to the changes is logical and appropriate. The stronger protection of inferences did not come slowly in response to then-present harms. In the 1980s, potential and anticipated technologies like videotex, as well as their potential harms, were enough to justify strong protection of PII in the Cable Act. Contemporary policy- and lawmakers should mimic their predecessors and consider the potential harms that sophisticated, machine-learning-powered inferences could represent in the future when finding justifications for the strong protection of inferences in contemporary data protection regimes.

B. INVASIONS OF PRIVACY INVOLVING INFERENCES MAY INVOLVE HARMS TO INTERESTS OTHER THAN THE REPUTATIONAL OR DIGNITARY

When justifications for privacy regulation are offered, they often focus on the individual interests that privacy rights seek to protect. According to Paul Schwartz and Karl-Nikolaus Peifer, William Prosser conceived of the four modern privacy torts (intrusion upon seclusion, public disclosure of private facts, false light, and appropriation) as intended to protect the rights-holder from offensive behavior, attacks against their reputation, mental distress, and exploitation of their image for financial gain.[142] The forefathers of privacy law, Samuel Warren and Louis Brandeis, thought of the right to privacy as protecting a person's "inviolate personality," which Edward Bloustein

142. *See* Paul M. Schwartz & Karl-Nikolaus Peifer, *Prosser's* Privacy *and the German Right of Personality: Are Four Privacy Torts Better Than One Unitary Concept?*, 98 CALIF. L. REV. 1925, 1941–42 (2010).

expanded upon to mean "the individual's independence, dignity, and integrity."[143]

The aforementioned justifications within the text of the GDPR and CCPA, by their mention of technological advancements and the unprecedented types of information businesses can access, hint at efforts to protect the same dignitary and reputational interests that Prosser, Warren, and Brandeis enumerated, but with an eye towards accounting for other interests that might be unforeseeably invaded. By virtue of the way they are created, the types of sensitive information they can include about users, and the way they are used by data-reliant organizations, inferences can result in invasions of privacy that violate interests outside of the ones suggested by most privacy scholarship to date. In fact, invasions of privacy involving inferences can result in immediate or near-immediate threats to the physical integrity, autonomy, civil rights, and even financial prospects of members of certain vulnerable populations.

The incident that introduces this Note is a prime example. The young woman whose pregnancy was involuntarily disclosed to her family by Target suffered an invasion of privacy as a result of an inference. Target's disclosure by implication of the young girl's pregnancy status, while not illegal, goes against established medical consensus regarding a minor's rights to not involve her parents in matters relating to her pregnancy.[144] The fact that she was robbed of the decision of when and how to inform her family of her pregnancy also represents the compromise of her autonomy, and depending on her family dynamic, her physical safety might also have been imperiled.

Social networking platforms like Facebook, which collect massive amounts of user data, are able to make inferences about users' sexual orientation, putting LGBTQ people's physical and emotional safety at risk.[145] Until recently, Facebook made it possible for advertisers to target users based on

143. *Id.* at 1943–45; Edward J. Bloustein, *Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser*, 39 N.Y.U.L. REV. 962, 963 (1964).

144. *See* AAP Committee on Adolescence, *The Adolescent's Right to Confidential Care When Considering Abortion*, 139(2) PEDIATRICS e20163861, 2 (2017) ("The American Medical Association, the Society for Adolescent Health and Medicine, the American Public Health Association, the American College of Obstetricians and Gynecologists, the AAP, and other health professional organizations have reached a consensus that a minor should not be compelled or required to involve her parents in her decision to obtain an abortion, although she should be encouraged to discuss the pregnancy with her parents and/or other responsible adults.").

145. *See* Carter Jernigan & Behram F.T. Mistree, *Gaydar: Facebook Friendships Expose Sexual Orientation*, 14(10) FIRST MONDAY, https://www.firstmonday.org/article/view/2611/2302 (last visited Jan. 14, 2021).

their sexual orientation.[146] In 2012, a British teen was kicked out of his home after his parents found out he was gay via Facebook.[147] The teen had not submitted any information to Facebook about his sexual orientation, nor had he joined any gay groups, yet claimed that Facebook had begun displaying gay-interest ads on his profiles, which were seen by his parents when he accidentally left his computer unattended.[148]

According to Human Rights Watch, there are at least seventy countries around the world where same-sex relations are criminalized, and nine where certain "forms of gender expression that target transgender and gender nonconforming people" are also crimes.[149] In these countries, LGBTQ people face punishments ranging from prison, to corporal punishment, to the death penalty.[150] In a jurisdiction where certain sexual or gender minorities are criminalized, a misplaced LGBTQ-related targeted ad or recommendation on a platform that has inferred a user's sexual orientation could present harms ranging in severity from that experienced by the British teenager to threats to the physical integrity and lives of LGBTQ people. According to Human Rights Watch, during the anti-gay purge perpetrated by local authorities in Chechnya, Russia, in 2017, one of the ways in which police would identify gay men to capture and torture was to search captured victims' cell phones, "looking for contacts of other men who might be gay."[151] If presence in an LGBTQ person's contact list is enough of a proxy for authorities in queerphobic regimes, inference-based relevant advertisements and recommendations in an individual's social media profile—like Instagram's "Suggested For You," which shows suggestions for "similar profiles" to a particular user's[152]—could be as well.

146. Alex Kantrowitz, *Facebook Has Blocked Ad Targeting By Sexual Orientation*, BUZZFEED NEWS (Mar. 21, 2018), https://www.buzzfeednews.com/article/alexkantrowitz/facebook-has-blocked-ad-targeting-by-sexual-orientation.

147. *See* Kenneth C. Werbin, Mark Lipton & Matthew J. Bowman, *The Contextual Integrity of the Closet: Privacy, Data Mining and Outing Facebook's Algorithmic Logics*, 2(1) QUEER STUD. IN MEDIA & POPULAR CULTURE 37 (2017).

148. *Id.*

149. *#Outlawed: "The Love That Dare Not Speak Its Name,"* HUMAN RIGHTS WATCH, http://internap.hrw.org/features/features/lgbt_laws/ (last visited Jan. 14, 2021).

150. *Id.*

151. *"They Have Long Arms and They Can Find Me:," Anti-Gay Purge by Local Authorities in Russia's Chechen Republic*, HUMAN RIGHTS WATCH (May 26, 2017), https://www.hrw.org/report/2017/05/26/they-have-long-arms-and-they-can-find-me/anti-gay-purge-local-authorities-russias.

152. *See People Are Getting Suggestions to Follow Other People After They Follow Me on Instagram. How Do I Turn This Off?*, INSTAGRAM, https://help.instagram.com/530450580417848 (last visited Apr. 18, 2021).

At a minimum, the potential for such targeted content on the internet adds a persistent burden to the online activities of LGBTQ users, in that they must either always attempt to monitor what information certain platforms know about them, or they must keep from accessing LGBTQ-related content on the internet entirely for fear that platforms will infer their sexual orientation. LGBTQ people have historically depended on the relative anonymity that the internet provides in order to access important educational content about their sexual orientation or gender identities without fear of adverse consequences. The existence of inferences over which the user has little control can effectively prevent users from extracting value from the wealth of information online.

In some instances, the use of inferences by certain types of organization can be illegal. In her landmark study *Discrimination in Online Ad Delivery*, Latanya Sweeney exhibits as part of her problem statement how Google Images has learned to associate, perhaps from user data it has collected or acquired, certain proper names with certain races.[153] Image searches for "Latanya" and "Latisha" return results for Black women, while searches for "Kristen" and "Jill" return images of white women.[154] Google Images's inference-making is not illegal, but a problem arises when platforms that collect large amounts of user information, armed with knowledge such as Google Images's, can make inferences about users' race and other traits and include them in profiles about the users. These profiles may be shared with data brokers who will themselves share them with business who use these profiles to target advertisements to the user. An investigation by ProPublica showed that Facebook allowed advertisers to exclude users based on their race or gender, even in cases of ads where exclusion from viewing could be illegal, like housing ads.[155] A credit card company hoping to advertise a new product to a specific audience could find, on Facebook, the ability to exclude users based on race, a practice that in the United States is outlawed by the Equal Credit Opportunity Act (ECOA).[156]

153. Latanya Sweeney, Discrimination in Online Ad Delivery, DATA PRIVACY LAB (2013), http://dataprivacylab.org/projects/onlineads/1071-1.pdf.

154. *Id.*

155. Terry Parris, Jr. & Julia Angwin, *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.

156. 15 U.S.C. § 1691(a)(1); *see also* CONSUMER FINANCIAL PROTECTION BUREAU, EQUAL CREDIT OPPORTUNITY ACT (ECOA) EXAMINATION PROCEDURES (2015), https://files.consumerfinance.gov/f/documents/201510_cfpb_ecoa-narrative-and-procedures.pdf (". . .a creditor may not advertise its credit services and practices in ways that would tend to encourage some types of borrowers and discourage others on a prohibited basis. In addition, a creditor may not use prescreening tactics likely to discourage potential applicants on a prohibited basis.").

Users, however, have no way of knowing what ads are being hidden from them due to certain traits, and as such are unaware of these violations to their civil rights. Credit-lending is not the only context in which such exclusionary targeting is possible; the practice is possible in the housing and employment contexts as well. The threat to a users' financial status and prospects that these practices represent amounts to digital redlining, a term that hearkens back to the sort of practices legislation like ECOA were passed to prevent. Christopher Gilliard defined digital redlining as "the creation and maintenance of technology practices that further entrench discriminatory practices against already marginalized groups."[157] As more aspects of everyday life, especially the management of personal finances, are handled online, digital redlining can perpetuate and in many cases worsen economic inequality, especially as it affects people of color and other historically disadvantaged communities.

These threats made possible by the combination of inference creation and user targeting represent types of harms that not only implicate new types of interests but could also disproportionately affect marginalized groups and some of society's most vulnerable populations. New types of data that can result in new types of harm justify unprecedented levels of protection for such data, bolstering the argument for strong user control over inferences created about them. If users are to be granted stronger control over their inferences, both the European and California data privacy regimes ought to be amended to fully incorporate inferences into the definition of sensitive personal data so as to trigger the strongest protections afforded by both statutes, as previously mentioned.

V. PROPOSALS FOR THE STRONG PROTECTION OF INFERENCES MAY STILL LEAVE SOME ISSUES UNRESOLVED

In this Note, I endeavored to lay a foundation for how to think about the issue of inferences while keeping in mind the sheer importance, power, and value of this type of data in current times. The muddled state of the data protection of inferences, however, still raises a host of unresolved questions that could adversely impact any attempt at stronger protection of inferences, including the approach I propose in this Note.

Firstly, data-reliant businesses could lobby against the stronger protection of inferences by arguing that the inferences are created by using technologies

157. *Banking on Your Data: the Role of Big Data in Financial Services: Hearing before the Comm. on Financial Services Task Force on Financial Technology*, 116th Cong. 4 (2019) (statement of Dr. Christopher Gilliard, PhD, Professor of English, Macomb Community College and Digital Pedagogy Lab Advisor).

the controllers themselves have developed to serve their own needs, and which these businesses consider to be trade secrets. Their argument would be, essentially, that inferences are created by the sweat of the businesses' brows, and, as such, any rights granted to the user over inferences—especially the right to access—risk disclosure of the trade secret and a threat to the business's ability to profit off of it. Sonia Katyal has written on the perils of such reasoning, contending that trade secrecy in the context of consumer technology is closely intertwined with civil rights, and arguing that transparency should be incorporated into trade secrecy law in recognition of the modern threats algorithms can represent for individual's civil rights.[158] Katyal puts forth as a remedy to this issue the federal whistleblowing protections in the Defend Trade Secret Act (DTSA) of 2016, by which whistleblowers who disclose source code they suspect leads to biased decision-making are protected, and so is the trade secret while the claims are investigated.[159] However, depending on whistleblowers to disclose issues related to the handling of sensitive personal data is an inefficient and insufficient remedy given the scale at which data collection and inference creation are occurring. Katyal remarks as such in the context of trade secrets by making reference to such a solution's administrative costs as well as the difficulty of initial detection of malfeasance.[160]

Secondly, data-reliant businesses might argue that, since inferences are information about a user that the controller creates (instead of collecting it), preventing the creation of inferences or limiting their use restricts businesses' speech and violates their First Amendment rights. Legal challenges based on this argument could hinge on whether it can be determined that protecting the privacy of data subjects by granting them rights over inferences is a substantial state interest. In *Sorrell v. IMS Health*, the Supreme Court held that restricting the use of doctors' personal information to target marketing at them was a content-based restriction, and, since it did not advance a substantial state interest, it violated the speaker's First Amendment rights.[161] Additionally, in *Central Hudson Gas & Electric v. Public Service Commission*, the Court instituted a test for whether restrictions on commercial speech violate the First Amendment.[162] The *Central Hudson* test states that if (1) the speech at issue relates to lawful activity and is not misleading, (2) the government interest is

158. Sonia Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 120 (2019).

159. *Id.* at 130.

160. *Id.* at 140.

161. Sorrell v. IMS Health Inc., 131 S. Ct. 2653 (2011).

162. Central Hudson Gas & Electric Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 557 (1980).

substantial, (3) regulation of the speech directly advances the government interest, and (4) the regulation is "no more extensive than necessary," the restriction is lawful.[163] In the case of inferences, granting users the right to prevent processing of their sensitive personal information for ad-targeting purposes might constitute a content-based restriction on businesses' commercial speech. Provided that the inferences relate to lawful activities and do not contain misleading or inaccurate facts about users, the constitutionality of any stronger protections will hinge on courts. Judges will have to decide, hopefully with access to the academic literature on inferences, whether users having control over inferences and their use is a substantial government interest, and whether and to what extent restrictions on data-reliant businesses' use of inferences advance such an interest.

Lastly, data-reliant organizations might object to the classification of inferences as personal data, which would trigger stronger protections under both the GDPR[164] and the California data privacy frameworks.[165] Since the inferences that these organizations create are often predictions about the likelihood of a fact being true at present,[166] they can argue that any one set of inferences or profile they create are not about a specific user, but rather a hypothetical individual whose data points and created inferences are very similar to the user's, and therefore the user has no rights over the information. This argument could be weakened by two factors. On the one hand, in making this argument these organizations might create more issues and legal exposure for themselves, given that the argument implies decisions of all levels of importance are being made about specific users on the basis of potentially inaccurate data. Additionally, even if these data-reliant organizations are not a hundred percent certain of the inferences they draw unless they seek confirmation from the user, the inference is often made on the basis of data collected from or about a user. Even if any identifiers are removed from both collected data and inferences—in a process called de-identification or anonymization—contemporary data sets can include inferences that are so numerous, unique, detailed, or sophisticated that they run a high risk of being re-identified.[167] The risk of re-identification is significant, and it increases as

163. *Id.*

164. *See* GDPR, *supra* note 16.

165. *See* CAL. CIV. CODE § 1798.100.

166. *See supra* discussion in Part II.

167. *See generally* Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye, *Estimating the Success of Reidentifications in Incomplete Datasets Using Generative Models*, 10:3069 NATURE COMMUNICATIONS 1 (2019) ("…we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes. Our results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for

these organizations gain access to more and different kinds of data points.[168] In the age of Big Data, different types of organizations collect or derive countless different or overlapping types of data points depending on the information that might be of importance to them, making users vulnerable to re-identification attacks, and as such underscoring the need for the protection of inferences.[169]

These issues are only three out of numerous possible ones, and they are likely exponentially more complex than this space allows. My hope is that each will garner enough attention from data protection scholars to merit further in-depth study. That is also my wish for the issue of inferences at large.

VI. CONCLUSION

The current treatment of inferences under European and Californian data protection statutes is confusing. This uncertainty harms all users, but especially historically vulnerable populations.

Everything from the collection of the data that drives inferences, to their creation, to their use is shrouded in a veil of secrecy and nebulousness. Nonetheless, inferences are the same type of data that the field has historically concerned itself with the most, only collected, organized, and handled in unprecedented ways that can bring about new kinds of harms. As daunting as the task of addressing this issue seems, legislators, policymakers, and regulators must strive to demystify and address the issue if they are strongly committed to the protection of the most sensitive types of personal data and, by extension, the most vulnerable people in our increasingly digital societies and economies.

anonymization set forth by GDPR and seriously challenge the technical and legal adequacy of the de-identification release-and-forget model.").

168. *See id.* at 2 ("With population uniqueness increasing fast with the number of attributes available, our results show that the likelihood of a re-identification to be correct, even in a heavily sample dataset, can be accurately estimated, and is often high.").

169. Khaled El Emam, Elizabeth Jonker, Luk Arbuckle & Bradley Malin, *A Systematic Review of Re-Identification Attacks on Health Data*, 6(12) PLOS ONE 1, 2 (2011).