What Makes Data Personal?

Maria Lilla Montagnani[†][*] *& Mark Verstraete*[**]

Personal data is an essential concept for information privacy law. Privacy's boundaries are set by personal data: for a privacy violation to occur, personal data must be involved. Furthermore, an individual's right to control information extends only to personal data. However, current theorizing about personal data is woefully incomplete. In light of this incompleteness, this Article offers a new conceptual approach to personal data. To start, this Article argues that personal data is simply a legal construct that describes the set of information that an individual should be able to control, or circumstances where an individual should be able to exercise such control.

After displacing the mythology about the naturalness of personal data, this Article fashions a new theory of personal data that more adequately tracks when a person should be able to control specific information. Current approaches to personal data rightly examine the relationship between a person and information; however, they misunderstand what relationship is necessary for legitimate control interests. Against the conventional view, this Article suggests that how the information is used is an indispensable part of the analysis of the relationship between a person and data that determines whether the data should be considered personal. In doing so, it employs the philosophical

[*] Associate Professor of Commercial Law, Bocconi University; Director of the LL.M. in European Business and Social Law, Bocconi Law School; Transatlantic Technology Law Forum Fellow, Stanford Law School.

[**] Research Fellow, UCLA Institute for Technology, Law, and Policy. J.D. Harvard Law School, B.A. University of Georgia. We thank BJ Ard, Derek Bambauer, Jane Bambauer, Elettra Bietti, Sebastian Benthall, Maurizio Borghi, Federico Ferretti, Brett Frischmann, Kat Geddes, Tomer Kenneth, Michael Karanicolas, Aniket Kesari, Gabe Nicholas, Helen Nissenbaum, Ira Rubinstein, Katherine Strandburg, Thomas Streinz, Peter Swire, Salome Viljoen, John Villasenor, Tal Zarsky, and Stav Zeitouni for helpful advice and support. Thank you to participants of the Privacy Research Group at NYU and UCLA Power and Accountability in Tech Conference where this work was presented.

concept of separability as a method for making determinations about which uses of information are connected to a person and, therefore, should trigger individual privacy protections, and which are not.

This framework offers a superior foundation to extant theories for capturing the existence and scope of individual interests in data. By doing so, it provides an indispensable contribution for crafting an ideal regime of information governance. Separability enables privacy and data protection laws to better identify when a person's interests are at stake. And further, separability offers a resilient normative foundation for personal data that grounds control interests in a philosophical foundation of autonomy and dignity values — which are incorrectly calibrated in existing theories of personal data. Finally, this Article's reimagination of personal data will allow privacy and data protection laws to more effectively combat modern privacy harms such as manipulation and inferences.

TABLE OF CONTENTS

***

INTRODUCTION

Personal data is an essential but deeply contested concept within information law. Delineating the boundaries of personal data is not merely a theoretical exercise but instead is a foundational task for any functional regime of information privacy law.[1] At an implementation level, privacy and data protection statutes depend significantly on an account of personal data to make key normative distinctions — the determination of whether information is personal data distinguishes violations that create liability from innocent disclosures of non-personal information.[2] Further, the set of control rights that privacy and data protection statutes provide are limited to only personal data.[3]

Because personal data is pivotal to the success of information privacy regimes, one would expect current privacy and data protection frameworks to encompass clear definitions of personal data mirroring the dignity and autonomy interests at stake.[4] Unfortunately, this is currently not the case. At present, personal data is inadequately theorized in both the United States and Europe, causing existing accounts of personal data to come untethered from the concerns that information governance laws are intended to remedy. Worse still, this disconnect causes these laws to misfire and provide rights over information in cases where they are not warranted. In addition, this disconnect leads existing information governance regimes to fail to protect against modern privacy harms, such as manipulation.[5]

[1] *See* Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1817 (2011) [hereinafter *The PII Problem*] (arguing for a new approach to personal data in the United States that includes information on either "identified" or "identifiable" individuals).

[2] *Id.* at 1816.

[3] *Id.* For the European approach, see Nadezhda Purtova, *The Law of Everything. Broad Concept of Personal Data and the Future of EU Data Protection Law*, 10 L., INNOVATION & TECH. 40, 43-45 (2018).

[4] *See* Bert-Jaap Koops, Bryce Clayton Newell, Tjerk Timan, Ivan Škorvánek, Tomislav Chokrevski & Maša Galič, *A Typology of Privacy*, 38 U. PA. J. INT'L L. 483, 510 (2017) (discussing a broad overview of the types of privacy that constitutional law seeks to protect in several jurisdictions).

[5] Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 2-4 (2019).

Against this backdrop, this Article offers a new theory of personal data that more faithfully tracks the dignitary and autonomy interests imperiled by the information economy.[6] In particular, it rethinks the relationship between people and information, which defines the conditions under which information should be considered personal data and, therefore, governed by individual rights of control. Instead of simply defaulting to analysis of the semantic relationship — or rather, whether information is *about* a person — this Article introduces *separability* as the primary means to evaluate whether the relationship between a person and information generates credible claims for individual control.

Separability describes the relationship that exists between a person and information, enabling identification of when information is connected (inseparable) to the person and warrants control by that data subject.[7] Conversely, separability also identifies when information is disconnected (separable) from the data subject and rationales for control falter.[8]

This analysis is inherently normative. Different uses of personal data have different moral valences, and privacy and data protection laws should be sensitive to these distinctions. Ultimately, uses of information that are inseparable risk using the data subject as a means to an end, undermining their dignity and autonomy. Control rights over inseparable uses create a buffer against these harms. As a result, information governance frameworks driven by separability are better equipped to protect individual dignity and autonomy in the digital age while also preserving the benefits that come from uses of information that are not meaningfully connected to any specific person.

This Article introduces a conceptual test for separability that examines both connection and use. For information to be inseparable and thus subject to individual control rights, it must have an *ex ante* connection to a person, including semantic connections, such as when information is

[6] *See* Alicia Solow Neiderman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 359-61 (2022); *see also* Tal Z. Zarsky, *Privacy and Manipulation in the Digital Age*, 20 THEORETICAL INQUIRES L. 157, 161-68 (2019) (discussing the limits of existing privacy and data protection paradigms in addressing the issue of consumer manipulation).

[7] Mark Verstraete, *Inseparable Uses*, 99 N.C. L. REV. 427, 430-32 (2021).

[8] *Id.*

about a person.[9] In addition, the information must be used in a way that depends on this connection to affect the person. Separability parts company from traditional conceptions of personal data that simply require that the information be about a person.[10] It requires two jointly sufficient conditions: connection and use.[11] When these two conditions are met, information should be considered personal data and subject to a robust suite of individual control rights.

To illustrate the centrality of use, consider two different uses of the same information. Information such as medical data can be used for research purposes or system optimization in ways that do not affect the person described by the data. Conversely, the same information can be used to infer new information about a person, ultimately enriching their user profile in order to influence their purchasing habits.[12] The underlying stakes of these two uses are markedly different. Therefore, the data subject's interest in controlling these uses is also quite disparate. The data subject has a stronger claim to regulate uses of information that can condition their choices and thus their autonomy.[13] At its core, separability provides robust criteria to make similar distinctions and helps resolve several contentious issues in information privacy law, including whether inferences are personal data and, by extension, should be governed by individual rights of control.

Grounding personal data in a theory of separability breaks new conceptual and normative ground in the debate over when control rights over data are warranted and casts new light on the ideal shape of information governance regimes. Because dignity and autonomy interests primarily attach when information is used — rather than simply collected or processed — separability fully captures individual interests over data. Consequently, an account of personal data based in separability marks an improvement over existing accounts of personal

[9] *Id.* at 471-76.

[10] *See* Jane Yakowitz Bambauer, *The New Intrusion*, 88 NOTRE DAME L. REV. 205, 213-29 (2012) (claiming that the existing American and European privacy and data protection frameworks provide rights of control to individuals over data that describes them).

[11] Verstraete, *supra* note 7, at 452-53.

[12] *See* Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 387 (2012).

[13] *See* Verstraete, *supra* note 7, at 435-39.

data in both the United States and Europe that largely fail to consider information uses.[14]

The rest of this Article unfolds as follows. Part I examines several fundamental aspects of the current debate over personal data. In particular, this Part maps competing approaches to personal data in the European Union ("EU") and the United States. It argues that both European and American theories of personal data embody incomplete views of privacy and are a poor fit for modern privacy and data protection harms.

In Europe, personal data, as embodied in the General Data Protection Regulation ("GDPR"),[15] fashions an incredibly broad notion of personal data that grants control rights in situations where they are not justified under privacy theories such as informational self-determination.[16] Conversely, the American approach to personal data, or personally identifiable information ("PII"), prioritizes identification, which embodies an outdated view of privacy as secrecy, limiting protections to disclosures rather than other potential harms that track autonomy concerns.[17]

Part II examines several different relationships between a person and data that are used to delineate personal data. It begins with an analysis of the triad of relations that each suffice for creating personal data under

[14] For a survey of the existing theories of privacy and data protection in Europe, see Plixavra Vogiatzoglou & Peggy Valcke, *Two Decades of Article 8 CFR: A Critical Exploration of the Fundamental Right to Personal Data Protection in EU Law*, *in* RESEARCH HANDBOOK ON EU DATA PROTECTION 11, 11 (Eleni Kosta, Ronald Leenes & Irene Kamara eds., 2022). For a broad overview of the range of theories that underlie American privacy law, see Daniel J. Solove, *Conceptualizing Privacy*, 90 CALIF. L. REV. 1087, 1099-126 (2002).

[15] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR].

[16] *See* Bart van der Sloot, *Legal Fundamentalism: Is Data Protection Really a Fundamental Right?*, *in* DATA PROTECTION AND PRIVACY: (IN)VISIBILITIES AND INFRASTRUCTURES 3, 8-9 (Ronald Leenes, Rosamunde van Brakel, Serge Gutwirth & Paul De Hert eds., L., Governance & Tech. Ser. No. 36, 2017); Purtova, *supra* note 3, at 43; Inge Graef, Raphaël Gellert & Martin Husovec, *Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data Is Counterproductive to Data Innovation* 5 (Tillburg L. & Econ. Ctr., Discussion Paper No. 028, 2018), https://ssrn.com/abstract=3256189 [https://perma.cc/2LJB-BFXT].

[17] *See* Schwartz & Solove, *The PII Problem*, *supra* note 1, at 1836-41.

the EU data protection framework: content, purpose, and result.[18] This Article shows that each of these relationships falters as a sufficient justification for personal data. In addition, this Part examines propertarian approaches to personal data that recognize a property-style relationship between a person and information.[19] This Part contends that property relationships overlook more traditional privacy concerns and prove a poor fit for determining when data is personal.

Part III offers a new theory of personal data grounded in separability. This Part demonstrates that separability provides a better approach to identify what makes data personal. This is largely because separability provides tools for an improved analysis of when the relationship between a person and data should implicate individual rights of control.[20] Moreover, this Part details the mechanics of when information is separable or not. To constitute personal data, information must be connected to a specific person and its use must rely on that connection to affect the person. To sharpen this concept, this Part sketches different combinations of connection and use to demonstrate when data is personal.

Part IV applies the theory of a separability-inspired personal data regime to resolve several open questions about control over inferred data in the information society.[21] Separability can chart a course that protects individuals against autonomy and dignity harms flowing from some uses of information while also preserving beneficial uses that do not affect the individual.

[18] *Article 29 Data Protection Working Party Opinion 04/2007 on the Concept of Personal Data*, at 10 (June 10, 2007), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf [https://perma.cc/W6Y4-DM39] [hereinafter *Art. 29 WP Opinion Concept of Personal Data*].

[19] *See* Gianclaudio Malgieri, *Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data*, 2016 PinG 133, 138-39 (suggesting that control over data should be scaled according to the amount of labor used to create it).

[20] Verstraete, *supra* note 7, at 435-39 (discussing the normative foundations of separability).

[21] *See* Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 Colum. Bus. L. Rev. 494, 498-99 (discussing the limits of the GDPR to provide control over inferred data).

I. THE STAKES OF PERSONAL DATA

This Part examines the relationship between personal data and the existing theories of privacy and data protection. In doing so, this Part makes several claims about the definitional stakes of personal data. First, personal data is not a natural concept. Instead, there are inherent policy choices embedded in choosing a definition of personal data. After demystifying the supposed naturalness of personal data, this Part canvases the existing European and American definitions of personal data and suggests that these definitions of personal data fail to mirror complete theories of information privacy.

A. *Personal Data Is Not a Natural Concept*

Personal data is not a natural concept. That is, there is nothing out in the world that is inherently personal data and, as a result, should be governed by a specific set of legal rules. Instead, personal data is necessarily a policy choice that is determined by a host of social, political, and ethical considerations. The circumstances where people ought to have control over information cannot be decided from some abstract concept of personal data, but must be derived from legal categories that should be informed by a broader normative vision of privacy and data protection law.

The idea that some legal concepts do not have a natural existence beyond their invocation in the law is not new. The American legal realists demonstrated that the idea of natural concepts in law is both inherently circular and impossibly vague.[22] More specifically, legal concepts are necessarily creatures of law so purely referring to their definition is circular.[23] Similarly, the legal realists critiqued the idea of "mechanical jurisprudence" which sought to derive legal conclusions by deductions based on abstract legal concepts.[24] Alongside the issue of circularity, the

[22] Felix S. Cohen, *Transcendental Nonsense and the Functional Approach*, 2 ETC: A REV. GEN. SEMANTICS 82, 91 (1945).

[23] *Id.*

[24] *See* Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 HARV. L. REV. 1685, 1710 (1976) (discussing the qualities of "classical legal thought" which attempted to deduce legal outcomes from abstract legal concepts); Joseph William Singer, *Legal Realism Now*, 76 CALIF. L. REV. 465, 497-98 (1988) (discussing formalist jurisprudence that sought to deduce outcomes from abstract legal concepts).

outcomes that follow from legal concepts are indeterminate.[25] In this grey space of indeterminacy, judges simply perform a policy analysis (or appeal to values outside of pure definitional deduction).

However, both academic and public conversation underestimate the extent to which personal data is inherently a policy choice about which data should be protected under privacy laws. The conventional view is that data that describes a person is necessarily connected to them in a significant way, which should generate legally recognizable interests over the data.[26] In the same way, both proposed and enacted legislation casually discuss data that describes a person in terms of property (e.g., "their" data) — moving quickly from a simple fact about a piece of information to a specific legal relation.[27]

In the next Section, we begin unpacking different definitions of personal data and assess what views of privacy and data protection they embody. Ultimately, we argue that current definitions of personal data fail to reflect a desirable and complete view of informational privacy. This discussion sets the stage for an introduction of our view of personal data that better maps to current privacy and data threats in the information economy.

B. *Personal Data and Theories of Information Governance*

This Section introduces two different conceptions of personal data — one European and one American — and offers a conceptual and normative critique of these definitions. Most importantly, this discussion shows that these definitions of personal data do not fully track the concerns of contemporary privacy and data protection scholars.

[25] JEROME FRANK, LAW AND THE MODERN MIND 128 (1930) (offering a "strong version" of the underdetermination theory in legal realism); *see* Brian Leiter, *Rethinking Legal Realism: Toward a Naturalized Jurisprudence*, 76 TEX. L. REV. 267, 295-97, 301 (1997) (discussing different possible legal realist positions about indeterminacy in law).

[26] *See* U.S. DEP'T OF HEALTH, EDUC., & WELFARE, ADVISORY COMM. ON AUTOMATED PERS. DATA SYS., NO. 73-94, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS 41-42 (1973) (introducing the concept of Fair Information Practice Principles ("FIPPs") which provide individuals with information rights over data that describes them); Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2056, 2058 (2004).

[27] *See* Bambauer, *supra* note 10, at 213-29. Moreover, Bambauer claims that legislative reforms in the U.S. and Europe provide rights of control to individuals over data that describes them simply because they are described in the data. *Id.* at 215.

Here, we focus explicitly on the underlying theories of privacy and data protection that seem to follow from European and American approaches to personal data conceptualizations.[28] The European approach, as embodied in the GDPR, provides a broad conception of personal data.[29] This expansive definition of personal data captures a dizzying array of data, provides control over information that does not raise salient privacy concerns, and often allows data subjects to block socially beneficial uses of information that do not implicate their interests.

By contrast, the American focus on PII only captures privacy harms that are linked to identification — or rather, situations where information reveals facts about a person. This exclusive focus on identification fails to consider the full spectrum of potential privacy harms largely because there are a host of "modern" privacy concerns around manipulation and autonomy that are not captured by reducing personal data to data that identifies the data subject.

1. The European Approach to Data Protection: The GDPR

The GDPR is a comprehensive piece of European legislation that sets the groundwork for data protection law in Europe.[30] Prior to the GDPR, data protection law pivoted almost exclusively around the notion of consent, requiring it for almost all processing.[31] The GDPR, instead,

[28] It is often difficult to talk about a unified American approach to personal data. Although American statutes seem to focus primarily on personally identifiable information ("PII"), the way in which PII is defined varies across different statutes. However, we believe that even though PII is implemented in different ways, PII as a concept is still coherent and drives the vision of American privacy law. That is, American privacy law is often focused on identification and many debates about the boundaries of personal data are debates about whether an individual is identified or not.

[29] Orla Lynskey, *Deconstructing Data Protection: The 'Added-Value' of a Right to Data Protection in the EU Legal Order*, 63 INT'L & COMPAR. L.Q. 569, 582 (2014); Purtova, *supra* note 3, at 43; Paul M. Schwartz & Daniel J. Solove, *Reconciling Personal Information in the United States and European Union*, 102 CALIF. L. REV. 877, 892 (2014) [hereinafter *Reconciling Personal Information*]; van der Sloot, *supra* note 16, at 8.

[30] GDPR, *supra* note 15.

[31] *See* ELENI KOSTA, CONSENT IN EUROPEAN DATA PROTECTION LAW 88 (2013). On the changed role of consent under the GDPR, see also Eur. Data Prot. Bd., *Guidelines 05/2020 on Consent Under Regulation 2016/679* (May 4, 2020) [hereinafter EDPB], https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf [https://perma.cc/Z25P-GMUB].

embodies a slightly different vision of control. This revised vision of control introduces a governance regime that provides a set of rights for data subjects and principles for data processing that more closely track the European idea of data protection as a fundamental right.[32]

a. The Concept of Personal Data in the GDPR

European data protection law rises and falls with personal data.[33] This is because the GDPR's rights, obligations, and protections only apply to personal data.[34] Moreover, the broad scope of European data protection law is a product of the fact that the GDPR provides a sweeping definition of personal data, which includes "any information relating to an identified or identifiable natural person."[35] Many commentators have been quick to point out that the GDPR invokes an incredibly broad notion of personal data.[36] Nadezhda Purtova, however, offers the slightly more controversial claim that the definition of personal data under the GDPR is so broad that almost any piece of information could be considered personal data.[37]

According to Purtova, the constituent parts of the GDPR's definition of personal data have been interpreted expansively, which, therefore, create a broad definition.[38] Yet the sweeping interpretations of the concepts underlying personal data began before the introduction of the GDPR. For example, the requirement that personal data relate to an

[32] *See* Chris Jay Hoofnagle, Bart van der Sloot & Frederik Zuiderveen Borgesius, *The European Union General Data Protection Regulation: What It Is and What It Means*, 28 INFO. & COMMC'NS TECH. L. 65, 66 (2019).

[33] *See* Purtova, *supra* note 3, at 43; *see also* Council Directive 95/46, art. 3(1), 1995 O.J. (L 281) 31 (EC) [hereinafter Directive 95/46]; GDPR, *supra* note 15, art. 2(1).

[34] Purtova, *supra* note 3, at 43.

[35] GDPR, *supra* note 15, art. 4(1). The Article is substantially unchanged with respect to the Directive 95/46, as it merely adds location and genetic data to what was already listed in Article 2 of the Directive 95/46. Yet the CJEU's jurisprudence has over time broadened the interpretation of the GDPR's notion of personal data. *See* Purtova, *supra* note 3, at 43.

[36] *See, e.g.*, Purtova, *supra* note 3, at 41-43 (arguing that the GDPR "is growing so broad that the good intentions to provide the most complete protection possible are likely to backfire"); Schwartz & Solove, *Reconciling Personal Information*, *supra* note 29, at 887 (discussing the breadth of the EU approach to privacy regulation).

[37] Purtova, *supra* note 3, at 41.

[38] *Id.*

"identified or identifiable" person had been construed broadly by the Article 29 Working Party ("Art. 29 WP")[39] and then expanded further by the Court of Justice of the European Union ("CJEU").[40] However, the GDPR cemented the expansiveness of the "identifiability" requirement by considering the technological possibility of re-identification — the process to re-establish the relationship between data and the subject to which the data refers — rather than the subjective ability for the data controller to re-identify.[41] Put more concretely, even when the data

[39] The Article 29 Data Protection Working Party ("Art. 29 WP") was an independent European body that dealt with issues relating to the protection of privacy and personal data. The Working Party's main goal was to provide expert advice to member States regarding data protection. The Working Party drafted recommendations about the implementation of data protection laws that are often cited in debates about the scope and content of data protection law in the EU. Since the adoption of the GDPR, the Art. 29 WP has been replaced by the European Data Protection Supervisor ("EDPS"), but the EDPS retains the same overall mission as the WP. *See Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 6.

[40] *See* Case C-434/16, Peter Nowak v. Data Prot. Comm'r, ECLI:EU:C:2017:994, ¶ 34 (Dec. 20, 2017); Case C-582/14, Patrick Breyer v. Bundesrepublik Deutschland, ECLI:EU:C:2016:779, ¶ 49 (Oct. 19, 2016). For further comment on the cases, see Frederick Zuiderveen Borgesius, *The Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition*, 3 EUR. DATA PROT. L. REV. 130, 131 (2017); Karolina Podstawa, Peter Nowak v Data Protection Commissioner*: You Can Access Your Exam Script, Because It Is Personal Data*, 4 EUR. DATA PROT. L. REV. 252, 254 (2018).

The CJEU is the highest authority on the interpretation of EU law. National judges in the EU can, and in some cases must, ask the CJEU how to interpret EU rules through the referral mechanism (Consolidated Version of the Treaty on European Union art. 19(3)(b), Jan. 3, 2020, 2020 O.J. (C 202) 27). The decisions above are the result of referrals to the CJEU.

[41] Purtova, *supra* note 3, at 44; *see also* GDPR, *supra* note 15, recital 26 ("To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.").

The upshot of Recital 26 is that personal data is a functional and malleable concept. Moreover, information may not meet the threshold for identifiability at one point in time because there are no existing technological measures to allow reidentification. However, once these measures are invented, the same information may be considered personal data. *See* Purtova, *supra* note 3, at 44; *see also* Schwartz & Solove, *The PII Problem*, *supra* controller has no interest or ability to re-identify information, it will still be personal data if there are technical measures that allow re-identification.[42]

Just as "identifiability" has been interpreted broadly, so too has the "relating to" requirement. In particular, the Art. 29 Working Party and CJEU offer multiple sufficient relationships that may satisfy the condition that data must *relate* to a person in order to be personal data.[43] Personal data may be related to a person through *content*, *purpose*, or *result*.[44] And further, any of these relationships satisfy the "relating to" requirement for personal data.[45]

Starting with *content* relationships, information related to a person through its content tracks the most intuitive idea about how information may be related to a person; that is, the information is about a person or describes them in some way. For example, a patient's medical diagnosis is related to them through its content because the diagnosis is about them and describes them. However, information that is about an identifiable individual is also considered personal data under the GDPR.[46] Rather than being about a person, identifiable information might refer to identifiers (such as an IP address) that can be combined with other data to determine a person's identity.[47]

Moreover, information can be related to a person through either its *purpose* or *result* — rather than its content — and still be considered

note 1, at 1818 (recognizing that the distinctions between non-identifiable, identifiable, and identified are often a matter of context and the current state of technology).

[42] *See* Graef et al., *supra* note 16, at 5; *see also Opinion of the Article 29 Data Protection Working Party on Anonymisation Techniques*, at 5, 9 (Apr. 10, 2014), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [https://perma.cc/UA68-3TDE].

[43] *See* Graef et al., *supra* note 16, at 5.

[44] *See Nowak*, Case C-434/16, ¶ 44; *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11.

[45] *See Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11.

[46] *See* GDPR, *supra* note 15, art. 4(1).

[47] *See* Schwartz & Solove, *Reconciling Personal Information*, *supra* note 29, at 905-08 (describing the difference between identified and identifiable data and the distinction attendant consequences for privacy law).

personal data.[48] Unlike when content connects a person to information, purpose and result are not characteristics of information but, instead, are statements about how information can be used.[49] For instance, information is related to a person through its *purpose* when data is used or likely to be used to evaluate, treat, or influence the status or behavior of an individual.[50] Similarly, information is sufficiently related to a person through its *result* when its use affects a person's rights or interests even when the effect is minor.[51] This means that whether information is personal data is, at times, contextual and depends on *ex post* considerations rather than purely *ex ante* ones.

This analysis follows from CJEU jurisprudence that has set the conceptual boundaries of personal data.[52] The result is a conception of personal data that goes well beyond information that *says* something about a natural person to also encompass information that can be used to change the status or behavior of an individual or influence that person's rights or interests. The GDPR encompasses the CJEU's developing interpretation of personal data, thereby offering protection to information about an individual as well as information that affects an individual.[53]

b. The Stakes of the GDPR's Definition of Personal Data

This Subsection considers the relationship between the GDPR's definition of personal data and the broader normative stakes of this regulation. Moreover, this subsection argues that the GDPR's definition of personal data fails to fully track quintessential privacy values such as control over personal information and self-determination, as well as

[48] *See Nowak*, Case C-434/16, ¶ 34; *see also* Purtova, *supra* note 3, at 69-70 (discussing the extensive interpretation that over time the CJEU's jurisprudence operated of the notion of personal data).

[49] *See* Purtova, *supra* note 3, at 69-70.

[50] Graef et al., *supra* note 16, at 5.

[51] *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11 ("It should be noted that it is not necessary that the potential result be a major impact. It is sufficient if the individual may be treated differently from other persons as a result of the processing of such data.").

[52] *See Nowak*, Case C-434/16, ¶ 4; Case C-582/14, Patrick Breyer v. Bundesrepublik Deutschland, ECLI:EU:C:2016:779, ¶ 5 (Oct. 19, 2016).

[53] *See Nowak*, Case C-434/16, ¶ 34.

broader pronouncements about fundamental rights within data protection. More specifically, the GDPR's definition of personal data sweeps too broadly and grants rights over information in cases that do not implicate these larger normative goals.[54] Ultimately, this discussion sets the stage to introduce our revised definition of personal data that more adequately tracks the normative goals of dignity and autonomy, which typically drive information law.

(1) The Evolution of Data Protection as a Fundamental Right

Data protection regulation in Europe is an evolving process. Initially, it began with a clear, limited objective which was principally about fashioning a set of rules to allow individuals to control their personal information and, by extension, determine their online identities.[55] Over time, data protection slowly began to take on a more significant role within European law as it became a fundamental right and incorporated within the European Charter of Fundamental Rights ("ECFR").[56] The evolution of data protection culminated in the GDPR which states that its aim is to protect the "fundamental rights and freedoms" of people and, in particular, their "right to the protection of personal data."[57]

Within the EU there is little guidance about what makes data personal and what type of protection is necessary. Article 8 of the ECFR declares that "everyone has the right to the protection of personal data concerning him or her."[58] However, the ECFR fails to offer a rigorous analysis of the conditions under which information *concerns* a person.

[54] *See* Valentin M. Pfisterer, *The Right to Privacy — A Fundamental Right in Search of Its Identity: Uncovering the CJEU's Flawed Concept of the Right to Privacy*, 20 GER. L.J. 722, 733 (2019) (analyzing the CJEU's jurisprudence to show that a lack of a consistent concept of fundamental rights to privacy and to data protection results in a lower degree of certainty, reliability, and predictability).

[55] *See* van der Sloot, *supra* note 16, at 5.

[56] Charter of Fundamental Rights of the European Union, Oct. 26, 2012, 2012 O.J. (C 326) 391 [hereinafter ECFR]. The Charter became legally binding when the Treaty of Lisbon entered into force on December 1, 2009, as the Treaty confers on the Charter the same legal value as the Treaties. *See* Stefano Rodotà, *Data Protection as a Fundamental Right*, *in* REINVENTING DATA PROTECTION? 77, 77-82 (Serge Gutwirth, Yves Poullet, Paul De Hert, Cécile de Terwangne & Sjaak Nouwt eds., 2009).

[57] GDPR, *supra* note 15, art. 1(2).

[58] ECFR art. 8.

The conditions required for information to concern a person are ultimately a philosophical question that implicitly relies on foundational ideas about when a person's rights and interests are implicated. It is no surprise, then, that the lack of consideration for the relationship between people and information leads the GDPR to falter conceptually and fail to appropriately track an individual's interests over information that concerns them.

The ECFR, however, does offer a limited set of clues about when information concerns a person and should be considered personal data.[59] Yet it is obvious from the structure and evolution of the ECFR that personal data concerns individuals in ways that are conceptually distinct from traditional privacy interests. This is evident for two principal reasons. First, privacy — like data protection — is discussed within its own Article within the ECFR, which indicates that privacy and data protection are motivated by different sets of rights and interests.[60]

Second, the evolution of personal data protection in Europe represents a dramatic split from privacy. Initially, personal data was protected under privacy frameworks within Europe,[61] but the turn towards digitization and data processing pressured EU institutions to adopt a set of rules to protect mundane information — like zip codes or car ownership — that did not implicate traditional privacy values.[62] Though mundane information does not implicate traditional privacy interests, control over this information is thought to provide individuals some level of informational self-determination; or rather, the ability to decide when and how to disclose information about themselves.[63]

Taken together, though, European laws provide little guidance about what is personal data and when it should be protected. However, this paucity of guidance may be a product of data protection's status as a fundamental right.[64] In principle, the GDPR protects personal data as a

[59] *See id.*

[60] *See id.* arts. 7-8.

[61] van der Sloot, *supra* note 16, at 5-6.

[62] *See id.* at 5-7.

[63] On the concept of self-determination, see Florent Thouvenin, *Informational Self-Determination: A Convincing Rationale for Data Protection Law?*, 12 J. INTELL. PROP., INFO. TECH., & ELEC. COM. L. 246, 248 (2021).

[64] *See* Hoofnagle et al., *supra* note 32, at 69; Lynskey, *supra* note 29, at 569. On the difficulties of identifying the underlying interest of data protection, see also Paul De Hert &

means to safeguard fundamental values such as human dignity, autonomy, and other fundamental interests.[65] In this sense, the right of data protection is instrumental to the protection of fundamental rights[66] and, for this reason, the GDPR does not rigidly define the concept of personal data. Instead, the GDPR offers a flexible standard that can be revised over time to track fundamental values, such as dignity and autonomy.[67] The next two Subsections demonstrate that the notion of personal data has been stretched so wide that it often fails to properly track data subjects' dignity and autonomy interests.

(2) Control over Personal Data

Traditionally, personal data is protected through individual control that is expressed through the data subject's consent. Through exercising their consent, individuals decide what information to disclose and how this disclosed information may be used.[68] However, as data protection has evolved, the notion of control — as well as consent — has also evolved.[69] The information economy enables the collection and processing of a dizzying amount of information. In order to counterbalance the exponential growth of information collection and processing, the GDPR grants data subjects control over a larger amount of personal information, which increases the scope of data protection.[70] In addition, the GDPR provides a nuanced suite of control rights that are more granular than simple binary consent options.[71] These two trends

Serge Gutwirth, *Data Protection in the Case Law of Strasbourg and Luxemburg: Constitutionalisation in Action*, *in* REINVENTING DATA PROTECTION?, *supra* note 56, at 3, 3-5.

[65] *See* Vogiatzoglou & Valcke, *supra* note 14, at 34-35.

[66] *See* Antoinette Rouvroy & Yves Poullet, *The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy*, *in* REINVENTING DATA PROTECTION?, *supra* note 56, at 45, 46.

[67] *See* Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 242 (2013).

[68] Thouvenin, *supra* note 63, at 250 (pointing out that consent is a "straightforward implementation of informational self-determination").

[69] *See* Lynskey, *supra* note 29, at 594-95.

[70] *See* GDPR, *supra* note 15, art. 4(1).

[71] For the more holistic interpretation of consent, see Yordanka Ivanova, *The Role of the EU Fundamental Right to Data Protection in an Algorithmic and Big Data World*, *in* DATA

determine a shift from individual control over personal data (via consent) to control over the processing of personal data (via control rights).

The crux of control over personal data in data protection follows from the ECFR, which proposes that individuals have a right to protect personal information *concerning* them.[72] The GDPR fills in the gaps left by the ECFR by delineating the scope and rights of this guarantee.[73] As stated earlier, the GDPR only regulates personal data, which is defined as any information relating to an identified or identifiable person.[74] And further, the CJEU has interpreted the constituent pieces of this definition quite expansively, so that almost any information could be considered personal data under the right circumstances.[75]

At its core, the GDPR operationalizes the fundamental right of data protection as laid out in the ECFR.[76] However, the GDPR's account of personal data potentially undermines its ability to effectively execute the ECFR's vision of data protection as a fundamental right.[77] The ECFR explicitly mentions that individuals have a right to protect information concerning them.[78] However, the GDPR marks a break from the idea of "concerning" a person and moves, instead, towards the idea of "relating" to a person.[79] Information that relates to a person is a broader category than information that concerns a person. Put another way, information or the use of some information can be related to a person but not concern them in any significant way.[80]

PROTECTION AND PRIVACY: DATA PROTECTION AND ARTIFICIAL INTELLIGENCE 145, 145-80 (Dara Hallinan, Ronald Leenes & Paul De Hert eds., 2021).

[72] ECFR art. 8.

[73] *See* van der Sloot, *supra* note 16, at 10-11 (referring to the GDPR as the implementation of Article 8 of the ECFR).

[74] GDPR, *supra* note 15, art. 4(1).

[75] Purtova, *supra* note 3, at 66.

[76] van der Sloot, *supra* note 16, at 10-11.

[77] *See* Thouvenin, *supra* note 63, at 256 (affirming that because of the aims of "mitigating largely unknown and unspecific risks, data protection law often fails to protect individuals against the realisation of these risks").

[78] ECFR art. 8.

[79] GDPR, *supra* note 15, art. 4(1).

[80] Separability, however, more faithfully tracks when information *concerns* a person and, therefore, realigns personal data with its conceptual foundations in the ECFR. *See infra* Part III.B.

Moreover, the idea of information that concerns a person inherently invokes an idea of rights and interests (or issues of concern) that perform richer normative work than simply the requirement that information relate to a person performs.[81] The result here is that the GDPR dilutes the normative valence of data protection and expands its sphere of operation beyond the initial decrees of the ECFR.

Finally, the GDPR's focus on process rather than substance misunderstands the nature of data protection and privacy wrongs. The GDPR introduces procedural rules that regulate personal data processing.[82] However, the focus on procedures — rather than substantive determinations about what data should be protected — reduces data protection to merely procedural protection.[83] The core of data protection should promote substantive, instead of procedural, protections. This is because violations of procedural rights of data processing only raise salient concerns if the procedures are themselves meaningful. Furthermore, whether the procedures are meaningful depends on the substance of what these procedures protect. Yet the GDPR mostly avoids this vital question by failing to recognize the inherent stakes of the definition of personal data.

(3) Informational Self-Determination

The foundational justification for privacy and data protection in Europe is informational self-determination, which provides a right and opportunity for individuals to determine the conditions under which information about them is disclosed and used.[84] Traditionally,

[81] *Infra* Part III (describing separability as a foundation for when information sufficiently concerns a person such that they should be able to exercise control over it).

[82] Some commentators have suggested the procedural nature of the GDPR is intended to promote fair processing and increase transparency. *See, e.g.*, Lorenzo Dalla Corte, *A Right to a Rule: On the Substance and Essence of the Fundamental Right to Personal Data Protection*, *in* DATA PROTECTION AND PRIVACY: DATA PROTECTION AND DEMOCRACY 27, 27 (Dara Hallinan, Ronald Leenes, Serge Gutwirth & Paul De Hert eds., 2020) (arguing that data protection has evolved away from privacy into a procedural, auxiliary, *sui generis* fundamental right as a response to technological developments and to the growing importance of secondary data protection legislation).

[83] *See id.* at 42.

[84] Thouvenin, *supra* note 63, at 248 ("[I]nformational self-determination refers to every individual's right and opportunity to determine which information about him or

informational self-determination provides the fundamental justification for granting individuals rights of control over their personal data. Supporters of informational self-determination often position the concept as a pre-requisite for preserving individuals' dignitary and autonomy interests.[85] Furthermore, informational self-determination is sometimes viewed as providing other corollary benefits, such as offsetting the perniciousness of information asymmetries in the data economy.[86]

While the right to informational self-determination stakes out lofty ideals about the relationship between control over information and fundamental values (like human dignity and individual autonomy), the efficacy of informational self-determination is a product of how well this right is operationalized in privacy and data protection regulation. Put differently, whether individuals are able to practice informational self-determination is determined by the features of regulations that purport to provide this right.

Yet the GDPR's concept of personal data is an uneasy fit with informational self-determination for a few principal reasons. To start, personal data under the GDPR is an incredibly broad notion that seems to include information that embodies a weaker claim of control under a theory of informational self-determination. For example, individuals have stronger claims of control over sensitive information (such as medical history) than mundane information (such as country of residence); however, both medical history and country of residence are protected equally under the GDPR as they both are instances of personal data.[87] In other words, informational self-determination is more strongly

herself is disclosed to others and for what purposes such information may be used."); *see* Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy in Europe: Initial Data on Governance Choices and Corporate Practices*, 81 GEO. WASH. L. REV. 1529, 1539 (2013) (quoting ALAN F. WESTIN, PRIVACY AND FREEDOM 7 (1967)) (defining informational self-determination as "the claim of individuals . . . to determine for themselves when, how, and to what extent information about them is communicated to others").

[85] *See* Paul De Hert & Serge Gutwirth, *Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power*, *in* PRIVACY AND THE CRIMINAL LAW 61, 63-64 (Erik Claes, Antony Duff & Serge Gutwirth eds., 2006); Lynskey, *supra* note 29, at 589; Rouvroy & Poullet, *supra* note 66, at 51.

[86] Lynskey, *supra* note 29, at 592.

[87] *See* GDPR, *supra* note 15, art. 4(1) (providing a broad definition of personal data).

implicated in a subset of the information that is currently protected under the GDPR.

Second, informational self-determination — as it is currently theorized — should extend only to information that is *about* a person. This is because informational self-determination protects against unwanted disclosures of information that determine how data subjects are viewed.[88] To that end, violations of informational self-determination will follow from information that is about a person. However, data subjects are granted control rights over information in situations where data is not *about* them. For instance, the GDPR's definition of personal data extends to information that is related to a person through its purpose or result, rather than its content.[89] Thus, exercising control over this information does not square easily with informational self-determination.

In addition, the GDPR's definition of personal data employs a broad definition of identification, which includes data that does not identify a person but could potentially identify them.[90] Again, by including identifiable information within the ambit of personal data, the GDPR extends control to information that will not determine how a person is viewed — unless somehow the data subject becomes identified. Hence, merely identifiable information does not pose concrete risks to information self-determination because it does not affect how a person is viewed. For this reason, control justifications for identifiable information that are rooted soundly in self-determination are largely unconvincing.

2. The American Approach: Personally Identifiable Information ("PII")

The American analogue to personal data in the GDPR is personally identifiable information. This Subsection analyzes PII and its relationship to the normative stakes that justify its protection. Unlike Europe, which offers a unified definition of personal data, PII is a fractured concept within the United States, as it is defined in competing

[88] *See* Rouvroy & Poullet, *supra* note 66, at 70.

[89] *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11.

[90] *See* GDPR, *supra* note 15, art. 4(1).

ways across the American privacy law landscape.[91] After describing specific instances of how PII is conceived in several American privacy laws, the following Subsections examine how the decision to prioritize PII interacts with ideas about the goals and values that motivate information privacy law in the first instance.

a. PII Background

The advent of protecting PII through privacy legislation arguably marks the beginning of information privacy — rather than simply privacy — law in the United States.[92] Akin to personal data in Europe, PII defines the scope of privacy laws and demarcates the set of information which individuals have a credible claim to control.[93] Under most American privacy statutes, privacy violations can only occur when PII is improperly collected or used.[94] To that end, American privacy law generally regulates the collection, processing, and disclosure of PII, while leaving non-PII generally unprotected.[95]

American privacy statutes fail to offer a uniform definition of personal data.[96] While there is an obvious lack of definitional harmony about PII, commentators have recognized several existing approaches to defining it: the tautological approach, the non-public approach, and the specific types approach.[97] Ultimately, all these approaches suffer from conceptual difficulties that potentially undermine the effectiveness of the legislation in which they are defined.

The tautological approach provides one potential pathway to define the scope and content of PII. [98] Under this approach, PII is any information that identifies a person.[99] Paul Schwartz and Daniel Solove rightly criticize the tautological approach for failing to offer clear

[91] Schwartz & Solove, *The PII Problem*, *supra* note 1, at 1825-27.

[92] *Id.*

[93] *Id.* at 1816.

[94] *See id.*

[95] *Id.*

[96] *Id.* at 1828.

[97] *Id.*

[98] *Id.* at 1829.

[99] *Id.*

guidance about when information identifies a person.[100] Without clear guidance about what constitutes PII, privacy laws are necessarily vague and fail to provide adequate notice about what information is protected. However, this criticism may be muted slightly if we understand the tautological approach to invoke a flexible view of PII that can be further refined as privacy litigation makes its way through the court system. Although, until PII is clarified beyond mere tautology, it will invariably suffer from its ambiguity.

The non-public approach relies on the distinction between private and public information to define what constitutes PII within American privacy laws.[101] The "non-public" approach to PII simply claims that PII is information that is not public.[102] This approach attempts to sketch the boundaries of PII by showing what it is not, rather than providing a positive definition. The Gramm Leach-Bliley Act ("GLBA") provides one instance of the non-public approach as it defines PII as "non-public personal information."[103] The non-public approach has been criticized for its failure to track identification; or rather, whether information is identifying independent of its status as public or private information.[104]

The specific types approach simply identifies which information is worthy of protection and labels it PII.[105] When information belongs to one of the predetermined categories included in a statute, then it is protectible as PII.[106] For example, the Massachusetts data breach notification statute requires notification if personal information is compromised.[107] However, the statute defines personal information as specific types of information, such as social security numbers, driver's license numbers, and financial information.[108] Though the specific types approach offers clear guidance about what information is protected

[100] *Id.*

[101] *Id.* at 1830.

[102] *Id.*

[103] Gramm-Leach-Bliley Act of 1999, 15 U.S.C. § 6809(4)(A) (2018).

[104] *E.g.*, Schwartz & Solove, *The PII Problem*, *supra* note 1, at 1830.

[105] *Id.* at 1831.

[106] *Id.*

[107] MASS. GEN. LAWS ANN. ch. 93H, § 3 (2022).

[108] *Id.* § 3(b).

within a statute, it fails to provide general principles about what information should be protected as PII in other circumstances.

American privacy statutes provide an array of competing approaches to defining what constitutes PII. While many statutes offer divergent conceptions of PII, the general concept of PII prioritizes the identifiability characteristic of information that makes it worthy of protection. Though protecting directly identifiable information is still the dominant approach in the United States, there is some movement towards invoking a more general concept of personal data.[109] This American movement for a general definition follows the GDPR where both identified and identifiable information are protected as personal data. For instance, California's recent privacy statute ("CCPA") mirrors the GDPR's definition of personal data in several important respects, rather than relying on quintessentially American conceptions of PII.[110]

b. The Stakes of PII

The focus on identification as the central analysis for PII restricts privacy claims to those based around identification harms. Of course, privacy law should serve to deter and remedy the harms that accompany wrongful disclosures of identifying information. However, the prioritization of protecting against harms stemming from disclosures of identifying information causes privacy regimes centered on PII to lapse into two controversial views of privacy: privacy as secrecy and privacy as control over personal information.

(1) PII and Privacy as Secrecy

Both the non-public and the specific types approaches to PII falter by recreating some of the conceptual errors that underlie privacy as secrecy.

[109] *See, e.g.*, California Consumer Privacy Act, CAL. CIV. CODE § 1798.140(o)(1) (2022) (broadly defining personal information).

[110] *See id.* The CCPA provides a broad definition of personal information that includes any information that "identifies, relates to, describes, is reasonably capable of being associated with, or could be reasonably linked, directly or indirectly, with a particular consumer or household." *Id.* For a comparison between the GDPR's definition of personal data and the CCPA's, see Jordan M. Blanke, *Protection for 'Inferences Drawn': A Comparison Between the General Data Protection Regulation and the California Consumer Privacy Act*, 1 GLOB. PRIV. L. REV. 81, 85 (2020).

Consider, again, the non-public approach to PII that extends protection only to private information.[111] Here, the analytical symmetry is quite plain. Secrets are, by definition, not public information, so by cabining privacy claims to non-public information, American privacy law prioritizes secret (or non-public) information as the principal source of redressable privacy claims. Yet legitimate privacy claims extend to information that is not private.[112] Hence, the non-public approach to PII fails to account for the ongoing privacy interests that obtain even after information is disclosed.

The specific types approach also tracks the core considerations of privacy as secrecy. When statutes single out specific types of information (such as financial information) for protection,[113] these statutes imply that this information should remain undisclosed and, further, that there are harms that come from unwanted disclosures of this information. While this structure may make sense for some narrow types of information (such as passwords) where disclosures offer little benefit and could potentially cause downstream harms, limiting PII to non-public

[111] *See* Schwartz & Solove, *The PII Problem*, *supra* note 1, at 1830 (discussing the non-public approach to PII).

[112] Several privacy theorists have critiqued the idea that privacy interests are exhausted after information is disclosed. Helen Nissenbaum's influential theory of privacy as contextual integrity analyzes privacy according to whether information flows are appropriate based on the norms of the context in which they occur. HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 129-58 (2009). Nissenbaum's theory recognizes that some disclosures of information may be legitimate in one context but not another and, further, that whether information has been previously disclosed is largely irrelevant for this analysis. *Id.* at 142. Similarly, Edward Bloustein recognizes that people may have privacy interests in selective disclosures of information and the fact that information is not a secret does not necessarily entail that there is no longer a privacy interest. *See* EDWARD J. BLOUSTEIN, INDIVIDUAL AND GROUP PRIVACY 123-86 (1978); *see also* Solove, *supra* note 14, at 1108-09 (discussing the limitations of privacy as secrecy). The decline of the third-party doctrine in Fourth Amendment jurisprudence is another recognition that people have legitimate expectations of privacy in information that is not secret. *See* Carpenter v. United States, 138 S. Ct. 2206, 2221-22 (2018) (holding that individuals retain reasonable expectations of privacy in cell site locations even though this information is disclosed to cell phone providers).

[113] *See, e.g.*, Fair Credit Reporting Act, 15 U.S.C. § 1681 (2018) (requiring consumer reporting agencies to adopt protection procedures for consumer credit and insurance information).

information represents an incomplete accounting of the interests that privacy should protect.

While privacy law should certainly consider the potential harms that come from disclosures of private information, an exclusive focus on these types of privacy wrongs is incomplete.[114] The concerns of privacy are more capacious than safeguarding private information. Privacy also provides space to develop different interests and identities without the conforming force of public scrutiny.[115] Broadly, privacy protects our personal interest in autonomy and self-development that transcends mere protection of private — often commercial — information.[116]

(2) PII and Privacy as Control over Personal Information

Though PII forces privacy law to be especially sensitive to privacy as secrecy claims, it also relies on theories of privacy as control over personal information. Privacy as control over personal information claims that privacy is principally focused on granting individuals control over information that is about them. However, privacy as control over personal information offers a shaky normative foundation for PII and privacy law for several reasons.

First, the theory offers little guidance about what personal information is when, in fact, PII's role should help determine what information should be protected by law (and thus deemed personal information). Privacy as control over personal information puts the cart before the horse because it fails to offer a unified theory of when information is personal and instead leaves this central concept to intuition alone.

Second, not all information that describes a person (or has a semantic connection to a person) implicates interests of control.[117] For this reason, PII is too broad of a concept when it includes all information that identifies a person. There is a normatively significant difference between

[114] *See* Solove, *supra* note 14, at 1108-09.

[115] Julie E. Cohen, *What Privacy Is for*, 126 HARV. L. REV. 1904, 1906-12 (2013).

[116] *Id.*; *see also* Margot E. Kaminski, *Regulating Real-World Surveillance*, 90 WASH. L. REV. 1113, 1132-35 (2015).

[117] van der Sloot, *supra* note 16, at 5 ("However, because data processing often does not deal with private and sensitive data, the right to control by the data subject was felt undesirable . . . .").

having the ability to control information about our religious affiliation or sexual orientation and what brand of seltzer we regularly purchase.[118]

Commentators have attempted to alleviate these concerns by introducing the concept of sensitivity.[119] Sensitivity divides information into two classes: sensitive information and non-sensitive information, with sensitive information potentially implicating more fundamental privacy interests.[120] We can revise control over personal information to be control over *sensitive* personal information. However, there has been little consensus on what constitutes sensitivity,[121] so the definitional problem merely reemerges. Control theories of privacy are left to grapple with uncertain boundaries about what information people must control for privacy to be preserved.

* * * * *

As this Part demonstrates, the existing definitions of personal data fail to mirror complete theories of information privacy. Part II discusses several different attempts to examine the relationship that connects a person to their information. Identification alone fails to provide a strong reason for a person to be able to control this information. Relationships such as purpose and result do a better job of tracking a person's interest in uses of information but largely fail to capture what makes informational privacy unique. Ultimately, we demonstrate that these accounts fall short, and we set the stage to offer a unique vision of personal data based on separability that is more accurately attuned to the harms that come from the personal information economy.

[118] Of course, these distinctions may break down if computational tools allow firms to infer "sensitive" information from mundane information. *See, e.g.*, Mason Marks, *Emergent Medical Data: Health Information Inferred by Artificial Intelligence*, 11 U.C. IRVINE L. REV. 995 (2021) (explaining how private sensitive medical information can be inferred from relatively innocuous information).

[119] *E.g.*, Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1128-30 (2015) (calling sensitive information a "critically important but undertheorized concept" in privacy law).

[120] *Id.* at 1129 (explaining how "sensitivity" is a fundamental concept within information privacy law).

[121] *Id.* at 1130.

II. EXISTING RELATIONSHIPS BETWEEN PEOPLE AND DATA

Personal data is simply shorthand for the circumstances in which people should have control over information that describes them in some way.[122] In order to figure out the boundaries of personal data, it is essential to determine when and where a person has a protectable interest in information that warrants granting them some level of control over this information. The crux of this analysis is the relationship between a person and the information.

In this Part, we examine a few attempts to clarify the types of relationships between a person and data that give rise to a control interest over information. To start, we discuss *content* as the core relationship that creates a personal interest in data. A content relationship occurs when information is about a person, that is, the person and data are related through the content of the information. Second, we address relationships of *purpose* and *result*.[123] A purpose relationship occurs when the data is used or likely to affect a natural person. A result relationship occurs when the consequences of using information affect a natural person. The crux of the distinction between purpose and result is the intention behind the use. Purpose relationships primarily describe circumstances where data collection and use are intended (or likely intended) to affect a person. Result relationships, by contrast, capture cases where the effects on a particular person are a consequence but not a primary purpose of data use. Finally, we discuss attempts to define control rights through *propertarian* ideals that assess the division of labor between data subject and data collector in creating the information.

At their core, these different relationships are intended to divide control over information along normative justifications for control rights more generally. Our analysis in this Part examines different relationships

[122] *See* Wachter & Mittelstadt, *supra* note 21, at 498-99 (claiming that personal data determines the set of information through which individuals can control how they are seen by others).

[123] Content, purpose, and result relationships are discussed in *Peter Nowak v. Data Protection Commissioner*, in which the word "effect" replaces "result," and the EU Data Protection Working Party's Opinion on Personal Data. Case C-434/16, Peter Nowak v. Data Prot. Comm'r, ECLI:EU:C:2017:994, ¶¶ 35-39 (Dec. 20, 2017); *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11.

that are thought to give rise to control rights over information. However, many of these relationships do not adequately parse information rights or fail to provide a suitable normative underpinning for dividing rights of control in the first instance. Ultimately, each of these relationships fails to support a full vision of information privacy and data protection law.

A. Content

Content relationships are the most common feature that give rise to control rights and, therefore, constitute personal data under several different privacy regimes. Recall, the CJEU's jurisprudence and Art. 29 WP guidance documents explicitly mention that data should be directly or indirectly related to a person through content to satisfy the relation requirement for personal data under the GDPR.[124]

Similarly, the American preference for PII gives priority to identification, which is primarily an analysis of whether information is about a person. The American approach to PII is often haphazard though, and not all information about a person is protected by American privacy law.[125] American privacy statutes sometimes simply delineate specific types of information that are protected or distinguish between private/public information and designate this information as PII.[126] However, the conceptual priority of identity in PII demonstrates that American privacy law is principally focused on content relationships, which also explains why de-identified data (or data that has its content relationship stripped through certain measures) is usually unprotected in American privacy regimes.[127]

A common intuition about personal data is that the person and data are linked by the content, or rather, the data is "about" the person. For instance, a person's medical records are linked to the person because the

[124] *See supra* notes 40–53 and accompanying text.

[125] Schwartz & Solove, *The PII Problem*, *supra* note 1, at 1828-36.

[126] *Id.* at 1829-31.

[127] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1740 (2010) ("In addition to HIPAA and the EU Data Protection Directive, almost every single privacy statute and regulation ever written in the U.S. and the EU embraces — implicitly or explicitly, pervasively or only incidentally — the assumption that anonymization protects privacy, most often by extending safe harbors from penalty to those who anonymize their data.").

records reveal medical information about them. Similarly, a person's religious affiliation reveals something about them and is protected as personal information by virtue of its revelatory potential. The logic of granting people control over information that describes them is rooted in the idea that identification is at the core of many privacy harms.[128] Control rights are intended to allow data subjects to determine how information is shared and, by extension, who has access to details about them.[129]

However, this internal structure falters for several reasons. First, not all uses of data that identify a person lead to recognizable privacy harms. For instance, medical studies may use data that describes specific individuals to learn about trends and transmission vectors for a virus. The goal of using this data is to identify broad patterns rather than learn about a specific individual. At bottom, these uses of identifying information fail to raise privacy concerns and do not seem to justify granting control rights in the first instance. Or, at minimum, these uses of information do not necessarily allow data controllers to learn information about a person and, thus, privacy rationales for control are weakened.

Conversely, not all information that fails to identify a person is devoid of potential privacy harms. For example, indirect identifiers, such as cookies, pixel tags, IP addresses, and phone numbers do not identify a single individual and, as a result, are often not considered PII.[130] Yet indirect identifiers can be used to affect the interests of people who are linked to them and, therefore, should be protectable as personal data.[131] Consider some uses of IP addresses, such as using them to support targeting individuals for advertisements. Although any single individual

[128] For an account of privacy and data protection harms, see *supra* Part I.B (delineating the interests protected by privacy and data protection frameworks in the U.S. and Europe).

[129] *See generally* Ruth Gavison, *Privacy and the Limits of Law*, 89 YALE L.J. 421 (1980) (developing an access theory of privacy).

[130] *See, e.g.*, Johnson v. Microsoft Corp., No. C06-0900RAJ, 2009 U.S. Dist. LEXIS 58174, at *12-13 (W.D. Wash. June 23, 2009) (holding that an IP address is not PII because it identifies a computer, rather than a person).

[131] This possibility is accounted for in both the CCPA and GDPR definitions of personal data. *See* Blanke, *supra* note 110, at 85-90.

is not identified, they still have an interest in controlling uses of their IP address that affect them individually.

Additionally, identification alone fails to justify control in the information economy particularly because data processing is largely automated, so information is almost exclusively accessed by machines and other automated processing systems.[132] The automation of the data economy potentially lowers the stakes of content relationships as the crux of privacy harms. For example, Judge Richard Posner claims that privacy violations require some human intervention.[133] In other words, purely automated systems cannot violate an individual's privacy. While Judge Posner's position may be open to critique, definitions that exclusively prioritize identification as the basis of control rights find it more difficult to justify control over automated — rather than human — systems.

In sum, content relations fail to perfectly justify control rights. Of course, we may be concerned about the collection of identifying information, but these worries are better treated by more robust data security protections or data minimization techniques, rather than personal data designations and their attendant rights of control.

B. Purpose and Result

Personal data may potentially be connected to a person through a non-content relationship. That is, even if information does not describe a person, it still may be considered personal data. More specifically, the Art. 29 Working Party and the CJEU in *Peter Nowak v. Data Protection*

[132] The sheer amount of information makes it incredibly unlikely that even a small fraction of it will be viewed by a human. *See* Bernard Marr, *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*, FORBES (May 21, 2018, 12:42 AM EDT), https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read [https://perma.cc/6E98-PJ99] (offering statistics on the incredible volume of information that is created on the internet).

[133] Richard A. Posner, *Privacy, Surveillance, and Law*, 75 U. CHI. L. REV. 245, 254 (2008) ("Computer searches do not invade privacy because search programs are not sentient beings. Only the human search should raise constitutional or other legal issues.").

Commissioner suggest that information can be linked to a person through purpose or result, rather than content.[134]

Under the Working Party's interpretation, content, purpose, and result are each independently sufficient to meet the "relating to" requirement for personal data.[135] Because each relationship is individually sufficient for personal data, either an *ex ante* or *ex post* relationship satisfies this requirement for personal data. Content relationships are *ex ante*; the relationship exists prior to any particular use of the information. Purpose and result relationships, by contrast, are *ex post*. There may not be an *ex ante* connection between a person and information, yet a relationship is created through the use of the information.

A *purpose* relationship occurs when information is used or likely to be used to evaluate, treat, or influence the status or behavior of a person.[136] This definition sweeps broadly. Data does not necessarily even need to be used to affect an individual, as the mere likelihood of a use with this purpose is sufficient.[137] There are several possible scenarios where there is a purpose relationship without a content relationship. For instance, information collected by smart devices (e.g., Fitbit) may be decoupled from any identifiers yet still have a feedback system that alters a person's behavior.[138] A smart device could automatically remind a person to work out or perform some activity based on past action without providing any information about what individual is receiving these reminders.

A *result* relationship occurs when the use of information is likely to impact a person's rights or interests.[139] This includes uses of data that lead to data processors treating individuals differently. While purpose relationships are focused principally on the behavior of individuals,

[134] *See* Case C-434/16, Peter Nowak v. Data Prot. Comm'r, ECLI:EU:C:2017:994, ¶¶ 35-39 (Dec. 20, 2017); *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11-12.

[135] *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11 ("These three elements (content, purpose, result) must be considered as alternative conditions, and not as cumulative ones.").

[136] *Id.* at 10.

[137] *Id.* ("That 'purpose' element can be considered to exist when the data are used or *are likely to be used*, taking into account all the circumstances surrounding the precise case, with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual." (emphasis added)).

[138] Purtova, *supra* note 3, at 55 (discussing smart devices' feedback loop).

[139] *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11.

result relationships focus more closely on how individuals are treated.[140] For instance, data that is used to distribute a good or services within society has a result connection to different data subjects because the information is used to make determinations about which individuals receive these benefits.

Purpose and result capture a broad array of uses in the information economy. Both governments and private sector organizations routinely use information in ways that employ these relationships to data subjects. However, using privacy and data protection law to regulate these uses is misguided and allows these regimes to sweep broadly and capture almost any piece of information. For instance, even mundane information that has no obvious connection to any individual — like weather data — can become personal data through its use.[141]

Expanding the core of privacy and data protection law to govern uses of information that lack any content connection to a person pushes these laws beyond what makes them unique in the first instance. Privacy law is distinctive because it regulates information that remains *connected* to specific people even after it is transferred. This unique feature is what gives rise to individual interests in downstream uses of information. When a person reveals information about themselves, this information retains the possibility of uses that affect the person through the ongoing semantic connection that remains.

By contrast, most things extinguish their connections to specific people after transfer. Paradigmatic commodities are sold and no longer retain connections to their previous owners. When a person sells a car to a neighbor, their interests in the car are exhausted. Uses of the car do not have the capacity to affect the previous owner merely because they were previously connected to it.

Information, however, is different. Data that describes people retains a connection after transfer and privacy law should focus control rights on this type of information. While we may have concerns about uses of information that are only connected through *ex post* relationships (like

[140] *Id.* (arguing that even without the "'content' or 'purpose' element, data can be considered to 'relate' to an individual because their use is likely to have an *impact* on a certain person's rights and interests, taking into account all the circumstances surrounding the precise case" (emphasis added)).

[141] Purtova, *supra* note 3, at 57-60.

purpose or result), these are not principally the concerns of privacy or data protection law. We may assess the decision-making process to allay fears about due process or similar normative goals, but it seems an odd fit to regulate these uses through control rights within privacy and data protection law.

C. Propertarianism

The propertarian approach appeals to property themes — most usually labor investment — in order to justify granting control rights over data.[142] Under this view, data subjects imbue information with their labor and, as a result, should be able to both control how this information is used and profit off of different uses.[143] At their core, labor theories of data consider the person and information mediated by a labor relationship.

[142] There is an important distinction to make between justification and entitlements within propertarian theories of data. Stav Zeitouni, How Information Privacy Is Propertized 3-4 (Jan. 20, 2022) (unpublished manuscript) (on file with author). Some propertarians invoke property themes in order to justify granting rights over information to particular people. *See, e.g.*, ERIC A. POSNER & E. GLEN WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY 205-49 (2018) (suggesting that data is a form of labor produced by data subjects and, by extension, militates in favor of granting data subjects rights over this information); Malgieri, *supra* note 19, at 135 ("[A] desirable solution to the conflict between consumers and companies on personal information would be entitling stakeholders to property rights on personal data.").

By contrast, other propertarians invoke property themes in order to describe the type of rights (property rights) that data subjects should have over data. *See, e.g.*, LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 159-62 (1999) (offering a property account of rights over data). The property entitlement position has been criticized for a host of reasons, primarily, though, for failing to offer suitable protections to data subjects. *See, e.g.*, Ignacio Cofone, *Beyond Data Ownership*, 43 CARDOZO L. REV. 501, 524-40 (2021) (arguing that the property conception of personal data fails to protect consumers); Jessica Litman, *Information Privacy/Information Property*, 52 STAN. L. REV. 1283, 1295-301 (2000) (critiquing property approaches to privacy); Pamela Samuelson, *Privacy as Intellectual Property?*, 52 STAN. L. REV. 1125, 1136-46 (2000) (arguing that various unintended consequences would result from adoption of the property view of personal data privacy); Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 622 (2021) (identifying impracticability and incentive to sell personal data as two criticisms of propertarian data reforms).

[143] POSNER & WEYL, *supra* note 142, at 205-49; *see also* Malgieri, *supra* note 19, at 134 ("Personal data is not only perceived as a specific domain of human personality, but also as a pivotal element of the data-driven economy and a strong tool of consumer power on the market.").

As a result, debates over whether data subjects or platforms should be able to control data are reduced to an analysis of which party invested more labor in the creation of the data.

At bottom, propertarian views are not only descriptive but also inherently normative because they rely on notions of just distribution and fairness in allocating control rights.[144] A system that partitions control according to labor rewards parties that invested their labor with a suite of rights.[145] Supporters of the propertarian approach claim that delineating control along labor investment protects data subjects because without legally significant ownership claims, platforms will simply be able to assert de facto rights over data because of their ability to exclude.[146]

Gianclaudio Malgieri offers the most sophisticated taxonomy of how control rights might be allocated through an analysis of labor.[147] Malgieri suggests that control rights can be divided between platforms and data subjects according to the "degree of ownership" that data subjects exert over different pieces of information.[148] However, these appeals to ownership are more simply appeals to labor and, by extension, ground control rights in the normative force of labor theories more generally. For example, Malgieri contends that data subjects should have the most control rights over information that they provide to platforms and the

[144] Propertarian justifications for control rights over information bootstrap their normativity from Lockean ideas of property that grant ownership in cases where a person "mixes" their labor with an object. JOHN LOCKE, *The Second Treatise of Government*, *in* TWO TREATISES OF GOVERNMENT 306 (Peter Laslett ed., Cambridge Univ. Press 2d ed. 1967) (1690); *see also* Stephen R. Munzer, *Acquisition of Property Rights*, 66 NOTRE DAME L. REV. 661, 662 (1999) (examining Locke's account of just acquisition of property in greater detail).

[145] The inherent appeals to fairness and desert are evident in Lockean theories of property which are often discussed broadly as "labor-desert" justifications of property rights. *See* Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287, 305 (1988) (discussing labor-desert justifications for IP rights).

[146] Malgieri, *supra* note 19, at 135.

[147] Malgieri's taxonomy contains three distinct levels (strong, intermediate, weak) that track the relationship between the individual and information and, by extension, create different tiers of control. *Id.* at 137.

[148] *Id.*

least amount of control rights over information that is created (such as inferred data) by the platforms.[149]

Propertarian frameworks — such as Malgieri's — offer several benefits yet still remain incomplete. An analysis of labor as the basis for control rights provides a suitable foundation for some of the inherent debates about access and control in the data economy.[150] Labor is an acceptable normative foundation for parsing competing claims between data subjects and data controllers over erasure and portability.[151] In determining claims of data erasure and data portability, an examination of which party invested labor intuitively tracks which party should have the ability to erase and transfer data. There is a compelling logic of fairness at play here — the party that labored to create the data should be the one who is given rights to transfer and erase the data. Labor investment offers a solid philosophical grounding for these two control rights because erasure rights and portability rights sound, at least partially, in ideas about competition and ownership which labor identifies quite well.[152]

While labor theories provide a suitable foundation for granting portability and erasure rights, these theories fail to suitably justify a host of other control rights that are typical within privacy and data protection frameworks. For instance, the rights of access[153] and rectification[154] fit uneasily with an examination of labor as the foundational analysis,

[149] *Id.*

[150] Viljoen, *supra* note 142, at 618-19 (describing how the data as labor movement attempts to code "data about the subject" as "wealth for the subject," which is intended to solve the unequal monetary gains that pervade the data economy).

[151] Data as labor rests on the strongest foundation when it is applied in the debate about data portability. *See* Gabriel Nicholas, *Taking It with You: Platform Barriers to Entry and the Limits of Data Portability*, 27 MICH. TECH. L. REV. 263, 269-72 (2021) (describing the mechanics of data portability).

[152] Data portability is often linked to debates about competition between platforms because, on many accounts, portability is offered as a potential solution to platform market power. Under the conventional view, data portability finds its normative valence in competition law rather than exclusively privacy law. *See, e.g.*, GABRIEL NICHOLAS & MICHAEL WEINBERG, ENGELBERG CTR. ON INNOVATION L. & POL'Y, DATA PORTABILITY AND PLATFORM COMPETITION (2019) (outlining data portability as a possible solution to reduce platform market power).

[153] GDPR, *supra* note 15, art. 15.

[154] *See generally id.* art. 16 (defining rectification in the GDPR).

largely because these rights track more traditional privacy concerns. Even in situations where data subjects have not invested in creating the data, there is still a compelling case to be made that they should be given rights of access and rectification. Further, propertarian theories will not be sensitive to interests that generate these additional control rights.

At its core, an analysis of labor has its strongest benefits in parsing competing control claims that are related to the nexus of labor and fair competition, such as data portability. This is largely because one justification for whether a person should be able to erase or move information is their role in the creation of the information itself. Again, the focus on the upfront investment of labor is more principally about allocating control rights through a notion of labor desert rather than more traditional privacy concerns that typically motivate data protection and privacy regimes.

In addition, the propertarian approach leaves itself open to several lines of deeper criticism. Alongside giving conceptual priority to labor, which often overlooks more traditional privacy concerns, rewarding labor for creating data may actually undermine privacy values rather than merely overlook them.[155] More specifically, data minimization is often heralded as a primary value in privacy and data protection policies, but propertarian approaches are inherently at odds with minimization techniques because these approaches reward data creation rather than minimization.[156] At bottom, propertarian interventions in the data economy implicitly assume that data creation is a valuable enterprise in the first instance.[157]

Finally, while propertarian justifications are strongest when they are dividing control claims between users and platforms, the data economy is often significantly more complex. The competing claims of control of use rights in the information economy are often between multiple users.[158] Competing claims of control between data subjects can occur in

[155] Viljoen, *supra* note 142, at 622.

[156] *See id.*

[157] *See id.* at 623.

[158] *See* Cameron F. Kerry & John B. Morris, *Why Data Ownership Is the Wrong Approach to Protecting Privacy*, BROOKINGS: TECHTANK (June 26, 2019), https://www.brookings.edu/blog/techtank/2019/06/26/why-data-ownership-is-the-wrong-approach-to-protecting-privacy [https://perma.cc/H6XH-JVAL] (explaining the problem of overlapping interests in data as property systems).

some cases. For example, there are competing control claims in a picture where multiple people's images are captured, but one of the people in the picture operated the camera and, therefore, caused the image to exist. Propertarians may take their approach from IP law and grant control rights in response to the labor of the camera operator, but this would ignore the privacy (and other non-propertarian) interests of others in the photo.[159]

Propertarian approaches offer some useful guidance for parsing control rights between data subjects and data controllers. However, they are vulnerable to critique for often overlooking privacy values in their analysis, failing to offer much guidance for competing claims by data subjects, and sometimes incentivizing the creation of data in cases where a better policy approach would favor minimization. As it stands, propertarian approaches are unlikely to hold sway over most privacy advocates and, by extension, remain a poor fit for grounding control rights in the data economy.

* * * * * *

As this Part demonstrates, existing relationships between a person and data are incomplete and open to an array of critiques. In the next Part, we provide a new understanding of the relationship between an individual and information that justifies control rights and, by extension, identify what data should be considered personal.

III. SEPARABILITY AS A FOUNDATION FOR PERSONAL DATA

This Part offers a novel conception of personal data based in the philosophical concept of separability. We argue that inseparable uses of data should be considered personal data and subject to individual rights of control. Separable uses, by contrast, should not be considered personal data and not governed by control rights.

Under an account of personal data grounded in separability, both *ex ante* and *ex post* relationships are necessary. More specifically, personal data must have a connection to a person and, in addition, must appropriate that connection to affect them. Ultimately, restricting

[159] *See generally* Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 60 (1884) (holding that the copyright vests in the camera operator, rather than the subject of the photograph).

personal data in this way creates a system that is better conceptually because it more faithfully highlights uses that raise individual concern. Separability identifies instances where a person is connected to information and this connection is exploited, thus demarcating the set uses where individuals have strong claims of control.

Normatively, this approach offers information privacy law a more rigorous foundation in moral philosophy as well because inseparable uses (or uses that are connected to a person) implicate Kantian concerns about treating a person as a thing and using a person as a means to an end.[160] As a result, separability provides a rich analysis to differentiate between uses of information that use the personal connection inherent in the information (and are inseparable) and uses that do not (and are benign).

This Part begins by providing a brief overview of separability. Next, it details the necessary conditions of inseparability and applies this framing to identify the boundaries of personal data. After that, it moves to discussing the philosophical stakes of separability. And finally, this Part develops a more rigorous approach for sympathetic judges and lawyers to implement these ideas through the prism of separability.

A. Brief Overview of Separability

Separability often works in service of distinguishing persons from things. This distinction is necessary because many philosophical approaches to property grant rights of control over things but contend that similar rights of control over people are impermissible.[161] Any approach to dividing the world into these two metaphysical categories requires some analysis to make these determinations, and conceptual accounts of separability are often invoked to make this determination.[162]

[160] *See* Robert Johnson & Adam Cureton, *Kant's Moral Philosophy*, STAN. ENCYCLOPEDIA OF PHIL. (Jan. 21, 2022), https://plato.stanford.edu/entries/kant-moral [https://perma.cc/7MJA-HQZQ].

[161] Kant dismisses the possibility of property claims in people because it violates the fundamental ethical norm that requires that people never treat another person as a means to an end. Kantian systems of property must identify the set of things that can be used as a means and, further, only these things are suitable for property claims. *See* ARTHUR RIPSTEIN, FORCE AND FREEDOM: KANT'S LEGAL AND POLITICAL PHILOSOPHY 35-37 (2009).

[162] *See id.*

There is a general consensus among both philosophers and property theorists that things are separable from people. While the general principle of separability is not contentious, the conditions of separability are deeply contested. Some theories of separability offer little content to make this determination. For instance, G.W.F. Hegel claims that only things that are external to the person are appropriate for property claims.[163] Hegel, then, appears to simply claim that things are separable when they can exist separately from the person.[164] Ultimately, these loose guidelines fail to provide much guidance at best and are purely tautological at worst.

Other theorists such as Immanuel Kant[165] and Margaret Jane Radin[166] appeal to physical separation to determine the conditions for separability. Parts of the body, or anything physically connected to the person, are inseparable and should not be treated as property. However, this view falters for several reasons. First, separability as physical separation rests on the fallacy of division. That is, just because someone treats a part of the body (consider hair used in a wig) as alienable property does not mean they treat the person that way.[167] Second, physical separation reduces the person to their physical boundaries and, by extension, prioritizes only physical connections. Yet some things that

[163] *See* G.W.F. HEGEL, ELEMENTS OF THE PHILOSOPHY OF RIGHT § 65, at 95, § 67, at 97 (Allen W. Wood ed., H.B. Nisbet trans., 1991); *see also* Neil Netanel, *Copyright Alienability Restriction and the Enhancement of Author Autonomy: A Normative Evaluation*, 24 RUTGERS L.J. 347, 359 (1993); Margaret Jane Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849, 1894 (1987).

[164] *See* HEGEL, *supra* note 163, §§ 65-66, at 95-96.

[165] *See* IMMANUEL KANT, LECTURES ON ETHICS 124 (Louis Infield trans., 1930) [hereinafter LECTURES].

[166] *See* Margaret Jane Radin, *Property and Personhood*, 34 STAN. L. REV. 957, 966 (1982) ("We have an intuition that property necessarily refers to something in the outside world, separate from oneself . . . . This intuition makes it seem appropriate to call parts of the body property only after they have been removed from the system.").

[167] *See* Stephen R. Munzer, *Kant and Property Rights in Body Parts*, 6 CAN. J.L. & JURIS. 319, 325 (1993); *see also* Verstraete, *supra* note 7, at 443-44 (addressing Munzer's critique of Kantian separability).

are foundational to a person are not connected physically but, instead, retain non-physical connections.[168]

Recognizing the shortcomings of earlier accounts of separability, J.E. Penner contends that anything that is contingently connected to the person is separable.[169] Penner claims that if a thing "might just as well be someone else's," then the relationship is contingent and separable.[170] Pushing against physical separation, Penner endorses the idea that organs or other parts of the body may be separable.[171] Penner claims that kidneys and other organs, in fact, can be someone else's and therefore can be considered property.[172]

While Penner is correct that the point of analysis should be about contingency, he misunderstands that contingency inheres at the level of use rather than being an inherent feature of something. Consider organ donation. Even though an organ can be transferred and become someone else's, there are still potential uses that are connected to the donor because of the genetic link that persists. Clearly, using a donated organ to perform routine bodily functions does not implicate the donor. Yet harvesting the genetic material remaining in the organ in order to find out the donor's potential for certain diseases and market medical products based on this information is inseparable from the donor, even though the organ is physically transferred.

Separability, then, is a product of both connection and use.[173] For something to be inseparable it must be connected to the person and the use must appropriate that connection to affect them.[174] This concept transcends privacy and data protection law. Many other things are potentially inseparable, such as creative works, body parts, and public

[168] Verstraete, *supra* note 7, at 443 ("[S]ome things that are physically separate may retain a nonphysical connection that can plausibly be leveraged to use the person, leading to underinclusion.").

[169] J.E. PENNER, THE IDEA OF PROPERTY IN LAW 111 (1997).

[170] *Id.* at 112 (emphasis omitted).

[171] *Id.*

[172] *Id.*

[173] *See* Verstraete, *supra* note 7, at 448-55.

[174] *Id.* at 452-53.

personas, which all retain a connection to a person that persists after transfer.[175]

The next Section applies this framework as the foundation of personal data. Individuals should retain control rights over uses of information that depend on the connection between person to the data and use this connection to affect the person who is described in the data. As a result, personal data identifies these situations where control rights are warranted because of a continued interest that remains even when information is passed downstream.

B. Personal Data Through the Lens of Separability

Rather than using separability to distinguish between persons and things, we can use it to demarcate which uses are connected to the data subject (and subject to control rights) and which uses are disconnected (and are a poor fit for individual rights of control). This is the work that personal data ought to be doing in both privacy and data protection. An acceptable account of personal data identifies the personal interests in data, which is exactly what separability highlights.

Applying separability to personal data yields this framework. Personal data must have a connection to the person *and* the use must rely on this connection in order to affect the person. Personal data, then, has two necessary conditions, one *ex ante* and one *ex post*. Both of these conditions are necessary to identify the continued interest that a person has over information and, as a result, are the basis for our account of personal data.

1. Connection

There are a number of potential connections that satisfy the first condition for inseparability.[176] With data, the connection is often a semantic connection. Or rather, information that is about a person or

[175] *Id.* at 432 ("This Article applies separability to a unique set of things — such as body parts, publicity rights, creative works, and personal data — that retain a connection to a specific person even after they are transferred. Inseparable uses depend on the connection linking the person to the thing. This Article argues that people retain an inalienable deontological interest in controlling inseparable uses of a thing even after it is transferred.").

[176] *See id.*

describes them in some way. The CJEU and Art. 29 WP describe this feature as a content relationship and, further, it is the most common way that people typically conceptualize personal data.[177] However, data can also be connected to a person even when they are not identified in the data. For instance, the use of identifiers — such as IP addresses and telephone numbers — can connect a person to information even when the information is *about* a computer (as in the case of IP addresses) rather than a person. This is still a connection as it allows the identification of a person.

The primary upshot of connection is that information that does not have some connection to a person will fail to meet the first threshold condition of personal data. Here, our approach departs from the GDPR in significant ways. First, under the GDPR, data not connected to an individual can still constitute personal data so long as its use is related to that person. For example, weather data — which has no connection to any particular person — can constitute personal data under the GDPR if it is used or likely to be used to affect a person's behavior or status. Rather than being related to a person through content, data can be related through use and satisfy the requirements of personal data under the GDPR.

Second, our approach conflicts with the GDPR in determining whether de-identified data are personal data. Commentators suggest that de-identified data fall within the ambit of personal data under the GDPR's framework because de-identification is not infallible.[178] Although identifiers are stripped from data sets to produce de-identified data, new data could potentially be introduced to re-identify the information.

By contrast, separability suggests that uses of de-identified data are not personal data and not subject to traditional data subject control rights. However, this does not necessarily entail that once data is de-identified that the people who were formerly described in the data lose all their rights and interests. Once de-identified data is re-identified, then

[177] *See* Case C-434/16, Peter Nowak v. Data Prot. Comm'r, ECLI:EU:C:2017:994, ¶¶ 35-39 (Dec. 20, 2017); *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11 (discussing the CJEU's jurisprudence and Art. 29 WP guidance documents' examination of content relationships that connect a person to information).

[178] Purtova, *supra* note 3, at 41-42; *see* Graef et al., *supra* note 16, at 6.

the condition of a connection is restored, and the information is potentially personal data depending on how it is used.

While separability departs from the GDPR when examining de-identified data, separability tracks similar conceptual terrain as anonymization under the GDPR. In commentary about the GDPR, anonymization typically describes de-identification techniques that are sufficiently robust to be irreversible.[179] Anonymized data is not considered personal data under the GDPR or our theory of personal data grounded in separability.[180]

Some data lacks any connection to an individual and, therefore, is not considered personal data under a theory grounded in separability. Other data is connected to a person, so its use must be considered in order to fully determine if the data constitutes personal data.

2. Use

The second condition for inseparability depends on how the data is used. In order for data to be used inseparably, the use must depend on the connection in order to affect the person. The crux of inseparability is not merely that some data is connected to the person (content relationships) or that data can be used to affect certain people (purpose or result relationships), but rather that some uses depend on that connection to bring about specific effects.

Most centrally, only some uses of data will meet this threshold and, as a result, only some secondary uses of identifying information will qualify as personal data. This framework provides a different point of analysis than many systems of information governance. Consider again, the GDPR. The GDPR restricts processing personal data for purposes other

[179] *Opinion 05/2014 on Anonymisation Techniques*, at 8 (Apr. 10, 2014), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [https://perma.cc/NM6H-2UWD].

[180] GDPR, *supra* note 15, Recital 26. However, the GDPR's approach to personal data seems at odds with its expansive approach to the "relating to" requirement for personal data. It is not obvious why anonymized data could not be related to an individual through purpose or result (*ex post* uses) and, therefore, track other cases where personal data does not have a content relationship to any individual and is still considered personal data. That said, separability suggests similar results in the case of anonymization. Anonymized data do not meet the threshold condition of retaining a connection to a specific person that persists after transfer.

than those for which the data was initially collected.[181] However, the GDPR explicitly makes exceptions for secondary uses including historical, statistical, or scientific purposes.[182]

One primary difference between a theory of personal data grounded in separability and the GDPR is the way in which information is related to the person. Recall that there are several different ways in which information can satisfy the "relating to" requirement under the GDPR. These different relations (content, purpose, result) are disjunctive; that is, they are independently sufficient.[183] Separability, by contrast, requires connection and effect. In logical terms, inseparable uses of information require content and (purpose or result) and further, the purpose or result must depend on the content connection.

There are benefits to basing control over secondary uses on separability rather than whether the purpose of the use is consistent with the reasons why the data was initially collected.[184] The GDPR seems to elevate the role of the purpose of collection as a moral baseline.[185] Secondary uses are assessed against this backdrop. The underlying

[181] For example, according to Recital 50 of the GDPR, scientific research purposes are considered to be compatible lawful processing operations. GDPR, *supra* note 15, Recital 50. The European Data Protection Supervisor has explained that the presumption of compatibility for research purposes depends on the requirement in Article 89(1) to ensure appropriate technical and organizational safeguards, such as pseudonymization and access limitations, and that the data should not be used to support measures or decisions regarding any particular individual. *Preliminary Opinion of the European Data Protection Supervisor on Data Protection and Scientific Research*, at 22 (Jan. 6, 2020), https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf [https://perma.cc/V8XA-E7P6].

[182] GDPR, *supra* note 15, Recital 50, arts. 5(1)(b), 6(4), 89; *see Opinion 03/2013 on Purpose Limitation*, at 11 (Apr. 2, 2013), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf [https://perma.cc/S6B3-JT4E] [hereinafter *Art. 29 WP Opinion Purpose Limitation*].

[183] *See Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 11.

[184] *See* GDPR, *supra* note 15, art. 5(1)(b) (requiring that personal data be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes").

[185] Rather, the purpose limitation requirement assumes that consent justifies uses for the purposes for which the information was initially collected. Different uses of information are not consented to and, therefore, lack lawful justification. However, the GDPR does explicitly mention lawful bases of processing data that do not require the consent of the data subject. *See* EDPB, *supra* note 31, at 25.

assumption here is that people are aware of the purpose of initial collection and consent to this purpose, so further processing of data that comports with this purpose does not raise autonomy concerns. While this may work in theory, the reality of the information economy is that data subjects are unaware of the original purpose of collection or the purpose is stated at a broad level of generality, so many secondary uses will be justified.[186]

By contrast, separability identifies uses which are connected to the data subject and uses which are not. This creates a more coherent framework for determining which secondary uses data subjects should be able to control. Consequently, separability serves as a better normative foundation for identifying when the data subjects' interests are at stake in a secondary use.

A short example helps clarify how separability tracks privacy interests in secondary uses more effectively than the GDPR. Consider the collection of information for banking purposes where the bank decides to use this information to make their services more efficient for all customers. Under the GDPR approach, personal information cannot be processed for system optimization because it was collected for banking.[187] Yet it is not immediately obvious why the data subject has a right to limit this use because it does not seem to implicate their interests in any significant way. Separability, by contrast, contends that this use of data for system optimization does not affect the person and, therefore, is not personal data subject to rights of control.

3. Connection and Use Combinations for Personal Data

This Subsection details the different combinations of connection and use to more coherently depict the conditions for personal data under a theory of separability.

[186] Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1005-06 (2017) (addressing the possibility that companies will collect data and attempt to define broad or vague purposes for collection).

[187] *Art. 29 WP Opinion Purpose Limitation*, *supra* note 182, at 53 n.122.

a. Connection & ~Use

There are several cases in which there is a connection between the person and data, but the use does not affect the person. For instance, information that describes a person may be used to train an algorithm that will not be used to alter the behavior or status of the person described in the data. Similarly, data that describes a person could be used to examine broad statistical trends, rather than make decisions or determinations about the person to whom the data refers.

While these uses will not be considered personal data under our theory of separability, because the use condition is not met, existing theories of personal data likely include this information within their ambit. Because the information identifies a person, the data will be considered personal data under PII, the GDPR, and the CCPA.[188] Consequently, data subjects within these information regimes will be allowed to control these uses even though they appear to lack strong justifications for doing so.

b. ~Connection & Use

There are other cases in which there is no connection between a person and the data, yet the use affects the person. An example of this combination is using weather data to influence an individual's behavior. For instance, an Internet-connected car may enable different driving features when it senses rain, which, in turn, affects the driver of the car. Or, similarly, the price of driving on certain roads could be made more expensive during rainy weather, thus affecting the individuals who choose to drive during this time.

While this use of information is not personal data under our theory of separability, it is more contentious under current accounts of personal data. The use of non-identifying weather data fails the first condition for inseparability; that is, it lacks any connection to a person. However, this use likely qualifies for protection as personal data under the GDPR through a purpose or result relationship.[189] By contrast, this use of weather data is likely not considered personal data under PII because it does not identify a person.[190]

[188] *See supra* Section I.B.1.a.

[189] *See supra* Section I.B.1.a.

[190] *See supra* Section I.B.2.

c. Connection & Use (*Use Does Not Depend on Connection*)

In some cases, there is a connection between the person and data, but the use does not depend on the connection. Admittedly, this case is a bit difficult when stated abstractly, but an example should help provide more clarity. For example, SmartBank collects data about how its customers use its online banking system. Moreover, this information is linked to a customer's name and bank account, so it meets the connection condition for inseparability. And further, this information is used by SmartBank to change its online banking interface to help its customers navigate the web page. So, the information use affects the customer by altering their online banking experience.

However, SmartBank's use of the information does not depend on the fact that it is any one particular customer's bank information. Put differently, the fact that it is Adam's bank information is irrelevant to the effectiveness of SmartBank's use; it just needs to be an SmartBank customer but not necessarily any specific one.

While the information has a connection to the person and the use affects the person, it is not personal data under our theory of separability. However, this use would be considered personal data under all other existing theories of personal data. The information is related to a person through both content and purpose; thus, it meets multiple sufficient conditions for personal data under the GDPR.[191] This use would be personal data under PII as well because the information describes an identified person.[192]

d. Connection & Use (*Use Depends on Connection*)

In some cases, there is a connection between the person and the data, but the use depends on that connection to bring about its effects. Continuing with the SmartBank example clarifies this distinction further. Rather than using the information for improving their online banking system, SmartBank decides to use the data to expand its newly formed eCommerce division. In doing so, SmartBank pools customer location data, financial history, and age in order to tailor product

[191] *See supra* Section I.B.1.a.

[192] *See supra* Section I.B.2.

advertisements to specific customers and, by extension, increase product sales.

SmartBank uses information that identifies specific customers according to unique identifiers (name and bank account) and, therefore, meets the connection condition for protection as personal data. In addition, SmartBank uses the information to affect these customers; that is, customers see different advertisements based on their data. And finally, the use depends on the connection because the advertisement's successfulness is a product of whether a specific person's data is used. Put another way, if a SmartBank customer accidentally logged into a friend's account who had different characteristics, the advertisements would be less effective. Thus, the use is partially dependent on the fact that it is a specific person's data.

While this use of information is personal data under our theory of separability, it is also personal data under existing accounts. Current accounts of personal data fail to interrogate the relationship between the connection a person has to their data and the use of the data. As a result, the analysis for this use under extant accounts of personal data is the same as in the previous Subsection where there was connection and use, but the use did not depend on the connection. Moreover, existing theories of personal data fail to recognize the different autonomy and dignity stakes in these two examples. Fortunately, our theory provides the groundwork for information governance regimes to capture these distinctions.

C. The Moral Philosophy of Separability

Having clarified the conceptual boundaries of separability, this subpart discusses the foundational moral principles that ground our theory of personal data. An analysis of separability identifies core values such as autonomy and dignity. Separability tracks these values primarily because it identifies the boundaries of the person, and by extension, identifies uses that are connected to the person (which they have an interest in controlling) and uses that are not connected (which they do not have an individual interest in controlling).[193] At the same time, information can

[193] For a more detailed summary of the philosophical stakes of separability, see Verstraete, *supra* note 7, at 435-39.

also be used for purposes that create value and allow people to pursue their own goals and interests. Separability offers a strategy to determine when different uses of information potentially undermine autonomy and dignity while also preserving the possibility of other beneficial uses that do not upend these values.

The roots of separability follow from earlier work in moral philosophy. In particular, G.W.F. Hegel[194] and Immanuel Kant[195] invoke separability to determine the analytical boundaries of property and persons. For both Hegel and Kant, property rights are only acceptable in things that are external to the person.[196] This distinction is essential for philosophical approaches to property. This is largely because property is necessary for people to develop and pursue projects; it provides stable expectations about resources that allow people to actualize their views of the good life.[197] However, property claims also need to be limited in some way and restricting property rights to things that are external to the person prevents the rights of control inherent in the property system to limit the autonomy of other people.[198]

While an analysis of separability that demarcates the boundary between persons and things is essential for both Hegel and Kant, Kant is more explicit about the normative valence that underlies this distinction. Kant's system of moral philosophy depends on distinguishing persons from things because we have different obligations towards people than we do mere things.[199] And further, Kant depends on separability to mark this distinction.[200] Persons have dignity and derive their value from their

[194] HEGEL, *supra* note 163, § 67, at 97.

[195] KANT, LECTURES, *supra* note 165, at 124.

[196] *See* HEGEL, *supra* note 163, § 65, at 95; Netanel, *supra* note 163, at 359 (discussing the use of the subject/object dichotomy in Kant and Hegel).

[197] *See* Henry E. Smith, *Property as the Law of Things*, 125 HARV. L. REV. 1691, 1693 (2012) (discussing the property system's purposes).

[198] Hegel asserts that self-actualization requires a property system where people can exert their will over things in the world in order to pursue projects and exercise their autonomy. HEGEL, *supra* note 163, § 41, at 73. However, Hegel also recognizes that overextending the property system to include persons would be a limitation on freedom and self-actualization through turning oneself into the property of another. *See id.* §§ 66-67, at 95-97.

[199] *See* IMMANUEL KANT, GROUNDWORK FOR THE METAPHYSICS OF MORALS 52-53 (Allen W. Wood ed. & trans., 2002) [hereinafter GROUNDWORK].

[200] *See* KANT, LECTURES, *supra* note 165, at 124.

rationality.[201] Put in finer detail, Kant claims that rational beings "are called *persons*, because their nature already marks them out as ends in themselves, i.e., as something that may not be used merely as means."[202] As we can see, rationality provides the moral worth of persons and provides guidelines about how they can be treated.

Because Kant divides the world into two entities, any entity that lacks rationality is not a person but merely a thing. While persons have rationality and derive an inherent value from this characteristic, things do not have inherent value. Instead, the value of things is merely instrumental; that is, they are valuable insofar as they are useful for the projects that people undertake.

Kant's metaphysical work is grounded in moral philosophy that has thick normative prescriptions. For instance, Kant's Humanity Formula requires that a person "[a]ct so that you use humanity, as much in your own person as in the person of every other, always at the same time as [an] end and never merely as [a] means."[203] Conversely, whenever a person uses another as a mere means they fail to acknowledge the other's inherent worth and thus undermine their dignity.[204]

Similarly, we follow Kant's lead and deduce normative boundaries that mark the limits of personal data. Returning to separability, uses of data that are still connected to the person are inseparable and risk using a person in a way that undermines their dignity. As a result, people ought to have control over these uses because they have a clear dignitary interest in these uses. Other uses, however, are separable from the person such that they do not implicate the dignitary interests of the person described by the data. In sum, inseparable uses of information are sufficiently related to the person described by the data that we can label these uses as personal data and provide attendant rights of control over them.

[201] *See* KANT, GROUNDWORK, *supra* note 199, at 46.

[202] *Id.*

[203] *Id.* at 46-47 (emphasis omitted).

[204] *See id.* at 46-48.

D. The Seeds of Separability in Privacy and Data Protection

The move towards an account of personal data that is based on separability is significantly different than current approaches; however, it does not represent an entirely radical departure. Similar to an analysis of personal data through the prism of separability, there are several elements within European data protection law which base determinations about personal data on contextual factors. Put another way, both separability and EU law demarcate personal data by appealing to features outside the information itself.

For instance, in Europe whether information relates to an "identified or identifiable" individual is based on the current state of technology.[205] As re-identification technology becomes more sophisticated and commonplace, the boundaries of personal data will be redrawn as the scope of "identifiability" shifts. As a result, a piece of data may not be personal data at the time of collection but will be personal data several years later.

Likewise, both separability and EU data protection law recognize the importance use for an analysis of the personal interests inherent in the data economy. For instance, the GDPR offers several distinctions about the role of secondary uses of information.[206] In some instances, secondary uses may be outside of the purpose for which the information was collected and, thus, require the consent of the data subject.[207] Further, some uses of information — even if they are outside the scope of the initial purpose of collection — do not require the consent of the

[205] GDPR, *supra* note 15, Recital 26.

[206] *See* GDPR, *supra* note 15, arts. 5(1)(b), 6(4); *see also* GDPR, *supra* note 15, Recital 50 ("The processing of personal data for purposes other than those for which the personal data were initially collected should be allowed only where the processing is compatible with the purposes for which the personal data were initially collected . . . . If the processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, Union or Member State law may determine and specify the tasks and purposes for which the further processing should be regarded as compatible and lawful. Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes should be considered to be compatible lawful processing operations.").

[207] *See Art. 29 WP Opinion Purpose Limitation*, *supra* note 182, at 32.

data subject. For example, research uses are specifically carved out as a class of uses that do not require consent.[208]

Both separability and EU data protection law appeal to considerations outside an analysis of the information itself. However, the EU's principles about when and why contextual factors are relevant are theoretically unclear or, at best, only implicitly stated. The core problem that arises from the lack of theoretical clarity underlying the appeal to contextual factors within data protection is that the principle that is used to derive these outcomes is opaque. This opacity, in turn, leads to a common criticism of the GDPR's approach to personal data — that it is ambiguous and fails to provide sufficient guidance to firms attempting to differentiate personal from non-personal data.[209]

While the GDPR approach shares some similarities with our approach grounded in separability, there are obvious conceptual differences. Most centrally, the GDPR does not consider use to be a necessary feature for personal data. In other words, data may be considered personal without reference to any specific use. Moreover, introducing separability into an analysis of personal data invokes contextual factors, but it provides more rigorous guidance than the definition currently offered by the GDPR. This is largely because separability surfaces otherwise implicit theoretical considerations and offers firms a roadmap for determining when specific uses are connected to a person and when they are distinct.

The next Part applies our theory of personal data grounded in separability to the contentious case of inferred data. As it stands, the status of inferred data as personal data and whether rights of control should attach is contested. This discussion demonstrates that distinguishing separable from inseparable uses not only works in service of important dignity and autonomy values but also aligns data protection to the stark realities of the information economy.

IV. TESTING SEPARABILITY THROUGH INFERRED DATA

This Part applies our theory of personal data grounded in separability to inferred data — data about individuals (such as credit scores) that are

[208] *See* GDPR, *supra* note 15, art. 6(4).

[209] *See* Michele Finck & Frank Pallas, *They Who Must Not Be Identified — Distinguishing Personal from Non-Personal Data Under the GDPR*, 10 INT'L DATA PRIV. L. 11, 12 (2020).

derived from existing data.[210] Inferred data is increasingly important in the information economy. The development of cutting-edge data analytics and artificial intelligence ("AI") marks a shift from companies primarily collecting data to companies generating inferred data. And further, the status of inferred data as personal data is deeply contested within information governance frameworks.[211]

The growing importance of inferred data requires rigorous thinking about when and whether this information should be classified as personal data. Again, the classification of inferred data as personal data is centrally important because it triggers privacy law and data protection frameworks. We argue that our theory of personal data grounded in separability offers the best path towards cataloging the conditions under which inferred data should be considered personal data.

A. The Law and Practice of Inferred Data

Inferred data is the product of inferential and predictive analytics.[212] More recently, the development of AI has strengthened the processing and predictive power of existing analytical techniques, creating more

[210] *See* Niederman, *supra* note 6, at 29-30.

[211] *See, e.g.*, Blanke, *supra* note 110, at 92 (arguing that the GDPR should protect inferred data as personal data); Wachter & Mittelstadt, *supra* note 21, at 515-21 (arguing that inferred data should be construed as personal data); Devika Bansal, *Scope and Analysis of Inferred Data: Application and Implications*, CONTEMP. L.F. (Dec. 27, 2021), https://tclf.in/2021/12/27/scope-and-analysis-of-inferred-data-application-and-implications [https://perma.cc/UPS3-2X4C] (complaining of the exclusion of inferred data from the GDPR's scope of application); Howard Yu, *GDPR Isn't Enough to Protect Us in an Age of Smart Algorithms*, CONVERSATION (May 29, 2018, 10:25 AM EDT), https://theconversation.com/gdpr-isnt-enough-to-protect-us-in-an-age-of-smart-algorithms-97389 [https://perma.cc/H9A8-ECS3] (advocating for the inclusion of inferred data under the GDPR in order to more fully protect data subjects). The sole clear exception of this contestation is the CCPA which simply states that inferred data is personal data. *See supra* Part II.A.

[212] Inferential analytics are used to deduce existing attributes or preferences such as gender or political opinions. By contrast, predictive analytics are used to make predictions about future outcomes using historical data combined with statistical modeling, data mining techniques, and machine learning. *See* Blanke, *supra* note 110, at 81-82 (providing examples of predictive analytics used by companies like Target and Facebook).

granular inferences about individuals.[213] The turn toward inferential data is here to stay as inferences are the source of immense commercial value and companies are likely to increase their capacity to create them.[214]

There are a few steps that companies use to create inferences. To start, the underlying data (source data) that is used to generate an inference is collected. Next, the source data is prepared to be inputted into an AI tool or analytical model. After that, the source data trains the AI tool or analytical model. Finally, the source data is inputted into an AI tool or analytical model to create an inference.[215]

Inferred data is used for a variety of purposes. One common use of inferred data is to supplement profiling. When used this way, inferred data fills gaps in incomplete datasets or serves as a check on the accuracy of available data.[216] Datasets enriched with inferred attributes are likely to have higher levels of completeness and accuracy; that is, they create better profiles.[217] In addition, inferred data can be reused as inputs for further analytics, thus creating more precise data analytics systems.[218]

Above all, the key characteristic and innovation of inferred data is that it is derived from all sorts of data, not just data that relates to a specific individual. For instance, in the case of group profiling, the data processor uses the inferential process to infer a characteristic of the group that is

[213] *See* Joe O'Callaghan, *Inferential Privacy and Artificial Intelligence - A New Frontier?*, 11 J.L. & ECON. REGUL. 72, 72 (2018) (discussing the challenges that the increasing volumes of inferential data about individuals raise for the existing regulatory models in terms of privacy and data protection).

[214] *See id.* at 74-75.

[215] *Id.* at 78.

[216] *See* Blanke, *supra* note 110, at 82 (showing how the accuracy of analytics evolved from the widely publicized episode concerning Target predicting, in 2012, that a teenaged girl was pregnant to a 2013 research carried out at Cambridge University where, with startling accuracy, a number of sensitive personal attributes were drawn on the basis of Facebook likes).

[217] Bart Custers, *Profiling as Inferred Data. Amplifier Effects and Positive Feedback Loops*, *in* BEING PROFILED: COGITAS ERGO SUM: 10 YEARS OF PROFILING THE EUROPEAN CITIZEN 112, 113 (Emre Bayamlioğlu, Irina Baraluic, Liisa Janssens & Mireille Hildebrandt eds., 2018).

[218] *See id.*

then applied to all members, as well as individuals outside the group that share some of the group's characteristics.[219]

For example, data mining techniques can be used to calculate the probability of defaulting on a loan for a group of individuals who share several characteristics (such as zip code, alcohol consumption, and monthly rent expenditures).[220] Data processors will create a group profile based on this information. Further, any individual who shares characteristics with this group — though not necessarily the characteristics initially used to create this group — will be given the same default rate as the group.[221]

B. Inferred Data and Traditional Conceptions of Personal Data

Whether inferred data constitutes personal data is controversial. The legal status of inferred data is crucial because data subjects' rights over inferred data attach if and only if it is personal data. At the same time, overprotection of inferred data may run the risk of unjustifiably limiting the use of data analytics, undermining the development of AI, and stifling beneficial innovation.[222]

This Section provides background on the current state of the law about whether inferred data is personal data. This discussion sets the stage for our normative analysis that considers whether and when protecting inferred data as personal data is desirable and socially beneficial.

1. European Approach

In Europe, the legal status of inferences is uncertain and often contradictory. The CJEU's jurisprudence and Art. 29 WP documents contribute to this uncertainty by offering different sets of analyses and conclusions about whether inferred data is protected as personal data under the GDPR.

[219] Elena G. González & Paul de Hert, *Understanding the Legal Provisions that Allow Processing and Profiling of Personal Data — An Analysis of GDPR Provisions and Principles*, 19 ERA F. 597, 610 (2019).

[220] *See id.*

[221] *See id.*

[222] *See, e.g.*, Graef et al., *supra* note 16, at 10 ("Over-compliance with data protection rules might . . . harm data innovation when it is used as a cover for strategic behaviour.").

Despite a lack of clarity, the Art. 29 Working Party's guidance documents indicate that inferred data is likely personal data.[223] In the 2007 Guidance "On the Concept of Personal Data," the Art. 29 WP states that personal data covers both objective information, such as "the presence of a certain substance in one's blood," and subjective information, such as "opinions or assessments."[224] Since most instances of inferred data are simply subjective opinions or assessments created through analytical techniques — like credit scores which are opinions about creditworthiness — inferred data would likely be considered personal data.[225]

However, in the Art. 29 WP's more recent guidance on data portability,[226] inferred data is not considered "portable" like the other personal data. The lack of portability rights over inferred data seems to indicate that it may not be considered personal data because data subjects are given portability rights over other personal data.[227] Yet data subjects are still granted other information rights over inferred data such as access and rectification,[228] which indicates that inferred data stands on ambiguous and uncertain terrain regarding its status as personal data.

Similarly, the CJEU's jurisprudence on the legal status of opinions, assessments, and analysis provides contradictory guidance on the status of inferred data. For example, in *YS, M and S v. Minister voor Immigratie, Integratie en Asiel*, the CJEU interpreted personal data restrictedly, claiming that the information about the data subject in a legal analysis was personal data.[229] However, the legal analysis itself — that similar to an inference relies on underlying data to create new information — was

[223] Wachter & Mittelstadt, *supra* note 21, at 521.

[224] *Art. 29 WP Opinion Concept of Personal Data*, *supra* note 18, at 6.

[225] *See id.*

[226] Art. 29 Data Prot. Working Party, *Guidelines on the Right to Data Portability*, at 8 (Dec. 13, 2016), https://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf [https://perma.cc/78UT-W6YT] (providing direction on how to implement data portability, that is, the possibility for data subjects to obtain and reuse their personal data for their own purposes across different services).

[227] *See id.*

[228] *Id.* at 8 n.ll.

[229] *See* YS v. Minister voor Immigratie, Integratie en Asiel, ECLI:EU:C:2014:2081, ¶ 70 (July 17, 2014).

not personal data.[230] By contrast, in *Nowak*, the CJEU considered both the exam (the underlying data) and the comments and opinions about the exam (the new information derived from the underlying data) to be personal data.[231]

And finally, the GDPR also expresses ambiguity about the status of inferred data. Rather than commenting on the status of inferred data explicitly, the GDPR regulates the process of automatic profiling[232] and provides data subjects with the right to object to automated decision making based on profiles.[233] In other words, the GDPR does not regulate the profiles themselves (which are often inferred data) but instead regulates how the profiles can be used. Even though the GDPR does not explicitly consider profiling outputs (inferred data), the majority of scholars maintain that inferred data is personal data and that the full set of informational rights should apply to this information.[234] Moreover, some commentators also point out that these rights are not designed for the peculiarity of inferred data and, as a result, fail to provide sufficient control over inferred data to data subjects.[235]

[230] *See id.* This distinction is fundamental to the categorization of inferences because inferences, like legal analyses, rely on underlying data to create new information. Therefore, the status of inferred data as personal data stands and falls with the status of legal analysis as personal data.

[231] *See* Case C-434/16, Peter Nowak v. Data Prot. Comm'r, ECLI:EU:C:2017:994, ¶ 62 (Dec. 20, 2017).

[232] GDPR, *supra* note 15, art. 22. Inferring is an umbrella term identifying the process of extracting information from previous information (source data), regardless of it being information on the present (profiling) or on the future (predicting). *See generally* Hideyuki Matsumi, *Predictions and Privacy: Should There Be Rules About Using Personal Data to Forecast the Future?*, 48 CUMB. L. REV. 149 (2017) (discussing how privacy and data protection laws should apply to predictions that are based on past or present facts or predictions that forecast an unvested future).

[233] GDPR, *supra* note 15, arts. 11, 21.

[234] *See, e.g.*, Damian Clifford, Megan Richardson & Normann Witzleb, *Artificial Intelligence and Sensitive Inferences: New Challenges for Data Protection Laws*, *in* REGULATORY INSIGHTS ON ARTIFICIAL INTELLIGENCE: RESEARCH FOR POLICY 19, 38 (Mark Findlay, Jolyon Ford, Josephine Seah & Dilan Thampapillai eds., 2022) (detailing the majority position about the protection of inferences under the GDPR).

[235] *See, e.g.*, Wachter & Mittelstadt, *supra* note 21, at 494-95 (claiming for the explicit recognition of a "right to reasonable inferences").

2. American Approach

Though the debate over the status of inferred data is less developed in the United States, it is equally ambiguous and unresolved. In particular, the status of inferred data under a theory of PII differs greatly from the guidance offered by the CCPA.

Under the American approach to personal data that prioritizes identification, inferred data will be PII as long as it identifies an individual. Therefore, only some subsets of inferred data will identify particular individuals and be granted protection as PII. Moreover, the legal battles around the status of inferences will track other legal disputes about personal data within the United States; that is, disputes will focus primarily on the conditions for identification of a particular person from an inference.[236] PII focuses exclusively on the inferred data and ignores the process used to create inferences; the sole criterion is whether the inferred data identifies an individual, irrespective of whether personal data or non-personal data was used to generate the inference.

While the status of inferred data under an idealized theory of PII pivots solely around the identification requirement, the CCPA focuses on the inferential process and explicitly states that inferences derived from personal information are personal data as well.[237] More specifically, the CCPA claims that inferences drawn from personal information to create a consumer profile mapping their "preferences, behavior, psychological trends, predispositions, attitudes, intelligence, abilities, and aptitudes" are personal data.[238] The underlying rationale for protecting inferred data as personal data is that it becomes a permanent and persistent part of

[236] *See, e.g.*, Schwartz & Solove, *The PII Problem*, *supra* note 1, at 1836-41 (discussing whether IP addresses identify a person and are personal data).

[237] Some commentators assert that because California had the benefit of "a couple of more years to see where technology was headed, it was able to include much tighter definitions." Blanke, *supra* note 110, at 91. Other commentators maintain, instead, that this is the result of the Cambridge Analytica scandal. *See* Clifford et al., *supra* note 234, at 40-41.

[238] Blanke, *supra* note 110, at 90 (quoting California Consumer Privacy Act, CAL CIV. CODE § 1798.140(o)(1) (2022)).

the data subject's profile, making it impossible to distinguish between factual, verified personal data and inferred data.[239]

In the United States, therefore, the protection of inferred data as personal data is polarized around two different positions. On one side, theories of personal data guided by PII will protect inferred data as long as it identifies a person. On the other side, the CCPA adopts a much more ample protection for personal data defined in a way that also encompasses inferences. It has been noted that by the time of the CCPA's adoption, it was clear that inferential analytics would become a powerful tool to shape the economy. This cleared the way for the Californian legislature to adopt a definition of personal data that mirrors the GDPR one but also included inferred data within its scope.[240]

3. A Comparison of Inferred Data Protections in Europe and the U.S.

The status of inferred data is not fully resolved within American or European privacy and data protection frameworks. Moreover, both the United States and Europe have adopted divergent methods for resolving open questions about the status of inferred data. In Europe, the GDPR does not contemplate inferred data as such but, instead, regulates profiling and provides the right to object to algorithmic decision making.[241] In line with its spirit, the GDPR is focused on the process of inferring — in particular, the conditions for inferences and rights of action data subjects retain over the inferential process.

By contrast, the American approach — as made clear in both PII and the CCPA — targets the outcome of the inferential process (inferred data) rather than the process itself. However, a salient difference between PII and the CCPA is noteworthy. While the CCPA protects inferences as personal data when they are derived from personal data, inferences protectable as PII only need to identify an individual.[242]

In attempting to resolve the status of inferred data, both the American and European systems largely fail to capture an essential feature about the relationship between an individual and information, which is how the

[239] *See id.* at 85.

[240] *See id.* at 91-92.

[241] *See* GDPR, *supra* note 15, arts. 21-22.

[242] *See supra* Section I.B.2.

information is used. Without a focus on how inferred data is used, both regulatory regimes do not adequately adjudicate when data subjects should be able to exercise control over inferred data.

On this score, the GDPR more controversially and the CCPA more clearly protect inferred data when the data subject's interests are not at stake. Similarly, the American approach of PII would grant protection when inferences "identify" an individual, without considering different uses of inferred data.

C. Inferred Data and Separability

This Section reimagines the conditions for when data subjects should be able to exert control over inferential data. Inferences are conceptually complex because both the source data and the inference itself raise questions about the relationship between the data and the individual. Inferences can be drawn from a variety of source data (including de-identified data) and used for a variety of purposes that may or may not be related to the person who is described in the inferred data.

In this complex scenario, the debate on inferred data has pivoted around the legal status of inferred data as personal data and the tools granted to data subjects to control inferences. However, better questions are *when* and *why* inferred data is personal data. Our theory of personal data grounded in separability provides a framework to begin answering these questions and resolving some of the debates around control rights over inferred data.

1. The Conditions for Separability: Inferred Data

For inferred data to be personal data (and subject to control rights), two conditions must be met: there must be a connection and a use that depends on that connection which affects the person.[243] Considering the connection, data is often inferred to describe an individual. For instance, inferred data is used to fill a gap in a person's profile and, therefore, describes the person in some way. In this case inferred data meets the connection threshold.

On the other hand, inferred data is often created with no connection to a specific person: inferences are crafted that forecast the level of traffic

[243] *See supra* Section III.B.3.

congestion in a city along a specific route. Inferences about traffic patterns are not connected to a specific person and, as a result, fail to meet the first necessary condition for inseparability and, by extension, personal data.

Assuming that a connection is present, the second step of the analysis examines how the information is used. More specifically, the use must rely on that connection and bring about effects on the person. By carrying out this contextual analysis, we identify the instances in which the law should grant control over select uses of information, rather than granting control over the information itself, as it stands under the current privacy rules. This is because separability contends that control interests obtain at the level of use, rather than at the level of the information itself.[244] We clarify this concept more fully by considering a concrete example of how separability responds to inferred data.

2. COVID-19 Inferences and Separability

An example that can we use to sharpen the line between separable and inseparable uses of information is deriving inferences about COVID-19 infection risk from a person's travel history.

Inferences about a person's COVID-19 infection and transmission risk can be drawn from a wide array of data. In particular, inferences about COVID-19 risk may be drawn from travel history, infection rate and weather data in recent places of travel, and information about potential future travel destinations. These pieces of information, taken together, may produce a COVID-19 risk score. This process is not all too dissimilar from credit scoring, where an array of data is used to create a risk score.

The status of COVID-19 risks scores as personal data varies according to which account of personal data we invoke. A general account of PII would consider these inferences as protectable personal data as long as they identify the person, irrespective of the underlying source data that was used to create them.[245] By contrast, the status of COVID-19 risk scores under the CCPA and GDPR is more ambiguous. Under both sets of regulations, inferred data is personal data only when it is derived from the data subject's personal data, such as the data subject's travel

[244] *See supra* Section III.B.2.

[245] *See supra* Section IV.B.2.

history.[246] However, COVID-19 risk scores may not be considered personal data when they are derived from non-personal data, such as weather data or COVID-19 data from countries where the person intends to travel.

The application of a theory of personal data grounded in separability would offer a better approach to determining if COVID-19 risk scores should be protectable. Moreover, separability better tracks situations where protection is necessary by making determinations on a use-by-use basis. Whether COVID-19 risks scores are protectable as personal data under separability depends on the connection between the information and the person as well as the use.

Assuming the existence of a connection between the person and information, we can consider different uses of these risk scores and whether a person's interests are implicated. For example, a health organization could compile travel history records and create inferences about COVID-19 risk to better understand large patterns of virus transmission. Here, the fact that a specific person is described in the data is unimportant because the use is focused on broad population patterns, not individuals. As a result, this use is separable and should not be subject to individual control rights.[247]

By contrast, an algorithmic tool that infers COVID-19 risk scores could be deployed at a country's border to control access into the country. The use of COVID-19 risk score, in this case, is markedly different than the health organization's use. This is because the tool is used to determine access for the person, thus potentially affecting their ability to travel. This use of COVID-19 risk is inseparable, and by extension, is a natural fit for control rights.[248]

The current European and American frameworks on data protection and privacy fail to provide the same granularity and precision that separability provides. Under PII, the analysis exclusively focuses on identification, so the COVID-19 risk score information would be protectable in both cases.[249] As we can see, PII risks over-protection as it does not consider the significance of different uses. The CCPA and GDPR

[246] *See supra* Part IV.B.

[247] *See supra* Section III.B.3.c.

[248] *See supra* Section III.B.3.

[249] *See supra* Section IV.B.2.

run into similar problems.[250] Both the CCPA and GDPR provide protection over inferred data (here, the COVID-19 risk scores) when the information is derived from personal data. However, both the CCPA and GDPR fail to offer a rigorous analysis of when data is personal. In addition, even when the underlying information is non-personal, inferences may still describe a specific person and be used to affect their rights and interests. This result appears to be a blind spot for both the CCPA and GDPR.

This discussion demonstrates that current frameworks fail to provide adequate guidance on the status of inferred data. By contrast, a reinvigorated analysis of use through the concept of separability offers a more robust framework to parse different uses of information along more philosophically rigorous criteria.

CONCLUSION

The ideal shape of information governance is the subject of active debate and contestation. Reform efforts splinter according to whether they understand the information economy to pose principally individual or social harms. Further, whether data processing practices are understood individually or socially determines whether potential harms should be offset through a system of individual control rights or collective, democratic mechanisms.

Traditionally, information privacy law has been conceptualized through the lens of individualism.[251] Privacy and data protection violations were principally considered individual wrongs that limited the person's ability to define themselves, thus undermining their dignity and autonomy.[252] Under the traditional view, then, the risks that data collection and processing pose are best remedied through personal rights

[250] *See supra* Sections III.B.2, IV.B.2.

[251] *See generally* ALAN F. WESTIN, PRIVACY AND FREEDOM (1967) (discussing rights to privacy and protection from surveillance); Rodotà, *supra* note 56, at 78-81 (discussing data protection with regards to individual freedom and privacy); Samuel Warren & Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890) (discussing how individuals should be protected by a right to privacy).

[252] *See* Rodotà, *supra* note 56, at 78-81.

of control.[253] Lately, however, information governance proposals that prioritize individualism at both the level of harm and institutional design have become contested.[254] Salome Viljoen, for instance, rebukes individualism at the conceptual level and claims that the harms of informational capitalism are primarily social and, further, that this requires democratic — rather than individual — systems of accountability over data extraction and processing.[255]

This Article offers a way out of this debate and provides an interpretative framework that forms the basis of effective future regulation. Separability offers an analysis to identify when uses of information implicate personal interests sounding in dignity and autonomy. A theory of personal data grounded in separability distinguishes between individual and collective interests over information. More concretely, inseparable uses of information risk using the data subject as a means to an end, which undermines their dignity and autonomy. These uses fundamentally concern individuals and, by extension, should be governed as personal data subject to individual rights of control.

Conversely, separability also identifies when information use is an issue of social concern. Separable uses do not risk undermining the dignity and autonomy interests of the data subject and, thus, are suitable to be regulated through democratic systems of accountability. As it currently stands, however, existing data governance models also seek to protect collective interests over information processing through a set of individual control rights.

Consider, for instance, the use of identified information to optimize traffic patterns within a city. In both the United States and Europe,

[253] *See generally* Marc Rotenberg, *Fair Information Practices and the Architecture of Privacy* (*What Larry Doesn't Get*), 2001 STAN. TECH. L. REV. 1, 29-30 (discussing Larry Lessig's failure to address how to protect privacy in an era of rapidly changing technology).

[254] Some commentators manifest skepticism about the efficacy of individual rights of control and, further, contend that the complexity of the information economy makes it difficult for these rights to be meaningfully exercised. *See, e.g.*, Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1881 (2013) (discussing structural problems that make exercising individual information rights difficult).

[255] *See generally* Viljoen, *supra* note 142 (describing democratic approaches to data governance).

information governance laws allow an identified individual the ability to prevent information about them to be used for the collective interest of improving city life. Separability, by contrast, contends that even though information is about an individual, the use determines whether individual interests should be protected through personal rights of control. Control rights should only apply when there is an individual dignity or autonomy interest at stake; however, existing models of data governance mistakenly assume that just because an individual is described by information, control over this information is an individual interest. Separability demonstrates that this analysis is significantly more complex; different uses of information determine whether the risk is individual or social and, in turn, determine the mechanisms of ideal information governance.

The ideal system of information governance would protect both individual and social interests. While separability identifies fundamental individual interests that should be protected by personal control rights, democratic theories of data governance — such as Viljoen's — demarcate social interests that should be protected by democratic systems. Both theories are necessary for a complete system of privacy and data protection.