Columbia Business Law Review – Vol. 2019 – Issue 2

A RIGHT TO REASONABLE INFERENCES: RE-THINKING DATA PROTECTION LAW IN THE AGE OF BIG DATA AND AI

Sandra Wachter* & Brent Mittelstadt**

Big Data analytics and artificial intelligence (AI) draw non-intuitive and unverifiable inferences and predictions about the behaviors, preferences, and private lives of individuals. These inferences draw on highly diverse and feature-rich data of unpredictable value, and create new opportunities for discriminatory, biased, and invasive decision-making. Data protection law is meant to protect people's privacy, identity, reputation, and autonomy, but is currently failing to protect data subjects from the novel risks of inferential analytics. The legal status of inferences is heavily disputed in legal scholarship, and marked by inconsistencies and contradictions within and between the views of the Article 29 Working Party and the European Court of Justice (ECJ).

This Article shows that individuals are granted little control or oversight over how their personal data is used to draw inferences about them. Compared to other types of personal data, inferences are effectively "economy class" personal data in the General Data Protection Regulation (GDPR). Data subjects' rights to know about (Articles 13–15),

* Corresponding author. E-mail: sandra.wachter@oii.ox.ac.uk. Oxford Internet Institute, University of Oxford, 1 St. Giles, Oxford, OX1 3JS, UK; the Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, UK.

** Oxford Internet Institute, University of Oxford, 1 St. Giles, Oxford, OX1 3JS, UK; the Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, UK. The authors would like to thank Prof. Viktor Mayer-Schönberger and Dr. Christopher Russell for their incredibly detailed and thoughtful feedback that has immensely improved the quality of this work. The authors would also like to thank Dr. Alessandro Spina, Prof. Manfred Stelzer, Prof. Lee Bygrave, and Dr. Patrick Allo for their insightful and considerate comments from which this Article greatly benefitted.

rectify (Article 16), delete (Article 17), object to (Article 21), or port (Article 20) personal data are significantly curtailed for inferences. The GDPR also provides insufficient protection against sensitive inferences (Article 9) or remedies to challenge inferences or important decisions based on them (Article 22(3)).

This situation is not accidental. In standing jurisprudence the ECJ has consistently restricted the remit of data protection law to assessing the legitimacy of input personal data undergoing processing, and to rectify, block, or erase it. Critically, the ECJ has likewise made clear that data protection law is not intended to ensure the accuracy of decisions and decision-making processes involving personal data, or to make these processes fully transparent. Current policy proposals addressing privacy protection (the ePrivacy Regulation and the EU Digital Content Directive) and Europe's new Copyright Directive and Trade Secrets Directive also fail to close the GDPR's accountability gaps concerning inferences.

This Article argues that a new data protection right, the "right to reasonable inferences," is needed to help close the accountability gap currently posed by "high risk inferences," meaning inferences drawn from Big Data analytics that damage privacy or reputation, or have low verifiability in the sense of being predictive or opinion-based while being used in important decisions. This right would require ex-ante justification to be given by the data controller to establish whether an inference is reasonable. This disclosure would address (1) why certain data form a normatively acceptable basis from which to draw inferences; (2) why these inferences are relevant and normatively acceptable for the chosen processing purpose or type of automated decision; and (3) whether the data and methods used to draw the inferences are accurate and statistically reliable. The ex-ante justification is bolstered by an additional ex-post mechanism enabling unreasonable inferences to be challenged.

I. Introduction 4
II. From Explanations to Reasonable Inferences..... 9
A. The Novel Risks of Inferential Analytics and a Right to Reasonable Inferences.....12
III. Are Inferences Personal Data? 22
A. Three-Step Model 25
B. Subjectivity and Verifiability 27
IV. Jurisprudence of the European Court of Justice 29
A. Joined Cases C-141/12 and C-372/12: *YS and M and S*.....29
1. Inferences as Personal Data 30
2. Remit of data protection law.....35
B. Case C-434/16: *Nowak*.....39
1. Inferences as Personal Data 40
2. Remit of Data Protection Law.....41
C. Lessons from Jurisprudence of the ECJ 46
V. Protection Against Inferences Under Data Protection Law.....50
A. The Right to Know About Inferences.....51
B. The Right to Rectify Inferences.....57
C. The Rights to Object to and Delete Inferences 59
D. Protections against Sensitive Inferences.....69
1. Can Inferences be Sensitive Personal Data?...70
2. Intentionality and Reliability 73
E. The Right to Contest Decisions Based on Inferences 77
VI. Re-Aligning the Remit of Data Protection Law in the Age of Big Data: A Right to Reasonable Inferences....81
A. Justification to Establish Acceptability, Relevance and Reliability 90
B. Contestation of Unreasonable Inferences.....98
VII. Barriers to a Right to Reasonable Inferences: IP Law and Trade Secrets 100
A. Algorithmic Models and Statistical Purposes in the GDPR.....102
B. Algorithmic Models and the EU's Copyright Directive.....109
C. Algorithmic Models and Outcomes and Intellectual Property Law 114

D. Algorithmic Models and Outcomes and Trade Secrets....116
VIII. Conclusion and Recommendations....120
A. Re-Define the Remit of Data Protection Law....124
B. Focus on How Data is Evaluated, Not Just Collected....125
C. Do Not Focus Only on the Identifiability of Data Subjects....126
D. Justify Data Sources and Intended Inferences Prior to Deployment of Inferential Analytics at Scale....128
E. Give Data Subjects the Ability to Challenge Unreasonable Inferences....129

I. INTRODUCTION

Big Data analytics and artificial intelligence ("AI") draw non-intuitive and unverifiable inferences and predictions about the behaviors, preferences, and private lives of individuals. These inferences draw on highly diverse and feature-rich data of unpredictable value and create new opportunities for discriminatory, biased, and privacy-invasive profiling and decision-making.[1] Inferential analytics methods are used to infer user preferences, sensitive attributes (e.g., race, gender, sexual orientation), and opinions (e.g., political stances), or to predict behaviors (e.g., to serve advertisements). These methods can be used to nudge or manipulate us, or to make important decisions (e.g., loan or employment decisions) about us. The intuitive link between actions and perceptions is being eroded, leading to a loss of control over identity and how individuals are perceived by others. Concerns about algorithmic accountability are often actually concerns about the way in which these technologies draw privacy-invasive and non-verifiable inferences that cannot be predicted, understood, or refuted.

[1] *See* Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter & Luciano Floridi, *The Ethics of Algorithms: Mapping the Debate*, BIG DATA & SOC'Y, July–Dec. 2016, at 1–2.

Data protection law is meant to protect people's privacy, identity, reputation, and autonomy, but it is currently failing to protect data subjects from the novel risks of inferential analytics. The broad concept of personal data in Europe could be interpreted to include inferences, predictions, and assumptions that refer to or impact an individual. If seen as personal data, individuals would be granted numerous rights under data protection law. However, the legal status of inferences is heavily disputed in legal scholarship, and marked by inconsistencies and contradictions within and between the views of the Article 29 Working Party[2] and the European Court of Justice. It is crucial to note, however, that the question of whether inferences are personal data is not the most important one. The underlying problem goes much deeper and relates to the tension of whether individuals have

[2] It is worth noting that as of the implementation of the General Data Protection Regulation ("GDPR") on May 25, 2018, the Article 29 Working Party has ceased to exist and has been succeeded by the European Data Protection Board ("EDPB"). *See* European Data Prot. Bd., *The European Data Protection Board*, Endorsement 1/2018 (May 25, 2018), https://edpb.europa.eu/sites/edpb/files/files/news/endorsement_of_wp29_documents.pdf [https://perma.cc/8H9A-RQR3] One of the first acts of the EDPB was to adopt the positions and papers drafted by the Article 29 Working Party pertaining to the GDPR. For a full list of adopted documents, see *id.* Only one set of guidelines produced by the EDPB between May 25, 2018 and April 2019 are relevant to the topics addressed herein. *See* European Data Prot. Bd., *Guidelines 2/2019 on the Processing of Personal Data Under Article 6(1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects* (Apr. 8, 2019), https://privacyblogfullservice.huntonwilliamsblogs.com/wp-content/uploads/sites/28/2019/04/edpb_draft_guidelines-art_6-1-b-final_public_consultation_version_en.pdf [https://perma.cc/R3GD-J75Y]. This Article therefore continues to focus on the opinions, guidelines, and working papers of the Article 29 Working Party, which remain a key source of interpretation for the GDPR and the preceding 1995 Data Protection Directive and have proven influential in standing jurisprudence of the European Court of Justice pertaining to data protection law. It is of course likely that in the future the EDPB will adopt additional positions in support of or contradictory to the views of the Article 29 Working Party, which may be relevant to the analysis carried out here.

rights, control, and recourse concerning how they are seen by others.

This Article will show that individuals are granted little control and oversight over how their personal data is used to draw inferences about them. Compared to other types of personal data, inferences are effectively "economy class" personal data in the General Data Protection Regulation ("GDPR"). Data subjects' rights to know about (Art. 13–15), rectify (Art. 16), delete (Art. 17), object to (Art. 21), or port (Art. 20) personal data are significantly curtailed when it comes to inferences, often requiring a greater balance with the controller's interests (e.g., trade secrets or intellectual property) than would otherwise be the case. Similarly, the GDPR provides insufficient protection against sensitive inferences (Art. 9) or remedies to challenge inferences or important decisions based on them (Art. 22(3)).

This situation is not accidental. In standing jurisprudence, the European Court of Justice ("ECJ")[3] and the Advocate General ("AG")[4] have consistently restricted the remit of data protection law to assessing the legitimacy of the input stage of personal data processing, including rectification and erasure of inputs, and objecting to undesired processing.[5] Critically, the ECJ has likewise made clear that data protection law is not intended to ensure the accuracy of decisions and decision-making processes involving personal data, or to make these processes fully transparent. In short, data subjects have control over how their personal data is collected and processed, but very little control over how it is

[3] *See* Case C–28/08 P, European Comm'n v. Bavarian Lager Co., 2010 E.C.R. I–6055, ¶¶ 49–50; Case C–434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶¶ 54–55; Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶¶ 45–47.

[4] Case C–434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-582, ¶¶ 54–58; Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2013 E.C.R. I-838, ¶¶ 32, 54–60.

[5] *See, e.g.*, Case C–553/07, College van burgemeester en wethouders van Rotterdam v. M.E.E. Rijkeboer, 2009 E.C.R. I-293, ¶¶ 48–52.

evaluated. The ECJ makes clear that if the data subject wishes to challenge their evaluation, recourse must be sought through sectoral laws applicable to specific cases, not data protection law.[6]

Conflict looms on the horizon in Europe that will further weaken the protection afforded to data subjects against inferences. Current policy proposals addressing privacy protection—the ePrivacy Regulation and the EU Digital Content Directive—fail to close the GDPR's accountability gaps concerning inferences. At the same time, the GDPR and Europe's new Copyright Directive aim to facilitate data mining, knowledge discovery, and Big Data analytics by limiting data subjects' rights over personal data. And lastly, the new Trade Secrets Directive provides extensive protection of commercial interests attached to the outputs of these processes (e.g., models, algorithms and inferences).

This Article argues that a new data protection right, the "right to reasonable inferences," is needed to help close the accountability gap currently posed by "high-risk inferences," meaning inferences drawn through Big Data analytics that are privacy-invasive or reputation-damaging, or have low verifiability in the sense of being predictive or opinion-based while being used for important decisions.[7] In cases where algorithms draw "high-risk inferences" about individuals, this

---

[6] *See supra* note 3.

[7] "Important" in this context refers to the existence of "legal or similarly significant effects" resulting from a given decision. This notion is derived from Article 22(1) of the GDPR regarding automated decision-making. *See* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) art. 22(1). The precise scope of "legal or similarly significant effects" remains unclear in practice, though it will be clarified as the GDPR matures via legal commentary, national implementation, and jurisprudence. *See generally* Sandra Wachter, *Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR*, 34 COMPUTER L. & SECURITY REV. 436 (2018); Sandra Wachter, *The GDPR and the Internet of Things: A Three-Step Transparency Model*, 10 LAW INNOVATION & TECH. 266 (2018).

right would require the data controller to provide ex-ante justification to establish that the inference to be drawn is reasonable. This disclosure would address (1) why certain data form a normatively acceptable basis from which to draw inferences; (2) why these inferences are relevant and normatively acceptable for the chosen processing purpose or type of automated decision; and (3) whether the data and methods used to draw the inferences are accurate and statistically reliable. The ex-ante justification would be bolstered by an additional ex-post mechanism enabling unreasonable inferences to be challenged. A right to reasonable inferences must, however, be reconciled with EU jurisprudence and counterbalanced with intellectual property ("IP") and trade secrets law, as well as with freedom of expression[8] and Article 16 of the EU Charter of Fundamental Rights[9]—the freedom to conduct a business.

Part II first examines gaps in current work on algorithmic accountability before reviewing the novel risks of Big Data analytics and algorithmic decision-making that necessitate the introduction of a right to reasonable inferences. For such a right to be feasible under data protection law, inferences must be shown to be personal data. Part III reviews the position of the Article 29 Working Party on the legal status of inferences. Part IV then contrasts this with jurisprudence of the European Court of Justice, which paints a more restrictive picture of the scope of personal data and the remit of data protection law. Part V then assesses the current legal protection granted to inferences under European data protection laws. With the legal status and limited protection granted to inferences established, Part VI then describes the aims and scope of the proposed "right to reasonable inferences." Part VII then examines barriers likely to be

[8] *See* JORIS VAN HOBOKEN, SEARCH ENGINE FREEDOM: ON THE IMPLICATIONS OF THE RIGHT TO FREEDOM OF EXPRESSION FOR THE LEGAL GOVERNANCE OF WEB SEARCH ENGINES 316–32 (2012); *see also* JORIS VAN HOBOKEN, THE PROPOSED RIGHT TO BE FORGOTTEN SEEN FROM THE PERSPECTIVE OF OUR RIGHT TO REMEMBER (2013).

[9] Charter of Fundamental Rights of the European Union, 2000 O.J. (C364) 1.

encountered in the implementation of the proposed right, drawing from data protection law, as well as IP law and the new EU Trade Secrets Directive. In Part VIII, the Article concludes with recommendations on how to re-define the remit of data protection law to better guard against the novel risks of Big Data and AI. In the same way that it was necessary to create a "right to be forgotten" in a Big Data world,[10] it is now necessary to create a "right on how to be seen."

II. FROM EXPLANATIONS TO REASONABLE INFERENCES

Recent years have seen a flurry of work addressing explainability as a means to achieve accountability in algorithmic decision-making systems.[11] This work has taken

[10] *See generally* VAN HOBOKEN, THE PROPOSED RIGHT TO BE FORGOTTEN, *supra* note 8; *see also* VIKTOR MAYER-SCHÖNBERGER, DELETE: THE VIRTUE OF FORGETTING IN THE DIGITAL AGE (2009).

[11] *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017); Tim Miller, *Explanation in Artificial Intelligence: Insights from the Social Sciences*, ARTIFICIAL INTELLIGENCE, Feb. 2019, at 1; Brent Mittelstadt, Chris Russell & Sandra Wachter, *Explaining Explanations in AI*, *in* FAT* '19: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAT* '19), JANUARY 29–31, 2019, ATLANTA, GA, USA 279 (2019); S. C. Olhede & P.J. Wolfe, *The Growing Ubiquity of Algorithms in Society: Implications, Impacts and Innovations*, PHIL. TRANSACTIONS ROYAL SOC'Y A, Aug. 6, 2018, at 8; Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation,* 7 INT'L DATA PRIVACY L. 76 (2017); Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 841 (2018); Finale Doshi-Velez & Mason Kortz, Accountability of AI Under the Law: The Role of Explanation (Berkman Klein Ctr. Working Grp. on Explanation and the Law Working Paper, 2017); *see also* Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan.–June 2016, at 1 (describing sources of algorithmic opacity).

many forms, including calls for regulation,[12] development of technical methods of explanation[13] and auditing mechanisms,[14] and setting of standards for algorithmic accountability in public and private institutions.[15] These diverse streams of work are essential in the quest to increase AI accountability and fortunately have made much progress in legal, ethical, policy, and technical terms. Yet each is still united by a common blind spot: a legal or ethical basis is required to justify demands for explanations and determine their required content.[16] As a result, much of the prior work on methods, standards, and other scholarship around explanations will be valuable in an academic or developmental sense, but will fail to actually help the intended beneficiaries of algorithmic accountability: people affected by algorithmic decisions.

---

[12] *See, e.g.*, Marion Oswald, *Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power*, PHIL. TRANSACTIONS ROYAL SOC'Y A, Aug. 6, 2018, at 1, 3; Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83 (2017).

[13] *See, e.g.*, Mittelstadt, et al., *supra* note 11.

[14] *See, e.g.*, Brent Mittelstadt, *Auditing for Transparency in Content Personalization Systems*, 10 INT'L J. COMM. 4991 (2016); Pauline T. Kim, Essay, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189 (2017).

[15] *See, e.g.*, European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2013(INL)), EUR. PARL. DOC. P8_TA(2017)0051, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN [https://perma.cc/9H5H-W2UE]; NAT'L SCI. & TECH. COUNCIL COMM. ON TECH., EXEC. OFFICE OF THE PRESIDENT OF THE UNITED STATES, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 30–34 (2016); HOUSE OF COMMONS SCI. & TECH. COMM., HC 351, ALGORITHMS IN DECISION-MAKING 24–31, 39–40 (2018) (UK); Corinne Cath, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo & Luciano Floridi, *Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach*, 24 SCI. & ENGINEERING ETHICS 505 (2018).

[16] For an exploration of norms around explanation, see Doshi-Velez et al., *supra* note 11, at 3–6.

Unfortunately, there is little reason to assume that organizations will voluntarily offer full explanations covering the process, justification for, and accuracy of algorithmic decision-making unless obliged to do so. These systems are often highly complex, involve (sensitive) personal data, and use methods and models considered to be trade secrets. Providing explanations thus imposes additional costs and risks for the organization.

Where a general legal or ethical justification for explanations of algorithmic decisions does not exist,[17] requests will require alternative grounds to be successful.[18] This Article refers to these potential grounds for demanding information about an automated decision-making process as legal or ethical "decision-making standards." Such standards define certain procedures that must be followed in particular decision-making processes and can be enshrined in individual rights, sectoral laws, or other regulatory instruments.

Decision-making standards are not typically embedded in an absolute right that would require the full decision-making procedure to be disclosed; it remains, for example, within the private autonomy of the employer to make hiring decisions. Rather, decision-making standards provide grounds to demand limited explanations detailing the steps of a decision-making process necessary to determine whether the procedures in question were followed. So, for example, a job applicant may have a right to certain standards being followed within that procedure, such as not basing the hiring decision

[17] The GDPR's right to explanation, even if legally binding, would be limited to decision-making based solely on automated processing with legal or similarly significant effects. These conditions significantly limit its potential applicability. *See* Wachter, Mittelstadt & Floridi, *supra* note 11, at 78; *see also* Article 29 Data Prot. Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, 17/EN, WP251rev.01, at 19 (Feb. 6, 2018), http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826 (on file with the *Columbia Business Law Review*).

[18] Doshi-Velez et al., *supra* note 11, at 4, for example, suggest that demands for explanation will not be justified unless accompanied by recourse for harm suffered.

on a protected attribute (e.g., ethnicity) because doing so would constitute discrimination.

Nonetheless, granting explanations is only one possible way forward in making algorithmic decision-making accountable. Explanations can provide an effective ex-post remedy, but an explanation can be rendered only after a decision has been made.[19] An explanation might inform the individual about the outcome or decision and about underlying assumptions, predictions, or inferences that led to it. It would not, however, ensure that the decision, assumption, prediction, or inference is justified.[20] In short, explanations of a decision do not equal justification of an inference or decision. Therefore, if the justification of algorithmic decisions is at the heart of calls for algorithmic accountability and explainability, governance requires both effective ex-ante and ex-post remedies. Individual-level rights are required that would grant data subjects the ability to manage how privacy-invasive inferences are drawn, and to seek redress against unreasonable inferences when they are created or used to make important decisions.

A. The Novel Risks of Inferential Analytics and a Right to Reasonable Inferences

The following Sections explain how European law is not equipped to protect individuals against the novel risks brought on by automated decision-making driven by inferential analytics. This Article argues that a new right—a right to reasonable inferences—might help to close the accountability gap currently posed by these technologies in Europe.[21]

---

[19] *See generally* Wachter, Mittelstadt & Floridi, *supra* note 11.

[20] *See* Miller, *supra* note 11, at 8; *see also* Mireille Hildebrandt, *Primitives of Legal Protection in the Era of Data-Driven Platforms*, 2 GEO. L. TECH. REV. 252, 271 (2018).

[21] *See* Wachter, *Normative Challenges of Identification in the Internet of Things*, *supra* note 7, at 448; Wachter, *The GDPR and the Internet of Things*, *supra* note 7, at 267–71.

To explain why this new right is essential, it is first necessary to establish the source of risks in Big Data analytics and algorithmic decision-making systems. Automated decision-making, profiling, and related machine-learning techniques pose new opportunities for privacy-invasive, discriminatory, and biased decision-making based on inferential analytics.[22] Modern data analytics has access to unprecedented volumes and varieties of linked-up data to assess the behaviors, preferences, and private lives of individuals.[23] Inferences can be used to nudge and manipulate us. The range of potential victims of these harms is diversified by the focus in modern data analytics on finding small but meaningful links between individuals,[24] and constructing group profiles from personal, third-party, and anonymized data.[25]

Numerous applications of Big Data analytics to draw potentially troubling inferences about individuals and groups

---

[22] *See* Mittelstadt et al., *supra* note 1, at 7–10. *See generally* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

[23] *See generally* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2013). *See also* Brent Daniel Mittelstadt & Luciano Floridi, *The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts*, 22 SCI. & ENGINEERING ETHICS 303, 304–06 (2016); Tal Z. Zarsky, *Understanding Discrimination in the Scored Society*, 89 WASH. L. REV. 1375 (2014).

[24] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 2–4 (2014); Peter Grindrod, *Beyond Privacy and Exposure: Ethical Issues Within Citizen-Facing Analytics*, PHIL. TRANSACTIONS ROYAL SOC'Y A, Dec. 28, 2016, at 10–12.

[25] *See* Alessandro Mantelero, *From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era*, *in* GROUP PRIVACY: NEW CHALLENGES OF DATA TECHNOLOGIES 139, 145 (Linnet Taylor, Luciano Floridi & Bart van der Sloot eds., 2017); Brent Mittelstadt, *From Individual to Group Privacy in Big Data Analytics*, 30 PHIL. & TECH. 475, 476 (2017).

have emerged in recent years.[26] Major internet platforms are behind many of the highest profile examples: Facebook may be able to infer sexual orientation—via online behavior[27] or based on friends[28]—and other protected attributes (e.g., race),[29] political opinions[30] and sadness and anxiety[31] – all of these inferences are used for targeted advertising. Facebook can also infer imminent suicide attempts,[32] while third parties have used Facebook data to infer socioeconomic

[26] *See, e.g.*, Christopher Kuner, Fred H. Cate, Christopher Millard & Dan Jerker B. Svantesson, *The Challenge of "Big Data" for Data Protection*, 2 INT'L DATA PRIVACY L. 47 (2012).

[27] *See* José González Cabañas, Ángel Cuevas & Rubén Cuevas, Facebook Use of Sensitive Data for Advertising in Europe (Feb. 14, 2018) (unpublished manuscript), https://arxiv.org/abs/1802.05030 [https://perma.cc/V2C8-FY3W].

[28] Carter Jernigan & Behram F.T. Mistree, *Gaydar: Facebook Friendships Expose Sexual Orientation*, FIRSTMONDAY.ORG (Oct. 5, 2009), https://firstmonday.org/ojs//index.php/fm/article/view/2611 [https://perma.cc/AMK2-QB8U].

[29] Annalee Newitz, *Facebook's Ad Platform Now Guesses at Your Race Based on Your Behavior*, ARS TECHNICA (Mar. 18, 2016), https://arstechnica.com/information-technology/2016/03/facebooks-ad-platform-now-guesses-at-your-race-based-on-your-behavior/ [https://perma.cc/H6SB-MSAE].

[30] Jeremy B. Merrill, *Liberal, Moderate or Conservative? See How Facebook Labels You*, N.Y. TIMES (Aug. 23, 2016), https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html [https://perma.cc/QNU7-YCBZ].

[31] Michael Reilly, *Is Facebook Targeting Ads at Sad Teens?*, MIT Tech. Rev. (May1, 2017), https://www.technologyreview.com/s/604307/is-facebook-targeting-ads-at-sad-teens/ (on file with the *Columbia Business Law Review*).

[32] Josh Constine, *Facebook Rolls Out AI to Detect Suicidal Posts Before They're Reported*, TECHCRUNCH (Nov. 27, 2017), https://techcrunch.com/2017/11/27/facebook-ai-suicide-prevention/?guccounter=1 [https://perma.cc/QF62-WJEH].

status[33] and stances on abortion.[34] Insurers are starting to use social media data to set premiums,[35] which is troublesome because research suggests that a person's social network can be used to draw acute and intimate inferences about one's personality.[36] Tendencies to depression can be inferred through Facebook[37] and Twitter[38] usage; Google has attempted to predict flu outbreaks[39] as well as other diseases and their outcomes[40]; and Microsoft can likewise predict

---

[33] *See* Astra Taylor & Jathan Sadowski, *How Companies Turn Your Facebook Activity into a Credit Score*, THE NATION (May 27, 2015), https://www.thenation.com/article/how-companies-turn-your-facebook-activity-credit-score/ [https://perma.cc/V4V5-7H55].

[34] *See* Sharona Coutts, *Anti-Choice Groups Use Smartphone Surveillance to Target 'Abortion-Minded Women' During Clinic Visits*, REWIRE (May 25, 2016), https://rewire.news/article/2016/05/25/anti-choice-groups-deploy-smartphone-surveillance-target-abortion-minded-women-clinic-visits/ [https://perma.cc/VE5A-D5S9].

[35] Leslie Scism, *New York Insurers Can Evaluate Your Social Media Use—If They Can Prove Why It's Needed*, WALL ST. J. (Jan. 30, 2019), https://www.wsj.com/articles/new-york-insurers-can-evaluate-your-social-media-useif-they-can-prove-why-its-needed-11548856802 (on file with the *Columbia Business Law Review*).

[36] *See* Kristen M Altenburger & Johan Ugander, *Monophily in Social Networks Introduces Similarity among Friends-of-Friends*, NATURE HUMAN BEHAVIOUR, Apr. 2018, at 284.

[37] *See* Megan A. Moreno et al., *Feeling Bad on Facebook: Depression Disclosures by College Students on a Social Networking Site*, 28 DEPRESSION & ANXIETY 447 (2011).

[38] *See* Moin Nadeem, Mike Horn, Glen Coppersmith & Sandip Sen, Identifying Depression on Twitter (July 25, 2016) (unpublished manuscript), https://arxiv.org/abs/1607.07384 [https://perma.cc/SKB6-WT6K].

[39] Donald R. Olson, Kevin J. Konty, Marc Paladini, Cecile Viboud & Lone Simonsen, *Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales*, PLOS COMPUTATIONAL BIOLOGY, Oct. 2013, at 1.

[40] *See* Anthony Cuthbertson, *Google AI Can Predict When People Will Die with '95 Per Cent Accuracy'*, INDEPENDENT (June 19, 2018), https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-ai-predict-when-die-death-date-medical-brain-deepmind-a8405826.html [https://perma.cc/D7RR-Y2M4]; Alvin Rajkomar et al., *Scalable and*

Parkinson's disease[41] and Alzheimer's disease[42] from search engine interactions. Amazon's Alexa might be able to infer health status based on speech patterns.[43] Other recent potentially invasive applications[44] include Target's prediction of pregnancy in customers,[45] researchers inferring levels of user satisfaction with search results using mouse tracking,[46] and, finally, China's far-reaching social credit scoring system.[47]

Accurate Deep Learning with Electronic Health Records, NPJ DIGITAL MED., May 8, 2018, at 2–4.

[41] *See* Ryen W. White, P. Murali Doraiswamy & Eric Horvitz, *Detecting Neurodegenerative Disorders from Web Search Signals*, NPJ DIGITAL MED., Apr. 23, 2018, at 1, 3; Liron Allerhand, Brit Youngmann, Elad Yom-Tov & David Arkadir, Detecting Parkinson's Disease from Interactions with a Search Engine: Is Expert Knowledge Sufficient? 1 (May 3, 2018) (unpublished manuscript), https://arxiv.org/abs/1805.01138 [https://perma.cc/SF5A-4VTW].

[42] *See* White, Doraiswamy & Horvitz, *supra* note 41.

[43] James Cook, *Amazon Patents New Alexa Feature That Knows When You're Ill and Offers You Medicine*, TELEGRAPH (Oct. 9, 2018), https://www.telegraph.co.uk/technology/2018/10/09/amazon-patents-new-alexa-feature-knows-offers-medicine/ [https://perma.cc/V346-HFWE].

[44] For an interesting overview of applications that infer sensitive information, see Christopher Burr, Nello Cristianini & James Ladyman, *An Analysis of the Interaction Between Intelligent Software Agents and Human Users*, 28 MINDS & MACHINES 735 (2018).

[45] *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [https://perma.cc/7Y84-6MWW]; MAYER-SCHÖNBERGER & CUKIER, *supra* note 23, at 57–58.

[46] Ye Chen, Yiqun Liu, Min Zhang & Shaoping Ma, *User Satisfaction Prediction with Mouse Movement Information in Heterogeneous Search Environment*, 29 IEEE TRANSACTIONS ON KNOWLEDGE & DATA ENGINEERING 2470 (2017).

[47] Simon Denyer, *China's Plan to Organize Its Society Relies on 'Big Data' to Rate Everyone*, WASH. POST (Oct. 22, 2016), https://www.washingtonpost.com/world/asia_pacific/chinas-plan-to-organize-its-whole-society-around-big-data-a-rating-for-everyone/2016/10/20/1cd0dd9c-9516-11e6-ae9d-0030ac1899cd_story.html [https://perma.cc/Z3KP-KK2T]. For a discussion of the challenges of regulating uses of non-traditional data, such as data generated by Internet of Things devices, for credit and similar decisions, see Scott R. Peppet,

None of these applications can claim to generate inferences or predictions with absolute certainty, and in several cases, they have suffered highly visible failures (e.g. Google Flu Trends).[48] Many are likewise used solely for targeted advertising. Justification for these invasive uses of personal data is crucial from an ethical[49] as well as legal[50] viewpoint to avoid inferential analytics that are privacy-invasive or damaging to reputation, particularly when these inferences are poorly verifiable or affected individuals receive no benefit. It is thus increasingly common to deploy inferential analytics at scale, based solely on the ability to do so and the perceived accuracy of the method or a belief that efficiency or revenue will improve.

From the perspective of the individual, the potential value and insightfulness of data generated while using digital technologies is often opaque. Counterintuitive and unpredictable inferences can be drawn by data controllers, without individuals ever being aware,[51] thus posing risks to

Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security and Consent, 93 TEX. L. REV. 85 (2014)

[48] *See* David Lazer, Ryan Kennedy, Gary King & Alessandro Vespignani, *The Parable of Google Flu: Traps in Big Data Analysis*, 343 SCIENCE 1203 (2014).

[49] For ethical approaches to AI accountability and justification, see Reuben Binns, *Algorithmic Accountability and Public Reason*, 31 PHIL. & TECH. 543, 548–52 (2018); Hildebrandt, *supra* note 20.

[50] Viktor Mayer-Schönberger & Yann Padova, *Regime Change? Enabling Big Data Through Europe's New Data Protection Regulation*, 17 COLUM. SCI. & TECH. L. REV. 315, 332 (2016) (considering moving away from consent-based data protection to governance of fair and ethical data uses); *see also* Alessandro Mantelero, *The Future of Consumer Data Protection in the E.U. Re-Thinking the "Notice and Consent" Paradigm in the New Era of Predictive Analytics*, 30 COMPUTER L. & SECURITY REV. 643, 653–55 (2014).

[51] *See* Mittelstadt & Floridi, *supra* note 23, at 312–13; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines,* 87 FORDHAM L. REV. 1085 (2018).

privacy[52] and identity,[53] data protection, reputation,[54] and informational self-determination.[55] As Tene and Polonetsky argue, "[i]n a big data world, what calls for scrutiny is often not the accuracy of the *raw data* but rather the accuracy of the *inferences* drawn from the data."[56] The Article 29 Working Party has recognised a similar challenge, arguing that, "[m]ore often than not, it is not the information collected in itself that is sensitive, but rather, the inferences that are drawn from it and the way in which those inferences are drawn, that could give cause for concern."[57] The European Data Protection Supervisor (EDPS) has likewise expressed concern over the privacy risks of inferences and the need for governance.[58] Similarly, NGOs and activist groups are aware

[52] Paul Ohm, *The Fourth Amendment in a World Without Privacy*, 81 MISS. L.J. 1309, 1316–18 (2012); *see also* Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857 (2017).

[53] Luciano Floridi, *The Informational Nature of Personal Identity*, 21 MINDS & MACHINES 549, 550 (2011); Mittelstadt, *supra* note 25, at 476.

[54] Sandra Wachter, Privacy: Primus Inter Pares—Privacy as a Precondition for Self-Development, Personal Fulfilment and the Free Enjoyment of Fundamental Human Rights (Jan. 22, 2017) (unpublished manuscript), https://papers.ssrn.com/abstract=2903514 [https://perma.cc/RLB3-G6SC].

[55] Urteil des Ersten Senats vom BVerfG [Volkszählungsurteil'], '15, Dezember 1983, 1 BvR 209/83 (Ger.), https://openjur.de/u/268440.html [https://perma.cc/DRS7-HNRZ]; Judgement of German Constitutional Court, BVerfG · Urteil vom 15. Dezember 1983 · Az. 1 BvR 209/83, 1 BvR 484/83, 1 BvR 420/83, 1 BvR 362/83, 1 BvR 269/83, 1 BvR 440/83 (Volkszählungsurteil). For a critical voice on this subject see Jan Klabbers, *The Right to Be Taken Seriously: Self-Determination in International Law*, 28 HUM. RTS. Q. 186 (2006).

[56] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 270 (2013) (emphasis in original).

[57] Article 29 Data Prot. Working Party, *Opinion 03/2013 on Purpose Limitation*, at 47, 00569/13/EN, WP203 (Apr. 2, 2013), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf [https://perma.cc/X6PC-825X].

[58] *See* European Data Prot. Supervisor, *EDPS Opinion on Online Manipulation and Personal Data* at 5, 8–16, Opinion 3/2018 (Mar. 19, 2018),

of these concerns and have recently submitted numerous complaints to fight for more clarity on the legal and ethical acceptability of inferential analytics.[59]

The unpredictability of the analytics behind automated decision-making and profiling can itself be harmful to individuals. As noted in jurisprudence of the European Court of Human Rights ("ECHR")[60], the use of untraditional data sources to make unpredictable and counterintuitive inferences about people can impact on the freedom of expression, the right to privacy and identity,[61] and self-determination of individuals.[62] The ECHR has a long-standing tradition of linking the right to personality to the

https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf [https://perma.cc/3KJ6-VSUD].

[59] Johnny Ryan, *Regulatory Complaint Concerning Massive, Web-Wide Data Breach by Google and Other "Ad Tech" Companies Under Europe's GDPR*, BRAVE (Sept. 12, 2018), https://www.brave.com/blog/adtech-data-breach-complaint/ [https://perma.cc/3DFW-JZTX]; *Our Complaints against Acxiom, Criteo, Equifax, Experian, Oracle, Quantcast, Tapad*, PRIVACY INT'L (Nov. 8, 2018), http://privacyinternational.org/advocacy-briefing/2426/our-complaints-against-acxiom-criteo-equifax-experian-oracle-quantcast-tapad (on file with the *Columbia Business Law Review*); *Privacy International Files Complaints Against Seven Companies for Wide-Scale and Systematic Infringements of Data Protection Law*, PRIVACY INT'L (Nov. 8, 2018), http://privacyinternational.org/press-release/2424/privacy-international-files-complaints-against-seven-companies-wide-scale-and (on file with the *Columbia Business Law Review*).

[60] For an overview on the jurisprudence on the right of privacy of the ECHR to 2017, see Council of Europe, *Case Law of the European Court of Human Rights Concerning the Protection of Personal Data*, T-PD(2017)23 (2017), https://rm.coe.int/case-law-on-data-protection/1680766992 [https://perma.cc/H4F2-9WVZ].

[61] For an in-depth discussion on identity and profiling, see PROFILING THE EUROPEAN CITIZEN (Mireille Hildebrandt & Serge Gutwirth eds., 2008); Antoinette Rouvroy, *Privacy, Data Protection, and the Unprecedented Challenges of Ambient Intelligence*, 2 STUD. ETHICS, L., & TECH. 1, 3–4 (2008).

[62] Nora Ni Loideain, *Surveillance of Communications Data and Article 8 of the European Convention on Human Rights*, *in* RELOADING DATA PROTECTION 183, 199–200, 202–03 (Serge Gutwirth, Ronald Leenes & Paul De Hert eds., 2014); Wachter, *supra* note 54, at 5.

right of privacy.[63] This link suggests that, to remain in control of their identity in the face of uncertainty, data subjects may alter their behavior (e.g. self-censorship) when using digital technologies.[64] Such chilling effects linked to automated decision-making and profiling undermine self-determination and freedom of expression and thus warrant more control over the inferences that can be drawn about an individual. Without greater control, inferences can operate as "an autonomy trap."[65] Therefore, there is also a public and collective interest in the protection of privacy.[66]

The tendency in mature information societies[67] to create, share, sell, and retain data, profiles, and other information about individuals presents additional challenges. Persistent records can be created through inferential analytics, consisting of unpredictable and potentially troubling inferences revealing information and predictions about private life, behaviors, and preferences that would otherwise remain private.[68] Compared to prior human and bureaucratic decision-making, the troubling change posed by the widespread deployment of Big Data analytics is that the profile or information "at the basis of the choice architecture

[63] *See generally* Wachter, *supra* note 54. For a critical view on guidelines of the Council of Europe's new privacy guidelines, see Alessandro Mantelero, *Regulating Big Data. The Guidelines of the Council of Europe in the Context of the European Data Protection Framework*, 33 COMP. L. & SECURITY REV. 584 (2017).

[64] PEN AMERICA, CHILLING EFFECTS: NSA SURVEILLANCE DRIVES U.S. WRITERS TO SELF-CENSOR 3–4 (2013); Jonathon W. Penney, Chilling Effects: Online Surveillance and Wikipedia Use (Sept. 8, 2016), https://papers.ssrn.com/abstract=2769645 [https://perma.cc/FGW8-WMVP].

[65] Tal Z. Zarsky, *"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE J.L. & TECH. 1, 35 (2002–03).

[66] *See generally* Priscilla M. Regan, *Privacy as a Common Good in the Digital World*, 5 INFO., COMM'N. & SOC'Y 382 (2002).

[67] Luciano Floridi, *Mature Information Societies—a Matter of Expectations*, 29 PHIL. & TECH. 1, 1 (2016).

[68] *See generally* Mittelstadt & Floridi, *supra* note 23.

offered" to individuals need not be held and used by a single third-party for a specific purpose, but rather "persists over time, travels with the person between systems and affects future opportunities and treatment at the hands of others."[69] These tendencies contribute to the solidification of identity and reputation, undermining the individual's right "to be allowed to experiment with one's own life, to start again, without having records that mummify one's personal identity forever."[70] Inferential analytics thus pose substantial and novel risks not only to identity, but to reputation and the choices offered to an individual by data-driven services.

While the potential harms of inferences have been recognized by European legal scholars and policy-makers, data protection law and its procedural approach have not yet caught up. Data subjects receive little help in coming to terms with the informativeness of the data they provide to controllers, who are generally not legally obligated to disclose or justify their criteria and methods used to draw inferences and make decisions based upon them.[71] Rather, the default procedural approach in European data protection law to protect the privacy of individuals is to grant oversight and control over how personal data is collected and processed. In other words, data protection law focuses primarily on mechanisms to manage the input side of processing. As will be explained below,[72] the few mechanisms in European data protection law that address the *outputs* of processing, including inferred and derived data, profiles, and decisions, are far weaker.

In the age of Big Data analytics, a myopic focus on input data in data protection law is troubling. The outputs of processing pose risks to individuals, yet data subjects are granted far less control over how these outputs are produced and used. Currently, individuals are not guaranteed

---

[69] Mittelstadt, *supra* note 25, at 482.

[70] Luciano Floridi, *Four Challenges for a Theory of Informational Privacy*, 8 ETHICS & INFO. TECH. 109, 112 (2006).

[71] *See infra* Part IV. *See generally* Tene & Polonetsky, *supra* note 56. (arguing that decision-making criteria of companies should be disclosed.)

[72] *See infra* Parts IV, V.

awareness of potentially problematic decision-making and will often lack a legal basis to examine the decision-making process for problems in the first place. This situation is a result of the uncertain legal status of inferences and the scope of applicable control mechanisms in data protection law. Transparency and consent mechanisms designed to manage input data are no longer sufficient; rather, the spread of inferential Big Data analytics requires a reaction in data protection law, by which meaningful control and choice over inferences and profiles are granted to data subjects.[73] As Judge Posner eloquently argues, "A seldom-remarked corollary to a right to misrepresent one's character is that others have a legitimate interest in unmasking the deception."[74] This Article argues that the introduction of a right to reasonable inferences is precisely the type of reaction required.

III. ARE INFERENCES PERSONAL DATA?

To grant data subjects broadly applicable, non-sectoral rights over their inferences under data protection law, inferences must be seen as personal data. This Part defines inferences as information relating to an identified or identifiable natural person created through deduction or reasoning rather than mere observation or collection from the data subject. The type of inference discussed here are "high risk inferences" which are created or used by data controllers or third parties, are privacy-invasive or harmful to reputation—or have a high likelihood of being so in the future—or have low verifiability in the sense of being predictive or opinion-based while being used for important

[73] *See* Serge Gutwirth & Paul De Hert, *Regulating Profiling in a Democratic Constitutional State*, *in* PROFILING THE EUROPEAN CITIZEN 271 (Mireille Hildebrandt & Serge Gutwirth eds., 2008); *see* also Ronald Leenes, *Addressing the Obscurity of Data Clouds*, *in* PROFILING THE EUROPEAN CITIZEN 293 (Mireille Hildebrandt & Serge Gutwirth eds., 2008) (also discussing the need for transparent decision-making processes).

[74] Richard A Posner, *The Right of Privacy*, 12 GA. L. REV. 393, 395 (1978).

decisions.[75] Several distinctions between "types" of personal data relevant to the legal status of inferences are evident in the GDPR itself as well as guidance issued by the Article 29 Working Party. Article 4 of the GDPR defines personal data as "any information relating to an identified or identifiable natural person."[76] Article 9(1) of the GDPR makes a further distinction between normal or non-sensitive personal data, and "special categories" of personal data that pertain to "racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation[.]"[77] Sensitive personal data incurs additional restrictions on processing under Article 9(2–4).[78] If inferences are personal data, this distinction between sensitive and non-sensitive types, and the higher standard of protection afforded to the former, will also apply.

The Article 29 Working Party further distinguishes between provided and observed data on the one hand, and derived and inferred data on the other.[79] Provided data includes any data that the data subject has directly provided to the data controller, for example the user's name or email address.[80] Observed data is also "provided by" the data subject, but indirectly or passively, including things such as

---

[75] *See supra* Section II.A; *see also infra* Part VI.

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

[77] *Id.* at art. 9(1).

[78] *Id.* at art. 9(2–4).

[79] Art. 29 Data Prot. Working Party, supra note 17, at 8; Article 29 Data Prot. Working Party, *Guidelines on the Right to Data Portability*, 16/EN, WP242rev.01, at 9–11 (Dec. 13, 2016), https://ec.europa.eu/newsroom/document.cfm?doc_id=44099 (on file with the *Columbia Business Law Review*).

[80] *Id.* at 9.

location data, clicking activity, or unique aspects of a person's behavior such as handwriting, keystrokes, or a particular way of walking or speaking.[81] In contrast, derived (e.g. country of residency derived from the subject's postcode) and inferred data (e.g. credit score, outcome of a health assessment, results of a personalization or recommendation process) are not "provided by" the data subject actively or passively, but rather created by a data controller or third party from data provided by the data subject and, in some cases, other background data.[82] The Article 29 Working Party's guidelines on data portability provide examples of personal data derived from non-traditional sources, such as data produced "from the observation of [a user's] behaviour," including clicking or browsing behavior and the inferences drawn from it.[83] Additionally, their guidelines on profiling and automated decision-making argue that "profiling . . . works [by] creating derived or inferred data about individuals – 'new' personal data that has not been provided directly by the data subjects themselves."[84] Clearly, if inferences can be considered personal data, they are of the latter type: derived or inferred.[85]

[81] Article 29 Data Prot. Working Party, *Opinion 4/2007 on the Concept of Personal Data*, 01248/07/EN WP136, at 8 (June 20, 2007) http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (on file with the *Columbia Business Law Review*).

[82] Article 29 Data Prot. Working Party, *supra* note 79, at 10–11.

[83] *See id* at 10, 10 n.20, 21. Note that inferences are not covered by Article 20, but rather by Article 15.

[84] *See supra* note 17, at 9; *see also* note 79, at 9–10 (referring to "observed data" such as "activity logs, history of website usage or search activities.").

[85] *See* Martin Abrams, The Origins of Personal Data and its Implications for Governance (Nov. 24, 2014) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2510927 [https://perma.cc/9YZ5-FT96] (discussing the differences between derived and inferred data).

A. Three-Step Model

To determine whether data is "personal data," the Article 29 Working Party[86] has proposed a three-step model. According to this model, the content, purpose, or result[87] of the data (processing) must relate to an identifiable person either directly or indirectly.[88] This approach allows for non-personal data to be transformed into personal data through linkage to an identified individual.[89] For example, the value of a house can become personal data used to assess individuals, such as the amount of their tax obligations.[90] Due to technical affordances, some commentators have argued that it is difficult to locate data that cannot potentially be transformed into personal data.[91]

The third step of the model, 'result', is key to the legal status of inferences.[92] The Article 29 Working Party argues that data being "likely to have an impact on a certain person's

---

[86] *See generally* Article 29 Data Prot. Working Party *supra* note 81; for an overview of EU jurisprudence on the definition of personal data, see Nadezhda Purtova, *The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law*, 10 LAW INNOVATION & TECH 40 (2018).

[87] *See* Article 29 Data Prot. Working Party, *supra* note 79, at 10 (defining purpose as "to evaluate, treat in a certain way or influence the status or behaviour of an individual.")

[88] *See id.* at 11.

[89] For an excellent overview of the concept of personal data, see Douwe Korff, Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments (Eur. Comm'n. Directorate-General Justice, Freedom & Sec., Working Paper No. 2, 2010), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1638949 [https://perma.cc/8JUL-5S6L].

[90] *See* Article 29 Data Prot. Working Party, *supra* note 81, at 9.

[91] Stefan Ernst, *Begriffsbestimmungen*, *in*, DATENSCHUTZ-GRUNDVERORDNUNG BUNDESDATENSCHUTZGESETZ (Boris Paal & Daniel A Pauly eds., 2018).

[92] *See* Korff, *supra* note 89, at 52–53. (arguing that profiles, understood as bundles of inferences and assumptions, should be treated as personal data.)

rights and interests"[93] is sufficient for it to be treated as personal data. In practice, this means that even if the data does not directly describe an identifiable person ("content"), or is not "used or . . . likely to be used . . . [to] evaluate, treat in a certain way or influence the status or behaviour"[94] of the person ("purpose"), it can still be classified as "personal data" based on its potential impact on an identifiable person's rights and interests ("result").[95] Information that is not directly readable from the data collected, but rather derived or inferred from it, can thus also be considered personal data.

This conclusion is further supported by the usage of the term "any information" in Article 4(1) of the GDPR; identical language was used to define "personal data" in the 1995 Data Protection Directive (95/46/EC), which the Article 29 Working Party has previously taken as evidence of legislators' intent to have a very wide definition of "personal data".[96] They argue that personal data includes 'subjective' "information, opinions, or assessments"[97] relating to an identified or identifiable natural person in terms of content, purpose, or result. Further, such information does not need to be "true or proven."[98] This position is implicitly supported by the Article 29 Working Party granting rights to data subjects "to access that information and to challenge it through appropriate remedies,"[99] for example by providing additional comments.[100] Several other guidelines issued by the Working Party similarly argue that certain individual rights apply to inferred and derived data, which by definition means these must be personal data.[101]

---

93 *See* Article 29 Data Prot. Working Party, *supra* note 81, at 11.

94 *Id.* at 10.

95 *Id.* at 10–11.

96 *Id.* at 4.

97 *Id.* at 6.

98 *Id.*

99 *Id.*

100 *Id.* at 6 n.5

101 *See Guidelines*, *supra* note 17, at 17–18. *Guidelines* clarifies that the rights to rectification, erasure, and restriction of processing apply to

B. Subjectivity and Verifiability

Inferences are often precisely these types of subjective and non-verifiable "information, opinions, or assessments"[102] created by a third-party through more than mere observation of the data subject. Several examples of such subjective or non-verifiable personal data are provided by the Article 29 Working Party. Concerning subjectivity, examples of subjective assessments are provided for several sectors: in banking, "assessment of the reliability of borrowers ("Titius is a reliable borrower"); in insurance ("Titius is not expected to die soon") or in employment ("Titius is a good worker and merits promotion")."[103] Such subjective third-party assessments can be considered a type of inference, as the

inferred and derived data. *Id.*; *see also supra* 81, at 11. Here, following the text of Article 20(1) of the GDPR, they clarify that the right to data portability covers only data "provided by" the data subject: "a personalisation or recommendation process, by user categorisation or profiling are data which are derived or inferred from the personal data provided by the data subject, and are not covered by the right to data portability." Derived and inferred data thus do not fall within the scope of data portability. In practice, this means that Art. 20 only covers data provided by the data subject or observed by the controller but not the profile itself or other inferred and derived data. This could be taken to suggest that derived and inferred data are not a type of personal data on the basis that an individual data protection right (Art. 20), which by definition applies to personal data, does not apply to these types of data. This interpretation is incorrect. Footnote 20 accompanying the preceding quote clarifies that although Art. 20 does not apply, Art. 15 and 22 still apply to inferred and derived data. By definition, for these other Articles to apply, the data being processed needs to be personal data. The Guidelines therefore endorse classifying inferred and derived data as personal data, albeit indirectly. These limits on data portability are sensible, as the right is designed as a competition tool, not a data privacy tool. *See also* Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay & Ignacio Sanchez, *The Right to Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services*, COMPUTER L. & SECURITY REV. 193 (2018).

102 *See* Article 29 Data Prot. Working Party, *supra* note 81, at 6.

103 *Id*.

assessment involves inferring a non-observed characteristic or subjective opinion of the subject from data already held[104]

Concerning non-verifiability, a second example is provided of a child's drawing depicting her family and her mood towards them.[105] Such a drawing, although created by the child, can allow for information about the behaviors of the child's parents to be inferred. As a result, the drawing itself, and any information about her parents' behavior inferred from it, is classified as the parents' personal data. Such inferences are not necessarily verifiable, and are subjective due to interpretation being required to derive information about the parents' behaviors.[106]

Each of these examples shows that the Article 29 Working Party believes opinions and assessments, understood here as inferences, do not need to be objective or verifiable to be considered personal data. Several legal commentators have reached similar conclusions. Ernst, for example, argues that predictions and inferences about a data subject constitute personal data irrespective of their timeframe or whether they address the past, present, or future.[107] By definition, predictions cannot be verified at the time they are made, but can nonetheless describe an identified or identifiable person. Klabunde similarly believes that assumptions and assessments are also personal data, irrespective of whether they are accurate or verifiable.[108]

---

104 For a discussion of opinions and assessments being classified as personal data under EU data protection law, see generally Korff, *supra* note 89.

105 *See* Article 29 Data Prot. Working Party, *supra* note 81, at 8. As such, the child's parents can exercise their right of access in relation to the drawing. *Id.*

106 *Id.*

107 Ernst, *supra* note 91, at 14–18.

108 Achim Klabunde, *Begriffsbestimmungen*, *in* DATENSCHUTZ-GRUNDVERORDNUNG BUNDESDATENSCHUTZGESETZ 7–8 (Eugen Ehmann & Martin Selmayr eds., 2017).

IV. JURISPRUDENCE OF THE EUROPEAN COURT OF JUSTICE

While the legally non-binding guidelines of the Article 29 Working Party clearly endorse the view that inferences are personal data, the legally binding jurisprudence of the European Court of Justice (ECJ) is less generous in its interpretation. Even though the ECJ also believes in a broad interpretation of the concept of personal data, the Court has historically held a more restricted view of the scope of "personal data" and applicable rights.[109] Two recent cases (YS. and M. and S.[110], and Nowak[111]) are particularly relevant to determining the legal status of inferences and the remit of data protection law more broadly.

A. Joined Cases C-141/12 and C-372/12: *YS and M and S*

YS and M and S addressed whether an applicant has a right to access the legal analysis (or "information about the assessment and application"[112]) underlying a decision of legal residency. The ECJ's judgement[113] and the associated opinion

---

[109] For an in-depth overview of the ECJ's concept of personal data, see Case C-101/01 Lindqvist [2003] E.C.R. I-12971, ¶ 24; Joined Cases C-465/00, C-138/01 and C-139/01 Österreichischer Rundfunk and Others [2003] E.C.R. I-4989, ¶ 64; Case C-73/07 Satakunnan Markkinapörssi and Satamedia [2008] E.C.R. I-9831, ¶¶ 35, 37; Case C-524/06 Huber [2008] E.C.R. I-9705, ¶ 43; and Case C-553/07 Rijkeboer [2009] E.C.R. I-3889, ☻0 62.

[110] *See supra* notes 3–4 and accompanying text.

[111] *See supra* notes 3–4 and accompanying text.

[112] Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I- 2081, ¶ 40.

[113] For in-depth analyses of the judgment, see Evelien Brouwer & Frederik Zuiderveen Borgesius, *Access to Personal Data and the Right to Good Governance During Asylum Procedures after the Cjeu's YS. and M. and S. Judgment*, 17 EUR. J. MIGRATION & L. 259 (2015); Xavier Tracol, *Back to Basics: The European Court of Justice Further Defined the Concept of Personal Data and the Scope of the Right of Data Subjects to Access It*, 31 COMPUTER L. & SECURITY REV. 112 (2015); *see also* Purtova, *supra* note 86.

of the Advocate General[114] in this case suggest a troubling direction of travel for the protection of data subjects for three reasons: (1) the limited scope of personal data; (2) the limited rights of access and rectification; and (3) the view that data protection law does not aim to ensure accurate or lawful decision-making, and thus does not govern how inferences are drawn in decision-making processes.

1. Inferences as Personal Data

The ECJ ruled "that the data relating to the applicant for a residence permit contained in the minute [a document containing the reasoning of the case officer] and, where relevant, the data in the legal analysis contained in the minute are 'personal data' within the meaning of that provision, whereas, by contrast, that analysis cannot in itself be so classified."[115] This ruling indicates that only the personal data contained or used within the legal analysis, but not the analysis itself, is personal data subject to protection under the 1995 Data Protection Directive. Specifically, the ECJ noted that only the "name, date of birth, nationality, gender, ethnicity, religion and language of the applicant,"[116] or only data that is "about" the data subject are personal data.[117]

This judgement is interesting because historically the Court has been predominantly asked to rule on the legal status of observations or verifiable data (e.g. "facts" about a person), not assessments or non-verifiable data.[118] Examples of personal data named in prior judgements include "telephone [numbers], and information about his/her working

[114] *See generally supra* note 4.

[115] *See* Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I- 2081, ¶ 48.

[116] Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I- 2081, ¶ 38.

[117] *See* Purtova, *supra* note 86, at 28.

[118] Of course, one must keep in mind that the Court can only rule on the cases referred to it, and thus the Court has no power to take views that fall outside the cases it considers.

conditions or hobbies,"[119] "the surname and given name of certain natural persons whose income exceeds certain thresholds" as well as "their earned and unearned income,"[120] "IP addresses,"[121] "fingerprints,"[122] "record of working time . . . and . . . rest periods,"[123] "data . . . collected by . . . private detectives,"[124] "image of a person recorded by a camera,"[125] "tax data,"[126] and "press releases."[127]

In contrast, in *YS and M and S*, the ECJ addressed whether legal analysis can be considered personal data. This determination is incredibly relevant for the legal status of inferences. A legal analysis is comparable to an analysis of personal data where new data is derived or inferred. Such analysis can consist of multiple inferences connected to an identified or identifiable individual (i.e. assessment of how the law applies to a case), leading to a final opinion, result, or

[119] Case C-101/01, Criminal Proceedings Against Bodil Lindqvist, 2003 E.C.R. I-12992.

[120] Case C-73/07, Tietosuojavaltuutettu v. Satakunnan Markkinapörssi Oy & Satamedia Oy, 2008 E.C.R. I-09831.

[121] Case C-70/10, Scarlet Extended SA v. Société Belge des Auteurs, Compositeurs et Éditeurs SCRL (SABAM), 2011 E.C.R. I-12006; Case C‑582/14, Patrick Breyer v. Bundesrepublik Deutschland, 2016 E.C.R. I-779(stating that "all the information enabling the identification" does not need to be in the "hands of one person").

[122] Case C-291/12, Michael Schwarz v. Stadt Bochum, 2013 E.C.R. I-670.

[123] Case C-342/12, Worten–Equipamentos para o Lar SA v. Autoridade para as Condições de Trabalho (ACT), 2013 E.C.R. I-355.

[124] Case C‑473/12, Institut professionnel des agents immobiliers (IPI) v. Geoffrey Englebert, 2013 E.C.R. I-715.

[125] Case C-212/13, František Ryneš v. Úřad pro Ochranu Osobních údajů, 2014 E.C.R. I-2428, ¶ 22.

[126] Case C-201/14, Smaranda Bara and Others v. Preedintele Casei Naionale de Asigurări de Sănătate, 2015 E.C.R. I-638, ¶ 29.

[127] LARAINE LAUDATI, EUROPEAN ANTI-FRAUD OFFICE, SUMMARIES OF EU COURT DECISIONS RELATING TO DATA PROTECTION 2000–2015, at 32 (2016), https://ec.europa.eu/anti-fraud/sites/antifraud/files/caselaw_2001_2015_en.pdf [https://perma.cc/DLF9-XBMG] (discussing Case T-259/03, Kalliopi Nikolaou v. Comm'n of the European Communities, 2007 E.C.R. I-254).

inference (i.e. the applicant does not meet the required standards of residency), and followed by a decision or action (i.e. denial of legal residency). Three issues arose: (1) is the legal analysis, and the inferences drawn within it, personal data, (2) are the final opinions, results, or inferences about an identifiable individual resulting from the analysis personal data; and (3) is the consequent decision or action personal data? The ECJ's judgement makes clear that the first question must be answered in the negative, meaning the analysis and constituent inferences are not considered personal data.[128] The ECJ, as opposed to the AG, does not distinguish between the legal analysis and the resulting opinions, results, or inferences created in the processing.[129] As a result, no answer is provided to the second question. Finally, the ECJ does not address the third question.

An alternative view, potentially inspired by the AG's distinction between medical analysis and results,[130] could be that the analysis is not equivalent to inferences, but rather the reasoning or logic that leads to the inference. First, it must be noted that this distinction only appears in a footnote in the opinion[131] and was not taken up by the ECJ in this case or in the Nowak case.[132] Second, the reasoning leading to an inference might be better conceived as a cognitive process, while the analysis is regarded as the recorded output of the reasoning. It is difficult to imagine the reasoning or logic in a "legal analysis" not involving the creation of inferences about the applicant's case. Even if one wishes to argue that this is not the case, meaning the legal analysis is merely the reasoning leading to inferences, the outcome of this Article's argument would not change as the problems remain the same. Regardless of how broadly one defines "inference," the rights

[128] Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶ 39, 48.

[129] Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2013 E.C.R. I-838, at ¶ 49 n.40.

[130] *Id.*

[131] *Id.*

[132] *See infra* Section IV.B.1.

granted over inferred or derived personal data are very limited.[133] The main concern addressed by this Article remains the limited rights, control, and recourse given to individuals over inferences, or how they are analyzed and assessed by third parties.

In this regard the judgement followed the opinion of the Advocate General (AG).[134] The AG defines legal analysis as "the legal classification of facts relating to an identified or identifiable person . . . and their assessment against the background of the applicable law,"[135] or "the reasoning underlying the resolution of a question of law."[136] Based on this definition, legal analysis cannot be considered personal data, as she argues that "only information relating to facts about an individual can be personal data,"[137] and thus a "legal analysis is not itself personal data."[138]

To unpack the distinction between facts (as personal data) and analysis, the AG used the example of information describing a person's weight. Allowing that "facts" can be described in "objective" (e.g. kilos) or "subjective" (e.g. "underweight," "obese" terms,[139] she argued that that "the steps of reasoning by which the conclusion is reached that a person is 'underweight' or 'obese' are not facts, any more than legal analysis is."[140] As a result, legal analysis, and more broadly "the steps of reasoning by which [a] conclusion is reached"[141] about an individual, cannot be considered personal data.[142]

---

133 *See infra* Section IV.A.2, Section IV.B.2, and Part V.

134 Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2013 E.C.R. I-838.

135 *Id.* ¶ 54.

136 *Id.* ¶ 59.

137 *Id.* ¶ 56.

138 *Id.* ¶ 61.

139 For a discussion of objective and subjective communication of facts, see *id.* ¶ 57.

140 *Id.* ¶ 58.

141 *Id.* ¶ 58.

142 *Id.* ¶¶ 58–59.

The distinction made here between describing a person as underweight or obese and "the steps of reasoning by which the conclusion is reached"[143] is important for answering the second question. Elsewhere in the opinion, the AG suggests that it is unhelpful "to distinguish between 'objective' facts and 'subjective' analysis,"[144] as "[f]acts can be expressed in different forms, some of which will result from assessing whatever is identifiable."[145] Assessments themselves, insofar as they can be considered a subjective expression of a fact, may therefore be considered personal data. Supporting this, the AG admits that she cannot "exclude the possibility that assessments and opinions may sometimes fall to be classified as [personal] data."[146] In this example, the AG clearly distinguishes between facts or outputs of an assessment process (i.e. an "assessment" or "opinion"), and the process itself (i.e. the "reasoning").[147]

The positions taken by the ECJ and AG in *YS and M and S* appear to be at odds with the view of the Article 29 Working Party.[148] According to their three-step model, personal data is not limited to data *about* an identified or identifiable individual. Rather, data that has the purpose to assess the data subject or results in having an effect on the data subject must also be considered personal data. In her opinion, the AG even refers to the Article 29 Working Party's guidelines on the concept of personal data (which she notes are not legally binding). She explains that the Article 29 Working Party document only attributes personal data status to "*results* of a medical analysis,"[149] but leaves open how the analysis or

143 *Id.* ¶ 58.

144 *Id.* ¶ 57.

145 *Id.*

146 *Id.*

147 *See id* ¶¶ 57–59. ("However, the steps of reasoning by which the conclusion is reached that a person is 'underweight' or 'obese' are not facts, any more than legal analysis is," and "[t]he explanation itself is not information relating to an identified or identifiable person.")

148 *See supra* Part III.

149 *Id.* ¶ 49, n.40 (emphasis in original).

reasoning leading to the assessment should be classified. Interestingly enough, the AG also leaves open how results of the analysis (the second question) should be classified, even though it seems highly unlikely that the outputs of analysis underlying a residency decision (i.e. inferences about the application) and the decision itself are not considered personal data.

The AG's definition of personal data as "facts about an individual,"[150] and the irrelevance of whether such facts are stated in objective or subjective terms, suggests that the she views verifiability as a necessary component of personal data. A troubling sort of test for personal data based upon verifiability can be inferred, wherein assessments and opinions can be classified as personal data only if they meet some unnamed threshold, or are sufficiently based upon verifiable facts to be considered a "subjective statement" of these facts. Where this threshold lies remains unclear.

2. Remit of data protection law

Another troubling aspect of the ruling is the position taken by the ECJ on the remit of data protection law. The ECJ argued that the purpose of data protection law is not to assess the accuracy of decision-making processes involving personal data. On this basis, the applicants' requests for access were denied, as their intention was to assess the accuracy of an assessment of personal data. Rather than being provided by data protection law, the ECJ argued that other laws applicable to the specific case should be consulted to assess whether the decision-making procedure is accurate. Specifically, the ECJ stated that:

> In contrast to the data relating to the applicant for a residence permit which is in the minute and which may constitute the factual basis of the legal analysis contained therein, such an analysis . . . is not in itself liable to be the subject of a check of its accuracy by that applicant and a rectification under Article 12(b) of Directive 95/46 . . . extending the right of access of

[150] *Id.* ¶ 56.

> the applicant for a residence permit to that legal analysis would not in fact serve the directive's purpose of guaranteeing the protection of the applicant's right to privacy with regard to the processing of data relating to him, but would serve the purpose of guaranteeing him a right of access to administrative documents, which is not however covered by Directive 95/46.[151]

YS and M and S is not the first time that the ECJ has claimed that data protection law (when personal data is processed by community institutions and bodies), and the right of access in particular, is not designed to provide access to or facilitate assessments of the accuracy for decision-making processes.[152] In *European Commission v. Bavarian Lager*, the ECJ ruled that:

> . . . when examining the relationship between Regulations Nos 1049/2001 and 45/2001 for the purpose of applying the exception under Article 4(1)(b) of Regulation No 1049/2001 to the case in point, it must be borne in mind that those regulations have different objectives. The first is designed to ensure the greatest possible transparency of the decision-making process of the public authorities and the information on which they base their decisions. It is thus designed to facilitate as far as possible the exercise of the right of access to documents, and to promote good administrative practices. The second is designed to ensure the protection of the freedoms and fundamental rights of individuals, particularly their private life, in the handling of personal data.[153]

In *YS and M and S*, the ECJ referred to Bavarian Lager and explained the overall aim, remit and purpose of data protection law

---

[151] Joined Cases C-141/12 & 372/12, YS v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶¶ 45–46.

[152] *See* Case C-28/08 P, European Comm'n v. Bavarian Lager, 2010 E.C.R. I-6055.

[153] *Id.* ¶ 49.

> Regulation No 45/2001 is not designed to ensure the greatest possible transparency of the decision-making process of the public authorities and to promote good administrative practices by facilitating the exercise of the right of access to documents. That finding applies equally to Directive 95/46, which, in essence, has the same objective as Regulation No 45/2001.[154]

Thus, data protection law in general, and the right of access in particular, are not designed to provide full transparency in decision-making involving personal data, or to guarantee "good administrative practices."[155]

These particular limits on the right of access are not one-off. In *College van burgemeester en wethouders van Rotterdam v. M. E. E. Rijkeboer*, the ECJ ruled that the right of access is limited to providing information regarding the scope of data undergoing processing (which is necessary to rectify or erase this data), to verify the lawfulness of processing, or to object to processing.[156] They covered similar territory in *YS and M and S*, arguing that full access to personal data does not need to be granted under the right of access.[157] Rather, as the ECJ held in *YS and M and S*, "it is sufficient that the applicant be in possession of a full summary of those data in an intelligible form, that is to say a form which allows that applicant to become aware of those data and to check that they are accurate and processed in compliance with that directive[.]"[158]

The AG, like the ECJ, views the remit of data protection law in a very limited way. She views legal analysis as not falling "within the sphere of an individual's right to

---

154 Joined Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en, 2014 E.C.R. I-2081, ¶ 47.

155 *Id.*

156 Case C-553/07, College van burgemeester en wethouders van Rotterdam v. M. E. E. Rijkeboer, 2009 E.C.R. I-3889, ¶¶ 51–52.

157 Joined Cases C-141/12 & 372/12, YS v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶ 44.

158 *Id.* ¶ (70)2.

privacy,"[159] and cannot see a "reason to assume that that individual is himself uniquely qualified to verify and rectify it and ask that it be erased or blocked."[160] She does admit that data subjects have a valid interest in "knowing exactly what circumstances were relevant to the decision taken,"[161] but believes this interest does not fall under the scope of data protection law because it does not "cover opinions and other measures taken during the preparation and investigation" of a case.[162] Instead, review of "the decision for which . . . legal analysis was prepared"[163] should be left to a relevant "independent judicial authority."[164] Data subjects are thus seen to have a valid interest in the accuracy of decisions taken about them, but lack an equivalent right of review.

This is a very troubling view and relates to the discussion above of legal and ethical decision-making standards.[165] First, a legal analysis contains the (interim) inferences, assumptions or opinions underlying final inferences and subsequent decisions. Excluding access and review of such analysis from the scope of data protection law means data subjects are unable to assess how potentially highly impactful inferences and decisions are made about them,[166] unless relevant sectoral laws allow them to do so.

---

159 Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en, 2013 E.C.R. I-838, ¶ 60.

160 *Id.*

161 *Id.* ¶ 36.

162 *Id.* ¶ 32.

163 *Id.* ¶ 60.

164 *Id.*

165 *See supra* Part II.

166 *See* Douwe Korff, *The Proposed General Data Protection Regulation: Suggested Amendments to the Definition of Personal Data*, EU LAW ANALYSIS (Oct. 15, 2014), http://eulawanalysis.blogspot.com/2014/10/the-proposed-general-data-protection.html [https://perma.cc/SRY9-JDW8]; Robert Madge, *Five Loopholes in the GDPR*, MEDIUM (Aug. 27, 2017), https://medium.com/mydata/five-loopholes-in-the-gdpr-367443c4248b [https://perma.cc/L8EM-8YPM]; Steve Peers, *Data Protection Rights and Administrative Proceedings*, EU LAW ANALYSIS (Jul. 17, 2014),

Second, requiring only a summary of personal data undergoing processing to be shared with the data subject via the right of access severely limits the data subject's ability to assess lawfulness of data processing and the accuracy of their personal data used to make the decision.

Third, the limited remit of data protection law is alarming. It might be the case that generally applicable decision-making standards exist in the public sector based on democratic legitimacy,[167] but comparable broadly applicable standards are less likely to govern the private sector. Even though the decision-making autonomy of private entities is bound by certain laws (e.g. anti-discrimination law), companies are less likely than the public sector to have legally binding procedures or rules they need to follow when making decisions. The spread of Big Data analytics and the resulting increase in the capacity of data controllers to infer information about the private lives of individuals, modify and solidify their identity, and affect their reputation, suggest that a higher level of protection is required than has previously been the case for human and bureaucratic decision-making.

Thus, according to the ECJ, when a private company draws inferences from collected data or makes decisions based on them, even if the final inferences or decisions are seen as personal data, data subjects are unable to rectify them under data protection law. Data subjects also lack access to the reasoning underlying the decisions, which is not considered personal data, as well as means to rectify the analysis under data protection law.

B. Case C-434/16: *Nowak*

The ECJ's view in *YS and M and S* seems to be partly at odds with its later ruling in *Peter Nowak v. Data Protection*

http://eulawanalysis.blogspot.com/2014/07/data-protection-rights-and.html [https://perma.cc/69YU-8U9H].

[167] *See generally* De Hert & Gutwirth, *supra* note 73, at 271, 276–77.

Commissioner[168] in December 2017. In the case, an exam candidate (Mr. Nowak) requested to exercise his right of access and "correction" in relation to his marked exam script.[169] As with *YS and M and S*, the case centered on the question of whether opinions and assessments, in this case an exam script and the comments of an assessor, constitute personal data.

1. Inferences as Personal Data

The ECJ determined that both the exam script and comments of the assessor are the candidate's personal data. In making this determination, the ECJ referred to a broad definition of personal data, which includes data "in the form of opinions and assessments, provided that it 'relates' to the data subject."[170] Specifically, the Court determined that an opinion or assessment that is "linked to a particular person" by "reason of its content, purpose or effect" counts as personal data.[171] Both the answers provided by the candidate and the comments made by an assessor on the exam script were deemed personal data on this basis.[172] The ECJ argued that the assessment, comments and evaluation of the candidate can have an "effect" on him and his private life, and are thus his personal data.[173] It is worth noting, however, that exam questions were not considered the candidate's personal data.[174]

The AG held a similar view, arguing that "the personal data incorporated in an examination script is not confined to the examination result, the mark achieved or even points

168 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 60; *see also* Purtova, *supra* note 86, at 66–67.

169 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, Opinion of Advocate General Kokott, ¶ 9–13.

170 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 34.

171 *Id.* ¶ 34–35.

172 *Id.* ¶ 42, 44.

173 *Id.*

174 *Id.* ¶ 58.

scored for certain parts of an examination. That marking merely summarises the examination performance, which is recorded in detail in the examination script itself."[175]

The ECJ also considered whether the interests of other parties can influence the classification of data as personal data. They responded in the negative, arguing that the fact that the assessment of the assessor also constitutes his or her personal data cannot block classification of the assessment as the candidate's personal data.[176] Further, both the ECJ and AG argued that the fact that certain rights like the right of access or rectification might be exercised due to the classification of the exam answers and the comments as personal data is, in fact, irrelevant to making such a classification, even if their exercise would otherwise be thought undesirable.[177] The status of personal data should therefore not be denied based on the data subject potentially exercising the right of rectification in an unintended way (i.e. correcting answers after the fact).

2. Remit of Data Protection Law

While the ECJ acknowledged in *Nowak* that opinions and assessments can be personal data, they did however note that the ability to fully exercise relevant individual data protection rights does not automatically follow from this classification. Rather, the ECJ argued that the scope of the rights attached to personal data have to be interpreted teleologically, with reference to both the aims of data protection law and the purpose for which the data was collected and processed.[178] In other words, the scope of data protection rights must be interpreted contextually, or with reference to the specific purposes for which data was collected, and the broader aims

[175] Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 27.

[176] Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 44.

[177] *Id.* ¶ 46; Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, Opinion of Advocate General Kokott, ¶ 31, 34.

[178] *Id.* ¶ 53.

of data protection law. This means that the reason for which this data is collected defines the data protection rights. In this context, someone was asking to be assessed, and therefore the situation is inherently antagonistic, which means that the data subject cannot rectify how they are being assed, apart from ensuring that their input data was complete.

For an exam script, the rights of access and rectification should not result in the candidate being allowed to correct answers a posteriori.[179] A sensible use of the right of rectification in this context allows the candidate to discover whether

> by mistake, the examination scripts were mixed up in such a way that the answers of another candidate were ascribed to the candidate concerned, or that some of the cover sheets containing the answers of that candidate are lost, so that those answers are incomplete, or that any comments made by an examiner do not accurately record the examiner's evaluation of the answers of the candidate concerned.[180]

Thus, the right of rectification was not taken to cover the content of the assessor's comments, which can be understood as a type of inference about the candidate's performance based on his answers.[181]

---

[179] *Id.* ¶ 51–52.

[180] *Id.* ¶ 54.

[181] *Id.* ¶ 56 ("In so far as written answers submitted by a candidate at a professional examination and any comments made by an examiner with respect to those answers are therefore liable to be checked for, in particular, their accuracy and the need for their retention, within the meaning of Article 6(1)(d) and (e) of Directive 95/46, and may be subject to rectification or erasure, under Article 12(b) of the directive, the Court must hold that to give a candidate a right of access to those answers and to those comments, under Article 12(a) of that directive, serves the purpose of that directive of guaranteeing the protection of that candidate's right to privacy with regard to the processing of data relating to him (see, a contrario, judgment of 17 July 2014, YS and Others, C‑ 141/12 and C‑ 372/12, EU:C:2014:2081, paragraphs 45 and 46)[.]"). This could give the impression that the assessment also falls under the right of rectification. However, considering

The AG's opinion aligned closely with the ECJ on the teleological interpretation of data protection rights. The AG argued that allowing the candidate to rectify answers after completing the exam would be nonsensical, as the purpose for which the data was collected was to evaluate the candidate's performance.[182] Rather, to be sensible, the right to rectification must be limited to assessments of whether the "script inaccurately or incompletely recorded the examination performance of the data subject. For example . . . [if] the script of another examination candidate had been ascribed to the data subject, which could be shown by means of, inter alia, the handwriting, or if parts of the script had been lost."[183]

While the AG acknowledged that assessments (i.e. the assessor's comments) can be personal data,[184] she remained dubious about the applicability of "a right of rectification, erasure or blocking of inaccurate data, under data protection legislation, in relation to corrections made by the examiner."[185] This narrower view is based on the AG's doubt "that comments made on the script could in fact refer to another script or not reflect the examiner's opinion,"[186] as "[i]t is precisely that opinion that the comments are meant to record."[187] Rectification would therefore be inappropriate, as "such comments would not be wrong or in need of correction

---

the examples provided for a sensible use of rectification, see Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 45, and the general goal of data protection—assessing the lawfulness of data processing—it is inconceivable that the right to rectification would also apply to the comments of the assessor.

[182] Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, Opinion of Advocate General Kokott, ¶ 35.

[183] Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 36.

[184] It is interesting to note the AG even points at the similarities between legal analysis and comments, and points towards the tension between interpretations in *YS and M and S* and *Nowak*, but ultimately refuses to address it. *See id.* ¶ 58–59.

[185] *Id.* ¶ 54.

[186] *Id.*

[187] *Id.*

even if the evaluation recorded in them were not objectively justified."[188] Here, the AG again indicates that the remit of data protection law is not to assess the justification behind an assessment or decision, in this case the mark on an exam script.

In contrast to the right to rectification, the ECJ acknowledged that the right of access must be granted "irrespective of whether that candidate does or does not also have such a right of access under the national legislation applicable to the examination procedure."[189] The ECJ did, however, explain that the right of access can be restricted by Member State laws or when the rights of freedoms of others are concerned.[190] This caveat reflects the ECJ's belief that the actual protection afforded by the right of access (and by extension, other data protection rights) must be determined contextually.[191]

These limitations on the rights of rectification and access align with several of the ECJ's prior decisions, which state that the remit of data protection law is not to ensure the accuracy of decision-making processes.[192] Other data protection rights not involved in the case were also addressed in the ECJ's judgement. The right of erasure was determined to be applicable to examination answers and the examiner's

188 *Id*.

189 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 56.

190 *Id.* ¶ 60–61.

191 *Id.* ¶ 60–61. Specifically, the ECJ suggests that "Member States may adopt legislative measures to restrict the scope of the obligations and rights provided for in, inter alia, Article 6(1) and Article 12 of that directive, when such a restriction constitutes a necessary measure to safeguard the rights and freedoms of others." *Id.* ¶ 60. The scope of rights is thus subject to restriction on the basis of purpose- or case-specific risks to the rights and freedoms of others.

192 *See* Joined Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶ 45–47; Case C-28/08 P, European Comm'n v. Bavarian Lager, 2010 E.C.R. I-06055, ¶ 49.

comments after an appropriate period of time.[193] The ECJ also explained that the candidate might have an interest that this data is not "being sent to third parties, or published, without his permission."[194]

In short, in *Nowak* the ECJ and AG seemingly broadened the scope of personal data to include opinions and assessments but followed their previous opinions in that only limited rights are granted over assessments (e.g. opinions, inferences). Further, data protection law was seen to not have the aim to evaluate whether these assumptions are accurate. Data subjects lack a right to rectify the comments (interim inferences) or the results of exams (final inferences) or exam questions.[195] Rather, other applicable laws and remedies need to be consulted, for example through examination procedures.[196] Finally, the remit of data protection law was again limited to discovery of the scope of data being processed, and to assess whether the processing is lawful. Assessment of the accuracy of inferential analytics and decision-making processes remains outside its scope.[197]

Owing to the fact that the rights in the GDPR have to be interpreted teleologically, it is not unthinkable that future jurisprudence will grant the right to rectification in relation to the content of assessments and inferences. However, in many cases people will request an assessment (e.g. to obtain employment, insurance, or a loan). In such cases the aim of processing will be to evaluate the person, which is often an inherently antagonistic situation in which a right to rectify one's assessment would defeat the purpose or telos of the assessment. Paired with the freedom of expression and freedom of contract, and following the ECJ's current

193 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 55.

194 *Id.* ¶ 50.

195 *Id.* ¶ 51–55.

196 *See id.*

197 *See id.* ¶ 57; Case C-553/07, College van burgemeester en wethouders van Rotterdam v. M. E. E. Rijkeboer, 2009 E.C.R. I-03889, ¶ 49.

thinking,[198] right to rectify inferences seems unlikely to be granted in such cases.

C. Lessons from Jurisprudence of the ECJ

The ECJ's rulings in *YS and M and S* and *Nowak* reveal a significant amount about how inferences are treated in data protection law based on the scope of "personal data" and the law's remit. The two judgements differ in their definition of personal data. In *YS, and M and S* the ECJ clearly interprets personal data in a limited way. Name, gender, and similar "facts" about a person are considered personal data, while opinions, reasoning and assessments that underlie decisions are not.[199] The AG even went so far as to argue that only (verifiable) facts constitute personal data.[200] In contrast, the ECJ determined in *Nowak* that opinions and assessments (i.e. comments of the assessor and underlying reasons for the mark) are personal data.[201]

Both court decisions leave open whether the result of an assessment (e.g. the final inference, a mark) and the subsequent decision (e.g. to fail someone at an exam, to refuse legal residency) are personal data. However, in both cases it seems inconceivable that the final assessment and decision, for example the decision to refuse residency or to fail someone at an exam, is not personal data. This seems to be especially true since the ECJ in Nowak[202] used the same terminology as the Article 29 Working Party's three-step definition of personal data, which includes the output data.[203]

---

198 *See supra* Part IV.

199 Joined Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶¶ 38–39.

200 Joined Cases C-141/12 & C-372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2013 E.C.R. I-838, ¶ 56.

201 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994.

202 *Id.*

203 This was even recognized by the Advocate General in *YS and M. and S*. She referred to the view of the Article 29 Working Party that the results of a medical analysis (regardless of verifiability) are personal data.

Despite the seeming widening of the scope of the definition of personal data in Nowak to include inferences, this shift lacks power. Only limited rights over inferences are granted. Further, evaluation of assessments and decisions is said to be outside of the intended purpose of data protection law. In *Nowak*, the ECJ noted that data protection rights do not automatically apply, but must be interpreted according to the purposes for which the data was collected.[204] So, for example, the right of access might conflict with the right to privacy of the assessor,[205] or using the right to rectification to correct answers after the fact would undermine the purpose of the exam to assess the candidate's performance, and thus the answers cannot be corrected.[206] The same holds true for the examiner's comments and assessment.[207]

This view parts with the position adopted by the Article 29 Working Party, according to which inferred and derived data enjoy the full protection of individual rights enshrined in Articles 15–18 and Article 21 of the GDPR.[208] Specifically, the Working Party appears to fully extend certain individual rights of the GDPR to derived and inferred data, including non-verifiable predictions.[209] This much is explicitly stated in relation to the Article 16 right to rectification, which is said to apply to "the 'input personal data' (the personal data used to create the profile) and the 'output data' (the profile itself or 'score' assigned to the person)", which is personal data relating to the person concerned.[210] The rights of access (Article 15), erasure (Article 17), restriction of processing

See Joined Cases C-141/12 & C-372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2013 E.C.R. I-838.

204 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 53.

205 *Id* ¶ 44, 55, 61.

206 *Id* ¶¶ 51–53.

207 *Id* ¶¶ 54–55.

208 Article 29 Data Prot. Working Party, *supra* note 17, at 17–19.

209 *Id*. at 18. For an example of a profiling containing the prediction that a patient will suffer from heart disease, see *id.*

210 *Id*.

(Article 18), and objection to processing (Article 21) are also said to apply.[211] Art. 18 is explicitly said to apply to any stage of the profiling process.[212]

The Article 29 Working Party's position appears to assume that data protection law aims to ensure accurate decision-making, which would allow inferences to be fully accessed, corrected, and erased (for example, if thought to be irrelevant). However, this view runs against the ECJ's decisions in *Bavarian Lager*, *YS and M and S*, and *Nowak*, and the AG in *YS and M and S* and *Nowak*.[213] The rulings and opinions in these cases clarify that the remit of data protection law is not to assess the accuracy of the reasoning behind decisions and assessments, or the accuracy of decisions and assessments themselves. Rather, other laws and governance mechanisms that are applicable to the specific case (e.g. an appeal process for residency or exam decisions) need to be consulted.

Moreover, the ECJ in *Bavarian Lager* and *YS and M and S* and the AG in *YS and M and S* and *Nowak* made it very clear that data protection law does not guarantee lawful decision-making (e.g. a right to good administration or correct marking).[214] The ECJ in Nowak did not disagree, even though reference was made to all these views. The limited way in which the right to rectification applies to the comments of the

211 *Id*. at 17–19.

212 *Id.* at 18.

213 *See supra* Sections IV.A–B.

214 The ECJ and AG's views of the remit of data protection law also contrasts with the Article 29 Working Party's concerns with biased and discriminatory decision-making in automated processing. The ECJ seemingly does not believe that such concerns fall within the scope of the GDPR. Admittedly, the judgments reviewed here were made prior to the GDPR coming into force in May 2018. However, in *Nowak* the GDPR was already acknowledged, so the ECJ's views have arguably already taken it into account. In fact, the ECJ stated that the new framework has even more generous clauses to restrict data access than the old Directive. *See* Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶¶ 44, 59, 61–62.

assessor was even mentioned by the ECJ in its judgment.[215] Based on these considerations and the examples of rectification that the court provided,[216] its agreement is implicit. In general, the ECJ in Nowak even noted how the GDPR allows broader exemptions to the right of access[217] and that Article 16 only aims to verify that the data undergoing processing is complete and accurate.[218]

The scope of Article 16 makes sense in this regard. It would be an odd situation if data protection authorities were competent to rule on the accuracy of immigration cases or examination disputes. In these cases, procedures are in place to deal with complaints (e.g. an examination procedure[219] or a higher court[220]). However, the same cannot always be said for inferences that the private sector draws. It is often left to the private autonomy of industry to assess and evaluate people. Companies are relatively free in how they assess people, except where there are laws (e.g. anti-discrimination law) that limit this freedom.

However, because the data protection rights and the scope of personal data have to interpreted teleologically, it is not impossible that data subjects might have rights to rectify assessments in the future. This will depend on the context. However, often people will ask to be assessed by others (e.g. for credit or loan applications) and therefore this is an inherent antagonistic situation where the self-perception of the data subject (e.g. being a reliable borrower) will not trump the assessment of a bank as making such a determination can fall within their private decision-making autonomy, freedom of contract, or even free speech.

As discussed in Part II, due to companies' widespread implementation of inferential analytics for profiling, nudging, manipulation, or automated decision-making, these "private"

---

215 *Id* ¶ 54.

216 *Id.*

217 *Id* ¶ 61.

218 *Id* ¶ 54.

219 *See supra* Section IV.B.

220 *See* Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶¶ 45–47.

decisions can, to a large extent, impact the privacy of individuals. Thus, a dialogue is necessary to determine the point(s) at which the right to privacy weighs more than the private autonomy of decision-makers, and individuals should have a right to be reasonably assessed.

At first glance, the ECJ's broadening of the scope of personal data in *Nowak* compared to preceding jurisprudence seems to move toward higher protection standards for inferences.[221] However, if the rights in the GDPR (e.g. Articles 15–17) do not apply to inferred and derived data at a level comparable to data provided by the data subject, it cannot be concluded that standards for protecting inferences have actually improved.[222] While it appears that inferences are "economy-class" personal data, this conclusion is not yet fully justified. First, the implementation of individual rights in the GDPR and related European law with regards to inferences must be examined—it must be determined whether data subjects will be able to assess the accuracy or reasonableness of inferential analytics and related decision-making processes. This will be the focus of the next Part.

V. PROTECTION AGAINST INFERENCES UNDER DATA PROTECTION LAW

While the ECJ and the Article 29 Working Party disagree on how many data protection rights enshrined in the GDPR apply to inferences, other European data protection frameworks (i.e. the GDPR,[223] Directive on the supply of

[221] See generally Purtova, *supra* note 86 (arguing that the data protection law becomes the law of everything due to the scope of "personal data").

[222] For a general discussion, see Hildebrandt, *supra* note 20 (expressing concern that data subjects have no control over inferences).

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

digital content[224] and the ePrivacy Regulation[225]) are also relevant to determine the full legal status of inferences in data protection law. This Part reviews the rights available to data subjects to manage how inferences are drawn and used to make decisions. In short, these frameworks offer insufficient protections against inferences.

A. The Right to Know About Inferences

Transparency rights can help individuals to know when and what inferences are drawn. Data subjects possess multiple transparency rights (Articles 13–15) under the GDPR, which aim to provide information about the scope and purposes of personal data collection and processing. In relation to inferences, transparency rights would inform data subjects about the existence and processing of inferred and derived personal data, or data that the data subject has not directly provided.[226] This type of oversight is an essential prerequisite for exercising other rights granted by the GDPR. Unfortunately, the GDPR's notification duties (Articles 13–14) are unlikely to fulfill this aim.

---

[224] Report on the Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content, Nov. 27, 2017, EUR. PARL. DOC. A8-0375/2017, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0375+0+DOC+PDF+V0//EN [https://perma.cc/DL8P-TBEN].

[225] *Proposal for a Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection of Personal Data in Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)*, COM (2017) 10 final (Jan. 10, 2017). For an assessment of the proposal, see Frederik Zuiderveen Borgesius, Joris van Hoboken, Ronan Fahy, Kristina Irion, Max Rozendaal, An Assessment of the Commission's Proposal on Privacy and Electronic Communications, EUR. PARL. DOC. PE 583.152 (2017), http://www.europarl.europa.eu/RegData/etudes/STUD/2017/583152/IPOL_STU(2017)583152_EN.pdf [https://perma.cc/FGZ2-7AP6].

[226] Concerning Article 15, see Article 29 Data Protection Working Party, *Guidelines on the Right to Data Portability*, *supra* note 79, at 10 n.20.

Article 13 describes numerous notification requirements for data controllers when they collect personal data directly from the data subject. At the time data is collected, the controller must provide the data subject with information about the purposes for which the data will be processed, and any potential third-party recipients or category of recipients.[227] Given this timeline, Art. 13 by definition covers only data provided by the data subject, including observed data.[228] Subsequently inferred or derived data thus cannot be included in the disclosure to the data subject as it has not yet been created.

In contrast, Article 14, which addresses notification requirements for personal data obtained from a third party, may be more helpful. Within one month of receiving data from a third party, controllers are required to disclose several pieces of information to the data subject: the categories of personal data collected, intended purposes of processing, recipients or categories of third-party recipients, the data controller's or third party's legitimate interests justifying processing (e.g. direct marketing),[229] and "from which source the personal data originate, and if applicable, whether it came from publicly accessible sources."[230] In practice, a data

[227] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) art. 13(1).

[228] *Id*. at 10.

[229] It is very important to note that "direct marketing" (Recital 47) is considered such a legitimate interest, which means data controllers do not require the data subject's consent. Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 1, 9.

[230] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) art. 14.

controller receiving inferred data (e.g. credit scores) from a third party would need to provide all the above information at the point the data is obtained.

These requirements leave open several gaps in relation to inferences. Even where Article 14 applies, the data controller only needs to inform data subjects about the categories of data involved. "Categories of personal data" is not defined in the GDPR, but it suggests that data controllers do not need to reveal details of the specific data they have received. Rather, providing abstract categories or a list of types of data is sufficient, meaning data subjects will not be aware of the specific data being processed.[231] Additionally, data subjects will not always receive a disclosure from each controller handling their data. If the controller transferring the data included information about (categories of) potential third-party recipients in the original disclosure to the data subject, the recipient controller is not required to make an additional disclosure regarding the transfer.[232] Finally, disclosures are not required if they are "impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes[.]"[233] The notion of "disproportionate effort" is particularly problematic, as the GDPR does not clarify its meaning beyond noting that the quantity of data subjects needing to be informed can be relevant.[234] Each of these gaps indicates that the data subject

[231] Rainer Knyrim, *Informationspflicht bei Erhebung von Daten, in* DATENSCHUTZ-GRUNDVERORDNUNG BUNDESDATENSCHUTZGESETZ 26–27 (Eugen Ehmann & Martin Selmayr eds., 2017).

[232] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) art. 14(5)(a); *see also* Knyrim, *supra* note 231, at 6–7.

[233] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) art. 14(5)(b).

[234] *Id.* at 12.

will not necessarily be informed when and what kind of inferred or derived data has been obtained from a third party.

Articles 13–14 leave open one final, very significant loophole that can result in data subjects being unaware of inferences drawn about them. In cases where inferred or derived data are not obtained via a third party, but rather created by the data controller itself, notification duties will never be triggered because the data is not gathered from the data subject (as necessary under Article 13) or a third party (as necessary under Articles 14). Controllers can thus avoid notification duties by drawing inferences themselves.

Article 15's right of access may provide a solution when the data subject lacks information about inferred and derived data being held for any of the above reasons. According to the Article 29 Working Party guidelines on the right to data portability[235] and on profiling[236], the right of access (Article 15) applies to inferred and derived data, including profiles built from such data by the data controller. Article 15 allows the data subject to request—at any time—information about the purposes of the processing, the categories of personal data held, the recipients or categories of recipients, and the source of the data obtained.[237]

Of course, the data subject must know the identity of the relevant controller to make such a request in the first instance, which poses an additional barrier. And even when such a request is lodged, the data subject may only be informed about the categories of data held, not specific details. However, the data subject may nonetheless be able to gain access to these details by requesting a copy of all data

[235] Article 29 Data Protection Working Party, *supra* note 79, at 10 n.20.

[236] Article 29 Data Protection Working Party, *supra* note 17, at 17.

[237] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 1, 43. It is worth noting that the requirement to provide information about the purposes of processing does not include information about the legal basis for processing.

undergoing processing.[238] This disclosure would include derived and inferred data if the definition of personal data provided by the Article 29 Working Party is followed, and to a lesser extent if the jurisprudence of the European Court of Justice is followed.[239]

While promising, Art. 15(3) is not an absolute right. Obtaining such a copy must "not adversely affect the rights and freedoms of others,"[240] which according to Recital 63 includes "trade secrets or intellectual property and in particular the copyright protecting the software."[241] As a result, inferred and derived data, even if considered personal data, may not need to be disclosed if disclosure could infringe IP law and trade secrets. This view of the limited scope of Article 15 is supported—even if not related to trade secrets—by the ECJ's judgment in *YS and M and S*, which confirms that only a summary of personal data undergoing processing needs to be provided.[242] Further, "rights and freedoms of others" also indicates that Art. 15 should not affect the right to privacy of other data subjects (i.e. third-party privacy). The ECJ confirmed as much in the *Nowak* ruling, stating that the GDPR has more generous clauses to restrict the right of access using Article 15(4) and Article 23[243] to protect the privacy of others (in *Nowak*, the examiner)[244] and other public interests.

The ECJ has thus revealed through these judgments that the right of access, particularly when addressing inferred and derived data, requires a balance of the interests of the data

---

238 *See id.* at art. 15(3).

239 *See supra* Parts III, IV.

240 Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 15(4).

241 *Id.* at 12; *see also infra* Part VII.

242 Joined Cases C-141/12 & C-372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶ 70(2).

243 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶¶ 59, 61.

244 *Id* ¶ 44.

subject making the request, data controllers, and other data subjects they serve, as well as other relevant public interests. As a result, even the right of access may not guarantee oversight of inferences.

One other potential source in European law for a right to know about inferences is worth noting. A new consumer protection package ("new deal") is currently under negotiation, which may require online marketplaces to inform consumers "about the main parameters determining ranking of the offers" presented to them.[245] Such disclosures may need to include information about inferences drawn about the user that underlie the rankings. While promising, it is still very early in the legislative process, so little more can be said about the package's potential at this point.

Data subjects thus face several barriers to oversight over inferences drawn about them. Assuming these barriers are overcome, the GDPR provides several other rights that can be exercised by the data subject: rectification (Article 16), erasure (Article 17), objection to processing (Article 21), and contesting automated decision-making, including profiling (Article 22(3)).[246] Together, these rights can provide data subjects with meaningful control over inferences that may breach their privacy or damage their reputation. However, several further barriers may limit the degree to which these rights can be exercised in relation to inferred and derived data.

---

[245] *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee: A New Deal for Consumers*, at 5, COM (2018) 183 final (Apr. 11, 2018). For a discussion of the relationship between consumer and data protection rights, see Natali Helberger, Frederik Zuiderveen Borgesius & Agustin Reyna, *The Perfect Match? A Closer Look at the Relationship Between EU Consumer Law and Data Protection Law*, 54 COMMON MKT. L. REV. 1427 (2017).

[246] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) arts. 16, 17, 21, 22(3).

B. The Right to Rectify Inferences

Article 16 grants data subjects the right to rectify inaccurate personal data or complete incomplete data "by means of providing a supplementary statement," the scope of which takes into account the purpose of the processing.[247] Rectification implicitly relies upon the notion of verification, meaning that a record can demonstrably be shown to be invalid (i.e. inaccurate or incomplete) and thus "corrected" by the data subject. The right is easy to implement when the data that is used or the inferences that are drawn have a factual basis, or in other words are verifiable (e.g. name, date of birth, marital status, income). For data provided by the data subject, some form of ground truth can be appealed to that demonstrates the flaw in the data held, be it the data subject's account of events, additional observations or records, or some other piece of information.

However, inferences can also be probabilistic assumptions that cannot be verified currently, or perhaps ever.[248] While some inferences can be verified through "ground truth," for example by asking the data subject whether her predicted income range is correct, others are inherently subjective (e.g. the data subject is a "high-risk borrower") or predictive (e.g. the data subject will apply for a mortgage within the next two years) and thus cannot be verified as such.

This distinction between verifiable and non-verifiable inferences has been linked to the applicability of the right to rectification to inferred and derived data, and the definition of personal data more broadly. Some argue that only data that can be verified counts as personal data and thus falls within the scope of the right to rectification, excluding unverifiable inferred data.[249] In contrast, Kamann and Braun suggest that

---

[247] *Id.* at arts. 16.

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

[249] Gianclaudio Malgieri, *Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data*, 4 PING PRIVACY IN GER. 133, 144 (2016) ("This, however, is a restrictive view on such data: they can constitute a complex list of information units (number

the right to rectification should not exclude inferences which cannot be verified, as the verifiability of an inference does not determine its effects on the data subject.[250]

A comparable position is taken by the Article 29 Working Party, which has argued that the definition of personal data does not depend on verifiability.[251] Going further, the Working Party explicitly attributes the right to rectification to opinions and assessments, using the example of a profile that predicts heart disease to which the subject could provide supplementary information.[252] Even though this profile is not verifiable, it is still considered the patient's personal data, at a minimum because it refers to an identifiable individual and can clearly impact his or her life. As a result of the "risk of inaccurate inferences" being drawn by controllers without input from data subjects, "it is also crucial that data subjects/consumers are able to correct or update their profiles if they choose to do so."[253]

The European Court of Justice has similarly (but not consistently) argued that opinions and assessments can constitute personal data.[254] However, as argued above, the ECJ does not see the remit of data protection law as guaranteeing the accuracy of decision-making. This view has major implications for legal protections against inferred data. It means that inferred data (assessments or opinions) and the underlying reasoning behind inferred data—even if considered personal data and objectively wrong—cannot be

and type of potential future illnesses; number and type of future car accidents or future professional misconducts; possible age of death; financial status at the end of one's professional career, etc.) and all these pieces of information could be defined as personal data if and only if they were 'true' or certain.").

250 Hans-Georg Kamann and Martin Braun, *Recht auf Berichtigung*, *in* DATENSCHUTZ-GRUNDVERORDNUNG BUNDESDATENSCHUTZGESETZ 20–21 (Eugen Ehmann & Martin Selmayr eds., 2017).

251 Article 29 Data Protection Working Party, *supra* note 79, at 6; Article 29 Data Protection Working Party, *supra* note 17, at 17–18.

252 Article 29 Data Protection Working Party, *supra* note 17, at 18.

253 Article 29 Data Protection Working Party, *supra* note 57, at 47.

254 *See supra* Sections III.B-C.

rectified under data protection law and can only be contested if there is a procedure in place to contest the evaluation.[255]

As discussed above[256], the ECJ advocates for a teleological approach for the application of the right to rectification. Future jurisprudence could thus hold that the right does apply to inferences in certain cases. However, such cases would likely require a non-adversarial relationship where the interests of the assessor (e.g. freedom of speech, freedom of contract, right to conduct a business) do not outweigh the interests of the assessed person (e.g. right to privacy, identity, reputation).

C. The Rights to Object to and Delete Inferences

The right to erasure may also serve as a remedy against inferences with which the data subject disagrees.[257] According to Article 17(1),[258] the data subject can request deletion of personal data *inter alia* where (1) processing is no longer necessary; or (2) consent is withdrawn and no other legal grounds or legitimate interests exist;[259] or (3) an objection to processing is entered that is not trumped by compelling legitimate grounds of the data controller.[260]

---

[255] *See* Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994 ("Any objections to the comments would consequently have to be dealt with as part of a challenge to the evaluation of the script.").

[256] *See supra* Section IV.B.

[257] For a discussion of why the right to be forgotten is essential in the connected world, see generally MAYER-SCHÖNBERGER, *supra* note 10.

[258] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 17(1).

[259] For an explanation of the concept of "legitimate interests," see Article 29 Data Prot. Working Party, *Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller Under Article 7 of Directive 95/46/EC*, 844/14/EN WP 217 (Apr. 9, 2014), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf [https://perma.cc/J3P5-GUL2].

[260] On the challenges of implementing the right to be forgotten for AI systems, see Eduard Fosch Villaronga, Peter Kieseberg & Tiffany Li,

Concerning point (2), from the controller's perspective, one potential source of legitimate interests is found in Article 16 of the EU Charter of Fundamental Rights: the freedom to conduct a business.[261] The GDPR does not prescribe a specific balance between data subjects' right to erasure and the legitimate interest of controllers. The Article 29 Working Party in relation to the 1995 Data Protection Directive has named, among others, "conventional direct marketing and other forms of marketing or advertisement," "prevention of fraud, misuse of services, or money laundering," "physical security, IT and network security," "processing for historical, scientific or statistical purposes," and "processing for research purposes (including marketing research)"[262] as areas where the data subject's interests may not prevail.

Concerning point (3), Article 21 grants data subjects the right to object to or stop data processing if the processing is based on Article 6(1)(e) (public interest or official authority) or 6(1)(f) (legitimate interests), which includes inferring or deriving new data from existing records.[263] In the case of profiling for direct marketing purposes (Article 21(2)),[264] an

Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten, 34 COMPUTER L. & SECURITY REV. 304 (2018).

[261] Charter of Fundamental Rights of the European Union, 2000 O.J. (C 364), art. 16. On the freedom to conduct a business as a type of legitimate interest, see Norbert Nolte & Christoph Werkmeister, *Recht auf Löschung ("Recht auf Vergessenwerden")*, *in* DATENSCHUTZ-GRUNDVERORDNUNG VO (EU) 2016/679, at 47–48 (Peter Gola ed., 1st ed. 2017).

[262] See Article 29 Data Protection Working Party, *supra* note 259, at 25, where some of the most common legitimate interests are listed.

[263] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 21.

[264] Article 21(2) and Recital 70 of the GDPR mention profiling for marketing purposes. *See id.; see also id.* at 13. In cases where profiling for marketing purposes amounts to online behavior marketing the Article 29 Working Party believes that this cannot be done without consent. *See* Article 29 Data Protection Working Party, *supra* note 259, at 26, 45–47; Article 29 Data Prot. Working Party, *Opinion 2/2010 on Online Behavioural Advertising*, 00909/10/EN WP 171 (June 22, 2010),

objection is guaranteed to be successful, meaning new inferences cannot be generated. It is worth noting that if direct marketing expands to include extensive profiling and tracking, then prior consent under Article 7 of the GDPR must be sought which can subsequently be withdrawn at any point,[265] unless data controllers can claim an alternative legitimate basis.[266] Any other purpose than direct marketing[267] must be weighed against the "compelling

https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp171_en.pdf [https://perma.cc/9ULD-2HQL]; Frederik J Zuiderveen Borgesius, *Personal Data Processing for Behavioural Targeting: Which Legal Basis?* 5 INT'L DATA PRIVACY L. 163 (2015). For a different view in the U.K., see INFORMATION COMMISSIONER'S OFFICE, DIRECT MARKETING: DATA PROTECTION ACT PRIVACY AND ELECTRONIC COMMUNICATIONS REGULATIONS, 20180306 Version: 2.3, at 14, https://ico.org.uk/media/for-organisations/documents/1555/direct-marketing-guidance.pdf [https://perma.cc/9RZV-929D].

265 Article 29 Data Prot. Working Party, *supra* note 259, at 26, 45–47; Article 29 Data Prot. Working Party, *supra* note 264.

266 Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 21(1)(b).

267 *Proposal for a Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection of Personal Data in Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)*, COM (2017) 10 final (Jan. 10, 2017). According to Article 4(1)(f) of the current draft of the e-Privacy Regulation, a *lex specialis* to the GDPR, "direct marketing communications" means any form of advertising, whether written or oral, sent to one or more identified or identifiable end-users of electronic communications services, including the use of automated calling and communication systems with or without human interaction, electronic mail, or SMS. The question remains whether personalized ads are considered "direct marketing," and therefore covered by the latest draft of the e-Privacy Regulation. The Article 29 Working Party has urged to expand the scope to include behavioral advertisements as the current draft seems too narrow. Article 29 Data Prot. Working Party, *Opinion 01/2017 on the Proposed Regulation for the ePrivacy Regulation (2002/58/EC)*, 17/EN WP247 (Apr. 4, 2017), at 20–21.

legitimate grounds" of the data controller.[268] Again, it remains unclear what such "compelling legitimate grounds" would look like.[269] However, if it is determined that the data subject has a stronger interest that allows processing to be stopped or consent is successfully withdrawn, Article 17 can then be effectively invoked to delete the inference.

Unsurprisingly, this has been a point of much discussion historically and now in relation to the GDPR, concerning both the right to erasure and other rights of the data subject. In recent commentary on the GDPR and handling of pseudonymised data, Nolte has argued that data controllers can use their legitimate interest to deny a request for deletion if the data is necessary for the "technical development" of their "app."[270] Requests for deletion may thus only be successful if the data controller no longer requires the data.[271]

Concerning inferences specifically, some commentators have cast doubt on the applicability of the right to erasure to inferences. Some scholars appear to suggest that Article 17 will not apply to inferences altogether,[272] while others argue

---

[268] Very often data controllers use consent for data processing, as lawfulness is easier to prove using Article 7. However, after withdrawing consent, the controller can continue processing if the same purpose is also covered under Article 6(1)(f) GDPR (legitimate interest). *See* Nolte and Werkmeister, *supra* note 263, at 13–15. This view seems to be partly at odds with the Article 29 Working Party's view that "swapping" between legal basis (i.e. from consent to legitimate interest) should not be allowed. They do, however, acknowledge that another lawful basis can justify further processing if this option is determined before data processing starts. *See* Article 29 Data Prot. Working Party, *Guidelines on Consent Under Regulation 2016/679*, 17 EN, WP259 rev.01 (Apr. 10, 2018), https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 (on file with the *Columbia Business Law Review*).

[269] Pointing at this loophole, see Wachter, *supra* note 21; Article 29 Data Protection Working Party, *supra* note 22, at 18, stating that "compelling legitimate grounds" are not defined.

[270] Nolte & Werkmeister, *supra* note 261, at 17–18.

[271] *Id.*

[272] Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18, 68–69 (2017–18). Edwards and Veale believe

that the financial expenditure of a data controller to create inferences will trump the data subject's request for deletion.[273]

These positions stand somewhat in contrast to prior jurisprudence of the European Court of Justice. The ECJ ruled in *Nowak* that the right of erasure applies to examination answers, examiner's comments,[274] and potentially even results[275] (i.e., provided and inferred data as well as the reasons for the inferences), although the right must be counterbalanced against other laws[276] (e.g., longer storage period of exam questions and comments).

Denying the right to erasure based on commercial interests and financial costs alone seems to erode the right to an empty shell, as these constraints will arguably apply to most data processing by commercial entities. Taken together, data subjects would only be allowed to delete personal data that they have provided, and only if this does not conflict with the business interests of the data controller. Further, inferences would face a higher bar than data provided by the data subject

---

that Article 17 is not designed to cover observed data, but they do not offer support. They also assume a complementary relationship between Article 17 and Article 20, which implies that inferred data would not be covered under Article 17 (as is the case with Article 20). The existence of such a relationship is, however, highly doubtful, as Article 20 states: "The exercise of the right referred to in paragraph 1 of this Article shall be without prejudice to Article 17." Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 1, 45.

273 Gianclaudio Malgieri, *Trade Secrets v. Personal Data: A Possible Solution for Balancing Rights*, 6 INT'L DATA PRIVACY L. 102, 115 (2016) ("The denial of access to some data, though creating an information asymmetry between consumers and companies, is necessary to respect economic freedom and freedom of the intellectual property of businesses.").

274 Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 55.

275 The Court did not state this explicitly, but it is reasonable to infer this from their position if the answers and comments can be deleted, provided other laws do not prohibit this. *Id.*

276 *Id.* ¶¶ 55, 60.

due to the additional costs to the data controller to generate the data. This approach seems to miss the balancing act required by the ECJ.[277]

An additional problem remains with the right to erasure. Even if the inferred data is deleted, the data controller might already have shared it with other third parties. Data controllers have limited obligations to inform third parties about deletion. Article 19 requires disclosure "to each recipient to whom the personal data have been disclosed, unless this proves impossible or involves disproportionate effort,"[278] or if the data was made publicly available by the controller, in which case "reasonable steps" have to be taken to inform other controllers who process these data that deletion had been requested.[279] The latter case is unlikely to apply to inferences, as profiles and inferences are not routinely made public. However, in both cases (Article 19 and Article 17(2)), even if inferences are deleted by one data controller, these deleted inferences could still be in use by third parties with whom they were shared. Data subjects bear the burden of identifying and requesting deletion with these third parties. These tasks may not be simple as the GDPR's notification duties (Article 13–14), and right of access (Article 15) give data controllers the option to disclose only categories of recipients with whom personal data has been or will be shared, as opposed to a list of specific recipients.[280]

---

[277] *Id.* ¶ 60. Regarding the view that privacy and business interests must be balanced, see Case T-353/94, Postbank NV v. Comm'n of the European Cmtys., 1996 E.C.R. II-921*; see also* Case T-198/03, Bank Austria Creditanstalt AG v. Comm'n of the European Comtys., 2006 E.R.C. II-1429.

[278] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 19. If the data subject requests it, data controllers must disclose with whom the data was shared, provided that Article 17 was successfully levied. *See id.*

[279] *Id.* at art. 17(2).

[280] *See infra* Section V.A.

Finally, even though "disproportionate effort" cannot be invoked by data controllers to deny a deletion request,[281] Article 11(2) allows exceptions from Articles 15 to 20 in cases where the data controller can prove not to be "in a position to identify the data subject."[282] Therefore, in cases where the data controller has de-identified the personal data (which is often the case in Big Data analytics), the controller does not need to re-identify the data in order to allow the data subject to exercise his or her rights.

It could be argued that other European laws relevant to data processing may provide a right to delete inferences. The current draft of the EU ePrivacy Regulation (EPR), however, does not offer additional support to delete inferences.[283] The framework states that "listening, tapping, storing, monitoring, scanning or other kinds of interception" (including monitoring of browsing behaviors[284]) shall not be allowed unless explicitly permitted under the regulation.[285] Consent to the processing of content data[286] or metadata[287] for one or more specified purposes is valid under this regime

---

[281] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 17; *see also* Nolte & Werkmeister, *supra* note 261, at 31–33.

[282] *Id.* at art. 17(2).

[283] *See Proposal for a Regulation*, *supra* note 225. For commentary on the current draft of the ePrivacy Regulation, see Frederik Zuiderveen Borgesius & Wilfred Steenbruggen, The Right to Communications Confidentiality in Europe: Protecting Trust, Privacy, and Freedom of Expression (Mar. 22, 2019) (unpublished manuscript), https://papers.ssrn.com/abstract=3152014 [https://perma.cc/DV27-8VYQ].

[284] *Proposal for a Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection of Personal Data in Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)*, at 14–15, COM (2017) 10 final (Jan. 10, 2017).

[285] *Id* at art. 5.

[286] *Id. at* art. 6(3)(a)–(b).

[287] *Id.* at 6(2)(c).

following the requirements for consent under Article 7 GDPR.[288]

According to Article 7 of the EPR, metadata and communication data must be erased or anonymized after the "receipt of electronic communication content," or if the metadata "is no longer needed for the purpose of the transmission of a communication." [289] Of course, if the data subject has given consent to further use of this data for other purposes, the data does not need to be deleted. However, even if consent is withdrawn, Article 7 of the EPR only refers to provided data (or content data, e.g. text, voice, videos, images, and sound) and observed data (metadata), but not derived or inferred data.[290] It will thus be unlikely to provide alternative means for data subjects to delete inferences drawn about them.

The framework is still under negotiation and thus it is not clear how the regulation will turn out. The latest Council draft has, however, introduced changes that could weaken data

[288] *Id.* at art. 9(3). The consent requirement covering both content data and metadata is also coupled with the "necessity requirement" as described in Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 7(4). This means that the evaluation of content and metadata is only valid if necessary for the purpose and cannot be fulfilled using anonymous data.

[289] *Id.* at art. 7.

[290] *Id.* at art. 7(1)–(2). Article 7(1) addresses "electronic communications content," defined in Article 4(3)(b) as "content exchanged by means of electronic communications services, such as text, voice, videos, images, and sound." *Id.* at art. 4(3)(b). Article 4(3)(c) defines "electronic communications metadata" as "data processed in an electronic communications network for the purposes of transmitting, distributing or exchanging electronic communications content; including data used to trace and identify the source and destination of a communication, data on the location of the device generated in the context of providing electronic communications services, and the date, time, duration and the type of communication." *Id.* at art. 4(3)(c).

protection, especially in relation to consent to third party tracking and metadata.[291]

The Proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content (Digital Content Directive; DCD)[292] is also unlikely to be helpful in this regard. The proposed framework governs the supply of digital content e.g., "video, audio, applications, digital games and any other software,"[293] excluding healthcare, gambling, and financial services. Article 3 of the DCD regulates the rights and duties of users and suppliers in relation to contracts on the supply of digital content for which "a price is to be paid or the consumer actively provides counter-performance other than money in the form of personal data or any other data."[294] The European Parliament supported the draft but changed the wording from "counter-performance" to "under the condition that personal data is provided by the consumer or collected by the trader or a third party in the interest of the trader" to alleviate concerns

[291] Proposal for a Regulation of the European Parliament and of the Council Concerning the Respect for Private Life and the Protection of Personal Data in Electronic Communications and Repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) – Examination of the Presidency Text, 2017/0003(COD) (Sept. 20, 2018), 20, 26 and 29, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:ST_12336_2018_INIT&from=EN [https://perma.cc/6PYR-64HH].

[292] Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content, COM(2015)634 final, 2015/0287 (COD), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015PC0634&from=EN [https://perma.cc/8EQY-F8L3]. For an overview of the drafting process and the current views of the European Parliament and Council, see generally Contracts for the Supply of Digital Content, EUR. PARL. DOC. PE 608.748 (2017), http://www.europarl.europa.eu/RegData/etudes/BRIE/2017/608748/EPRS_BRI%282017%29608748_EN.pdf (on file with the *Columbia Business Law Review*).

[293] Proposal for a Directive, *supra* note 292, at art. 2.

[294] *Id*. at art. 3.

held by the EDPS that personal data should be used as a currency.[295]

For data to be covered by the DCD, it must be actively provided by the data subject either directly or indirectly (e.g., access to photos or email addresses). Typical examples are "cloud storage services, social media or email accounts."[296]

The interesting segment of the framework concerns actions to be taken after a contract is terminated. Following termination of a long-term contract (under Article 16(4)b of the DCD), or due to a lack of conformity with the contract (under Article 13(2)c of the DCD), the consumer is granted the right "to retrieve all content provided by the consumer and any other data produced or generated through the consumer's use of the digital content," to prevent the supplier from using it,[297] and to render it anonymous.[298] This right covers deletion of user-generated data including "digital images, video and audio files, blogs, discussion forums, text-based collaboration formats, posts, chats, tweets, logos, podcasting, content created on mobile devices, content created in the context of online virtual environments, ratings and collections of links referring to online content."[299]

Concerning the deletion of inferences, the difficulty is that it is unclear whether observed and inferred data are also considered "user-generated data." The DCD explicitly excludes data collected to ensure the digital content conforms

---

[295] Contracts for the Supply of Digital Content, EUR. PARL. DOC. PE 614.707 (2018), at 8 http://www.europarl.europa.eu/RegData/etudes/BRIE/2018/614707/EPRS_BRI%282018%29614707_EN.pdf [https://perma.cc/N6JD-63LT].

[296] Press Release, European Commission, A New Deal for Consumers: Commission Strengthens EU Consumer Rights and Enforcement (Apr. 11, 2018) (on file with the *Columbia Business Law Review*).

[297] Proposal for a Directive, *supra* note 292, at Art 13(2)(b) (" . . . with the exception of the content which has been generated jointly by the consumer and others who continue to make use of the content.").

[298] On the potential impossibility of anonymising data, see Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010).

[299] Proposal for a Directive, *supra* note 292, at 17.

with legal and contractual requirements, including, for example, geolocation data for mobile applications, tracking cookies, and automatically generated data (e.g., IP addresses).[300] Given these constraints, it seems unlikely that inferences will fall in the scope of the law, at least when the user is not actively involved in their generation (and not just providing the underlying data).[301] The DCD's right to delete user-generated data after a contract is terminated thus does not appear to offer a right to delete inferences.[302] As a result, users that have paid for the content or service with their data will not be able to delete data that was derived or inferred based upon it.

D. Protections against Sensitive Inferences

While inferences appear to be "economy class" personal data, the protection of which is contextually bound and typically less than sensitive and non-sensitive data "provided by" the data subject, this trend does not apply to inferences describing special categories of data.[303] Compared to non-sensitive types of personal data, the threshold for collecting and processing sensitive personal data is comparatively high. As described in the preceding sections, requests to know about, transfer, rectify, and delete inferences often require a

[300] Proposal for a Directive, *supra* note 292, at 16–17.

[301] Note that this framework constitutes a "maximum harmonisation" preventing member states to have more consumer-friendly rules, but this minimum standard cannot be circumvented via contracts. European Parliament, *supra* note 242, at 11–12.

[302] For a favorable view, see Inge Graef, Martin Husovec & Nadezhda Purtova, Data Portability and Data Control: Lessons for an Emerging Concept in EU Law (Dec. 6, 2018), https://papers.ssrn.com/abstract=3071875 [https://perma.cc/E7E5-46JV]; *see also* Gianclaudio Malgieri, *'User-Provided Personal Content' in the EU: Digital Currency Between Data Protection and Intellectual Property*, 32 INT'L REV. OF L., COMPUTERS & TECH. 118 (2018).

[303] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 9.

balance to be struck between the interests of data subjects and controllers. However, when sensitive data is being processed, this balance is often not necessary, or at least becomes heavily skewed towards the interests of the data subject.

Take for example the requirements around objecting to processing. Objecting to "legitimate interests" of the data controller could be trumped by the controller's compelling legitimate grounds.[304] However, this is not the case for sensitive data.[305] Unlike non-sensitive data "legitimate interests" of the data controller cannot serve as a lawful basis for data processing. Other potential lawful bases (e.g. explicit consent or the data subject manifestly making their data public) for processing sensitive data or drawing sensitive inferences of course remain, but compared to non-sensitive data, one less route is available to controllers.

1. Can Inferences be Sensitive Personal Data?

While the special protections for sensitive personal data are clear in the GDPR, the extent to which inferences can be classified as such is not. Article 9 of the GDPR defines sensitive data processing as "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation."[306] It is important to note that gender, age, information about a person's financial situation, geolocation and personal profiles are not considered sensitive data under Article 9, despite often serving as grounds for

304 *See generally supra* notes 259, 261.

305 *See id.* (describing further exceptions). Most are coupled with some kind of public interest or require that the data was made public by the data subject. *See id.*

306 *Id.*

discrimination.[307] A general prohibition on sensitive data processing is established with several exceptions, including explicit consent, scientific or statistical purposes, and when "processing relates to personal data which are manifestly made public by the data subject."[308]

Concerns about inferences are implicit in the definition of "special categories of personal data." The phrase "personal data revealing" suggests that the definition is intended to cover data that both directly discloses and indirectly reveals protected attributes.[309] In a 2011 opinion, the Article 29 Working Party supported this position, arguing that the definition of special categories covers "not only data which by its nature contains sensitive information . . . but also data

[307] The sensitive nature of some of these categories and an expansion of "sensitive data" was discussed by the Article 29 Working Party, but did not find its way into the GDPR. *See* Article 29 Data Prot. Working Party, *Advice Paper on Special Categories of Data ("Sensitive Data")*, at 10, Ares(2011)444105-20/04/2011 (2011), https://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf [https://perma.cc/FV7G-VVS4].

[308] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 9.

[309] Sebastian Schulz, *Verarbeitung besonderer Kategorien personenbezogener Daten*, *in* DATENSCHUTZ-GRUNDVERORDNUNG VO (EU) 2016/679 11–12 (Peter Gola ed., 2017) (referring to Article 29 Data Prot. Working Party, *supra* note 307, at 6, where it reads, "The term 'data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership' is to be understood that not only data which by its nature contains sensitive information is covered by this provision, but also data from which sensitive information with regard to an individual can be concluded."; *see also* Edwards & Veale, *supra* note 272, at 37 (claiming it is uncertain whether non-sensitive data is transformed into sensitive personal data if it can be used to infer or reveal sensitive attributes.). However, this position does not account for several opinions and guidelines from the Article 29 Working Party which include such data within the scope of "sensitive data." *See* Article 29 Data Prot. Working Party, *supra* note 307; Article 29 Data Prot. Working Party, *supra* note 81; Article 29 Data Prot. Working Party, *supra* note 17.

from which sensitive information with regard to an individual can be concluded."[310] Similarly, in a later set of guidelines on profiling, the Article 29 Working Party noted that profiling activities can create sensitive data "*by inference* from data which is not special category data in its own right but becomes so when combined with other data."[311] While such proxy data, such as a postcode, is not sensitive by nature, the Article 29 Working Party clearly believes it must be treated as such if it "indirectly reveals" or can be used to infer sensitive attributes.[312]

Higher data protection standards afforded to sensitive data can apply to inferences in two senses. First, when inferred or derived data directly disclose protected attributes—for example when a processor infers a person's ethnicity from their education history—they must be treated as sensitive data. This is a direct form of application in which inferences are treated no differently than sensitive data "provided by" the data subject and is not interesting for purposes of this Article. Second, when personal data can be shown to allow for sensitive attributes to be inferred (i.e., 'indirectly revealed'), the source data from which sensitive inferences can be drawn can also be treated as sensitive data (e.g. last name or location of birth to infer race).

In light of the Cambridge Analytica scandal[313], the European Data Protection Board issued a statement explaining that data that reveals political opinions should be seen as special category data.[314] Further, the fact that the

310 Article 29 Data Protection Working Party, *supra* note 307, at 6.

311 Article 29 Data Prot. Working Party, *supra* note 17, at 15 (emphasis added).

312 Article 29 Data Protection Working Party, *supra* note 307, at 6.

313 Carole Cadwalladr & Emma Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, Guardian (Mar. 17, 2018), https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [https://perma.cc/5SCT-7RKV].

314 European Data Prot. Bd., Statement 2/2019 on the Use of Personal Data in the Course of Political Campaigns (Mar. 13, 2019),

data subject might have made this data publicly available, which is usually an exception from Article 9 of the GDPR, cannot be used as a justification to process data under the "legitimate interest" basis and thus without explicit consent. The statement also stressed that principles of lawfulness, fairness, and transparency also need to be respected.[315] This process can also fall under the restrictions of Article 22, if it is a solely automated process.

This fluidity of the categorization of personal data as sensitive reveals a fundamental problem with the distinction. Non-sensitive data can become sensitive if used to infer sensitive attributes, yet the content of the data remains the same. This suggests that the distinction between sensitive and non-sensitive data is fundamentally flawed, at least when used to govern the collection of personal data.[316] Put simply, the distinction is increasingly strained in the era of Big Data analytics, as seemingly any data can become sensitive personal data if a way can be found to infer information about protected attributes from it.[317]

2. Intentionality and Reliability

Despite the fragility of the distinction between sensitive and non-sensitive data, the higher level of protection afforded to the former in data protection law means it must be taken seriously. The fact that non-sensitive data can reveal

https://edpb.europa.eu/sites/edpb/files/files/file1/edpb-2019-03-13-statement-on-elections_en.pdf [https://perma.cc/5BS2-4VJE]. For a similar case in Austria where the postal office inferred political opinions about their customers without consent, see *Austrian Data Protection Authority Finalises Investigation into Österreichische Post AG* EUROPEAN DATA PROTECTION BOARD (Feb. 19, 2019), https://edpb.europa.eu/news/national-news/2019/austrian-data-protection-authority-finalises-investigation-oesterreichische_en [https://perma.cc/KZK3-AA4H].

315 European Data Prot. Bd., *supra* note 314.

316 *See* Sandra Wachter, *Data Protection in the Age of Big Data* (2019) 2 NATURE ELECTRONICS 6 (2019).

317 *See* Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995 (2017).; Mayer-Schönberger and Cukier, *supra* note 28.

information about sensitive category attributes through linkage and inference begs a question: Under what conditions should non-sensitive personal data be reclassified as sensitive personal data? Much academic discussion has been devoted to this question, according to which the classification of proxy data as "sensitive data" potentially depends on two conditions: (1) the *intention* of inferring sensitive attributes, and (2) the *reliability* of the data in question for inferring sensitive attributes.

Regarding intentionality, several legal commentators have argued that the classification of data as sensitive depends on the stated purpose of processing. Data controllers must have the intention of inferring sensitive information from a selection of data for it to be classified as sensitive.[318] Schulz gives the example of a pizzeria delivering to customers in a drug abuse center. Transaction records would not be considered sensitive data unless the pizzeria intended to infer information about the health status of their customers.[319] Similarly, Schiff argues that last names and location of birth—even though potentially reliable to infer race—are only sensitive if the data controller intends to infer race.[320] Nguyen goes so far as to argue that sensitive attributes coincidentally revealed by non-sensitive data do not require a re-classification of the source data as sensitive, such as a closed-circuit television image depicting a person wearing religious attire, which was not captured to assess the individual's religious beliefs.[321] The same holds true for photos that reveal disabilities or wedding photos at the church from which

318 Alexander Nguyen, *Videoüberwachung Insensitiven Bereichen* 35 DATENSCHUTZ UND DATENSICHERHEIT 715 (2011); Alexander Schiff, *Besonderer Kategorien personenbezogener Daten*, *in* DATENSCHUTZ-GRUNDVERORDNUNG 20–21 (Eugen Ehmann & Martin Selmayr eds. 2017).

319 Schulz, *supra* note 309, at 11–14.

320 Schiff, *supra* note 309, at 14–15.

321 Nguyen, *supra* note 318, at 715. Even though this does not refer to the GDPR, Nguyen's view is relevant as the definition of personal data (which includes inferences) has not changed since the Data Protection Directive; *see also* Schulz, *supra* note 309, at 11–12.

religion or sexual orientation can be inferred, unless the camera was purposefully placed (e.g. at a known meeting point for a particular protected group).

In contrast, although the Article 29 Working Party has not directly addressed intentionality, they have provided some indication that certain types of data can be sensitive without knowing how they will be processed. Photos, images, traffic cameras, and other surveillance devices are seen to raise particular concerns for their capacity to reveal, coincidentally or otherwise, sensitive attributes such as ethnic origin or health status.[322] The classification of these data sources, which are not self-evidently intended to reveal ethnicity or health status, appears to hinge on their content rather than the intention of subsequent processing.[323]

Regarding reliability, several of the same commentators have argued that the non-sensitive data should only be re-classified if it provides a reliable or statistically significant basis to infer sensitive information.[324] Schulz provides two examples: attendance records at union events and online browsing behavior of pornographic content cannot reveal trade union membership or sexual preferences with certainty, and thus do not need to be classified as sensitive data themselves.[325] However, this view is not uncontested. Even though Schiff believes that the intention to infer sensitive attributes is required, he believes that geolocation and browsing history have sufficient disclosive power to reveal political views and sexual orientation. He thinks that reliability does not equal certainty, but rather a good indication (e.g. religious attire).[326]

The General Court has similarly affirmed that data must reliably reveal sensitive information to be considered

---

[322] *See* Article 29 Data Prot. Working Party, *supra* note 307, at 8.

[323] *See* Douwe Korff, Comparative Study on Different Approaches to New Privacy Challenges, in Particular in the Light of Technological Developments 41 (European Comm'n Directorate-General Justice, Freedom and Security, Working Paper No. 2, 2010) .

[324] Nguyen, *supra* note 318; Schulz, *supra* note 309, at 13–14.

[325] Schulz, *supra* note 309, at 12–13.

[326] Schiff, *supra* note 318, at 26–27.

sensitive data, albeit without appealing to the "certainty" threshold advanced by Schulz. Rather, a claim that data reveals sensitive information must be substantiated for it to be treated as sensitive. The General Court used the example of knowing that an individual works as an assistant to a member of the European Parliament; this relationship was not taken as sufficient to infer the individual's political beliefs, suggesting that reliability is an essential attribute of "sensitive data."[327]

Two applications of intentionality and reliability as thresholds for classification of personal data as "sensitive personal data" should be avoided. Some types of data are known to act as a proxy for protected attributes (e.g., postcodes revealing ethnicity). Information about these attributes contained in proxy data can influence inferences or decisions down the line. This influence does not need to be intentional, meaning the proxy data was not intentionally processed as a proxy for the protected attribute, but revealed information about it nonetheless. In the case of proxy data, intentionality is thus unnecessary for sensitive attributes to influence decision-making.[328] Therefore, not even the idea of governing the use of potential sensitive data under the purpose limitation restriction of Article 5 of the GDPR would help.

Similar concerns apply to reliability. As discussed above in relation to the Article 29 Working Party's three-step model, personal data does not need to be verifiable (or accurate) to impact the data subject. Inferences that claim to describe a sensitive attribute, but in fact are drawn from an unreliable source or using unreliable methods would fail to meet the reliability requirement. This situation should not result in the inference or source data from being classified as sensitive personal data, as the accuracy of an inference does not

[327] Case T-190/10, Kathleen Egan & Margaret Hackett v. European Parliament, 2012 E.C.R. I-165.

[328] This is one reason why Germany's data protection law prohibits credit scores solely based on postcodes or addresses. *See* Philipp Richter*, Big Data, Statistik Und Die Datenschutz-Grundverordnung*, 40 DATENSCHUTZ UND DATENSICHERHEIT 581, 583 (2016).

constrain its potential impact on the data subject's life. In effect, if this approach were adopted, the burden of data protection would shift to the data subject to object, rectify, or delete further processing of inaccurate inferences. Successful exercise of these rights cannot be taken for granted, as inaccurate inferences would fail to be considered sensitive personal data due to a lack of reliability, meaning controllers could use legitimate interests as a basis for data processing.

The main issue is not whether the data is a reliable basis to infer sensitive information. Rather, the problem is that data controllers might start treating people differently based on their assumptions about them, even if these sensitive assumptions have not been drawn from a reliable source. In this regard it does not matter whether the inference is accurate or the source data was reliable.

To summarize, the definition of "special categories of personal data" in the GDPR clearly indicates that any personal data that directly discloses or contains information about a special category must be treated as "sensitive data'." In contrast, the classification of data, which indirectly reveals or can be used to infer sensitive information, is not so straightforward. The necessity of intentionality and reliability are a point of disagreement among commentators, the Article 29 Working Party, and the ECJ: one,[329] both,[330] or perhaps neither[331] condition must be met to re-classify non-sensitive source data as sensitive data capable of revealing sensitive information.

E. The Right to Contest Decisions Based on Inferences

Although there is no consensus about the legal rights over inferences, there is an argument to be made that even the

[329] Schiff, *supra* note 318, at 26–27 (arguing that intention is required, but reliability is not); *see also* Case T‑190/10, Kathleen Egan & Margaret Hackett v. European Parliament, ECLI:EU:T:2012:165. Here, the ECJ affirmed the necessity of reliability, but did not address intentionality.

[330] Schulz, *supra* note 309, at 11–14; Nguyen, *supra* note 318.

[331] Article 29 Data Prot. Working Party, *supra* note 307.

GDPR goes beyond procedural data control and management (informational self-determination), and provides safeguards against inferences and decisions based on inferences with the right to contest in Article 22(3).[332] The following Section will call into question that the right to contest can be meaningfully implemented without underlying decision-making standards.

Article 22(3) of the GDPR describes safeguards against decisions based solely on automated processing, including profiling, that produce legal or similarly significant effects for data subjects.[333] Data subjects are granted rights to express their views, contest decisions, and obtain human intervention. These safeguards suggest that the GDPR is moving beyond mere procedural data control and management (or informational self-determination, e.g. Article 15) to allow data subjects to evaluate and challenge automated decisions and profiling that can be based on inferences. Even though the right "to put his point of view" also featured in the 1995 Data Protection Directive (Art 15(2) DPD), the two additional safeguards in Art 22(3) suggest that data subjects' interests in how their data is evaluated are given increasing importance, at least in cases where processing is fully automated. Finally, even though not legally binding,[334] the right to explanation in Recital 71 similarly recognizes data subjects' interests in how they are evaluated.[335] This recognition of valid interests regarding the output of data processing distinguishes Article 22 from the majority of other

[332] Isak Mendoza & Lee A Bygrave, The Right Not to Be Subject to Automated Decisions Based on Profiling (Mar. 9, 2018), https://papers.ssrn.com/abstract=2964855 [https://perma.cc/NPK5-MGE2].

[333] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 22.

[334] *See* Wachter, Mittelstadt & Floridi, *supra* note 11.

[335] *See* Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 14.

mechanisms in the GDPR, which instead focus on management of input data.

The right to contest effectively provides data subjects with the ability to contest automated decisions in sectors where human-based decisions may not be contestable, or where relevant legal or ethical decision-making standards may not exist. But as shown in Part II, the greater protection afforded by the GDPR can be justified by the growing and novel risks introduced by usage of automated decision-making in areas such as "employment opportunities, credit or insurance, or targeting [data subjects] with excessively risky or costly financial products."[336]

At first glance, the right to contest appears to strengthen the protection afforded to data subjects against all types of legally or similarly significant automated decision-making in data protection law, regardless of whether other locally relevant laws apply that would constrain automated decision-making. Data subjects now have a right to contest fully automated decisions regardless of the sector in which the decision was made and without reference to its prevailing regulations and decision-making standards. However, the success of an objection lodged by a data subject turns on its ability to appeal to enforceable legal or ethical decision-making standards which have been violated. The right to contest alone offers little protection against automated decisions and underlying inferences without such complementary standards.

This weakness of the right to contest reflects—as already mentioned—the remit of data protection law, or at least as it has been interpreted by the European Court of Justice in prior jurisprudence and opinions.[337] The ECJ has argued that the remit of data protection law does not include assessment of the accuracy and content of decision-making. In *Nowak*, the ECJ and AG took this position, referring to both the Data Protection Directive and GDPR (which was forthcoming at the

[336] Article 29 Data Prot. Working Party, *supra* note 17, at 10.

[337] *See supra* Sections IV.A–B.

time of the judgment).[338] The ECJ denied data subjects an opportunity to assess the results of decision-making themselves, explaining instead that this evaluation rests with competent sectoral authorities that handle complaints (e.g. an examination procedure[339] or a higher court[340]). The interpretation of the two data protection rights addressed in these cases (i.e. access and rectification) limited them to assessing the accuracy and completeness of input data; for example, whether an exam script was complete, but not the reasoning behind an assessment.

If applied to the GDPR's right to contest (i.e., to nullify or amend an automated decision), this interpretation of the Directive suggests that a challenge will only be successful if the input data was incorrect or incomplete, or other data protection principles were infringed (e.g., the controller fails to demonstrate a lawful basis for processing). The reasoning or parameters behind decisions can only be contested if complimentary decision-making standards (e.g., anti-discrimination law) exist outside of data protection law, which itself does not establish standards concerning the content or outcomes of decision-making processes.

The right to contest thus appears to be a mere procedural right to reverse decisions or impactful profiling made using inaccurate or incomplete input data. It is unlikely to compel data controllers to revise automated decisions based on inferences unless sector-specific decision-making standards or other provisions in data protection law have been infringed. As a result, the private autonomy of the decision-maker will typically be upheld, meaning the choice of parameters used in the decision-making process does not have to be justified to the data subject. If this view is continued in the future, the protection will likely be an empty shell.

338 *See supra* Section IV.B.

339 *See supra* Section IV.B.

340 Joined Cases C–141 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶¶ 45–47

Future jurisprudence could, of course, adopt the advocated teleological approach to extend the rights to contest and to rectification to the content of assessments and inferences.[341] In any case, reasonable assessment standards will need to be established because contesting decisions or inferences will only be successful if a standard or rule is violated.[342] If there are no standards for making decisions, decision-makers will never be in violation of the law.

VI. RE-ALIGNING THE REMIT OF DATA PROTECTION LAW IN THE AGE OF BIG DATA: A RIGHT TO REASONABLE INFERENCES

As should now be clear, inferences receive less protection under data protection law than other types of personal data provided by the data subject. In many ways, the lower status of inferences reflects the limitations placed on the remit of data protection law by the ECJ.[343] Specifically, in standing jurisprudence the ECJ has argued that data protection law is not intended to assess the accuracy of decision-making processes or ensure good administrative practices.[344] Such assessments are instead deferred to sectoral and Member State law, and relevant governance bodies.

While the ECJ plays a key role in defining the remit of data protection law, the novel risks introduced by Big Data analytics and automated decision-making[345] suggest that the prescribed remit of data protection law may be too narrow to realize the law's original aims. In this Part, this Article argues that continuing to rely on sensitivity and identifiability as metrics for the level of protection to grant data is misguided. Rather, greater emphasis must be placed on management of output data, or inferences and decisions, to reconfigure privacy as a holistic concept. A right to reasonable inferences

341 *See supra* Sections III.C, IV.B.

342 *See supra* Part II.

343 *See supra* Part IV.

344 *See supra* Section IV.C.

345 *See supra* Section II.A.

is proposed as an accountability mechanism reflecting this re-configuration of data protection law.

Tensions between profiling, discrimination, privacy, and data protection law have long been acknowledged.[346] In this regard, the term "data protection" is misleading, as it suggests that the laws aim to protect the data, when in fact it is intended to protect people.[347] Data can both directly and indirectly reveal aspects of an individual's private life, which then, among other things, offer grounds for discrimination. The right to privacy offers protection against such disclosures which can lead to discrimination and irreversible harms, "and have long-term consequences for the individual as well as his social environment."[348]

The current limitations placed on the remit of data protection law can be detrimental to its broader aim of protecting privacy against the risks posed by new technologies. As Bygrave explains, privacy is about individuality, autonomy, integrity and dignity.[349] The broader right to privacy addresses personal and family life, economic relations, and more broadly an individual's ability to freely express her personality without fear of ramifications.[350] Protecting this right is a key aim of data protection law.

346 MAYER-SCHÖNBERGER, *supra* note 10; MAYER-SCHÖNBERGER & CUKIER, *supra* note 23; PROFILING THE EUROPEAN CITIZEN, *supra* note 61; Wachter, *Privacy*, *supra* note 54. *See generally* CASES, MATERIALS AND TEXT ON NATIONAL, SUPRANATIONAL AND INTERNATIONAL NON-DISCRIMINATION LAW 674 (Dagmar Schiek, eds., 2007).

347 *See* Mireille Hildebrandt, *Profiling: From Data to Knowledge*, 30 DATENSCHUTZ UND DATENSICHERHEIT 548 (2006); Wachter, *supra* note 54.

348 Article 29 Data Prot. Working Party, *supra* note 307, at 4.

349 LEE A BYGRAVE, DATA PROTECTION LAW: APPROACHING ITS RATIONALE, LOGIC AND LIMITS 128–129 (2002).

350 *See generally* Daniel J. Solove, *"I've Got Nothing to Hide" and Other Misunderstandings of Privacy*, 44 SAN DIEGO L. REV. 745 (2007).

Standing jurisprudence of the ECJ[351] and ECHR[352] has recognized that the aim of data protection law is to protect these broader aspects of privacy, or, in other words, to restrict the processing of personally identifiable data that impacts these areas. Data protection is thus only one segment of privacy.

Reflecting this, privacy and data protection have traditionally been seen as individual rights in the EU.[353] Stemming from the idea that an individual should have the right to be left alone by the state, the right to privacy was originally proposed as a defense mechanism against governmental surveillance.[354] Legal remedies addressing data protection provide tools that prevent individuals from being identified or unduly singled out. On the other hand, legal remedies against discrimination were created based on the experience during the Second World War, seen in Article 14 of the EU Convention of Human Rights.[355] Both aims are reflected in the 1995 Data Protection Directive and now the

---

[351] *See, e.g.*, Case C-101/01, Criminal Proceedings Against Bodil Lindqvist, 2003 E.R.C. I-12971; Case C‑434/16 Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994; Case C‑582/14, Patrick Breyer v. Bundesrepublik Deutschlan, 2016 E.C.R. I-779.

[352] Amann v. Switzerland, 2000-II Eur. Ct. H.R. 1, 20 § 65 (". . . the term 'private life' must not be interpreted restrictively. In particular, respect for private life comprises the right to establish and develop relationships with other human beings; furthermore, there is no reason of principle to justify excluding activities of a professional or business nature from the notion of 'private life' . . . . That broad interpretation corresponds with that of the Council of Europe's Convention of 28 January 1981 [.]"). *See also* COUNCIL OF EUROPE, CASE LAW OF THE EUROPEAN COURT OF HUMAN RIGHTS CONCERNING THE PROTECTION OF PERSONAL DATA (2017), https://rm.coe.int/case-law-on-data-protection/1680766992 [https://perma.cc/MP7S-2DKP].

[353] Alessandro Mantelero, *Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection*, 32 COMPUTER L. & SECURITY REV. 238, 243 (2016); Alessandro Mantelero & Giuseppe Vaciago, *Data Protection in a Big Data Society. Ideas for a Future Regulation*, 15 DIGITAL INVESTIGATION 104, 107 (2015).

[354] Mantelero, *supra* note 353, at 245.

[355] Grabenwarter, *supra* note 354.

GDPR, which restrict processing of personally identifiable information to prevent "singling out," with special provisions for processing of sensitive data due to concerns with discrimination.[356] Sensitive or protected attributes are linked to observable variables that have historically proven discriminatory (e.g. ethnicity, religion).

As the novel risks of automated decision-making and profiling suggest,[357] these systems disrupt traditional concepts of privacy and discrimination by throwing the potential value and sensitivity of data into question. A question thus becomes apparent: Are the fundamental aims of data protection law still being met in the age of Big Data, or is a re-alignment of the remit of data protection required to restore adequate protection of privacy?

To answer this question, it is necessary to evaluate whether individual-level rights can be effectively applied to inferences, and whether the distinction between types of data in data protection law based on identifiability and sensitivity is actually effective when applied to inferences. Concerning the first point, the preceding discussion revealed that data subjects are often unable to access or evaluate inferences drawn about them, as well as the processes that led to these inferences. At a minimum, inferences enjoy less protection under data protection law due to the necessity of balancing requests for access, erasure, or other rights with the interests of data controllers (e.g., trade secrets, intellectual property) and the rights and freedoms of others. Ironically, inferences receive the least protection of all the types of data addressed in data protection law, and yet now pose perhaps the greatest risks in terms of privacy and discrimination.[358]

Concerning the second point, if these distinctions break down when applied to inferences, protections under data protection law are arbitrarily applied, creating greater

[356] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 9.

[357] *See supra* Part II.

[358] *See supra* Section II.A.

opportunities for invasions of privacy and related harms (e.g., discrimination). Many inferences can be drawn from an individual's personal data, but this is not the only possible source. Third party personal data, anonymized data, and other forms of non-personal data can also be used to develop inferences and profiles. This background knowledge, built from anonymized, non-personal, or third-party data, can then be applied to individual data subjects.[359] The process of drawing inferences and constructing profiles can in this way be separated from their eventual application to an identifiable person.

As a result, a gap exists between the capacity of controllers or devices to collect data and draw inferences about people from it, and data protection law's capacity to govern inferential analytics not addressing an identifiable individual.[360] Ultimately, affected individuals are not (fully) able to exercise their data protection rights (e.g. access[361] or erasure[362]) until standalone inferences or profiles based on anonymized, non-personal, or third party data have been applied at an individual level.[363] By using data about people

[359] Wim Schreurs. Mireille Hildebrandt, Els Kindt, Michaël Vanfleteren, *Cogitas, Ergo Sum. The Role of Data Protection Law and Non-Discrimination Law in Group Profiling in the Private Sector*, *in* PROFILING THE EUROPEAN CITIZEN 241, 246 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

[360] *See supra* note 347 (taking the view that data subjects need to consent before data is anonymized.)

[361] Hildebrandt, *supra* note 347, at 550 (explaining "that citizens have no legal right to even access the knowledge that is inferred from these anonymised data and may be used in ways that impact their lives.").

[362] Rubinstein, for example, doubts that the right to be forgotten would apply to profiles built from anonymised or aggregated data. Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 80 (2013) ("[I]t is not even clear whether Article 17 [of the GDPR] would apply to predictive inferences based on personal data that may have been anonymized or generalized as a result of analytic techniques at the heart of Big Data.").

[363] *See also* Hildebrandt, *supra* note 347, at 550. On why exclusion of anonymous data from data protection law is a problem, see Schreurs et al., *supra* note 359, at 241.

not linked to a particular individual, or by purposefully anonymizing data prior to drawing inferences and constructing profiles,[364] companies can thus avoid many of the restrictions of data protection law. This is not to suggest that individuals should have rights over the data of others, or data which has not been applied to them. Rather, the difficulty is that individuals lack redress against the constituent third party or anonymous data and processing that have led to the inferences or profiles applied to them, unless relevant sectoral decision-making standards apply (e.g. anti-discrimination law). Identifiability thus poses a barrier to meaningful accountability for inferential analytics.

As an example, concerns have been raised about the classification of data collected by autonomous cars. Sensors can scan the road ahead, detecting objects to avoid, which may include pedestrians. Such data describing the car's surroundings does not clearly fall within the scope of "personal data" in data protection law.[365] Although undoubtedly data about people, such images do not normally allow for unambiguous identification of recorded individuals.

For data to be "identifiable," it does not need to identify an individual with absolute certainty. Rather, it seems to be enough that the person can be singled out from a group, even if, for example, his or her name is not known, but other characteristics describe the person sufficiently.[366] The

364 *See* Schreurs et al., *supra* note 359, at 248.

365 *See* Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Transparent, Explainable, and Accountable AI for Robotics*, SCI. ROBOTICS, May 31, 2017, at 1, 1. *See generally* Inge Graef, Raphaël Gellert, Nadezhda Purtova & Martin Husovec, Feedback to the Commission's Proposal on a Framework for the Free Flow of Non-Personal Data (Jan. 22, 2018) (unpublished manuscript).

366 *See* Korff, *supra* note 323, at 45. On why the distinction between identifiable and non-identifiable uses is important in the Big Data era, see Colin J. Bennett & Robin M. Bayley, *Privacy Protection in the Era of 'Big Data': Regulatory Challenges and Social Assessments*, *in* EXPLORING THE BOUNDARIES OF BIG DATA 205, 209–10 (Bart van der Sloot, Dennis Broeders & Erik Schrijvers eds., 2016); Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 WASH. L. REV. 703, 704–05 (2016).

possibility of identifying a person must be evaluated reasonably, considering "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."[367]

This can have major implications for assessing problematic behavior of the car, such as a crash, not least because such a definition of "identifiability" is fluid and changes with advances in technology.[368] Scholars have shown that anonymized data can often be linked back to individuals.[369] The driver, pedestrians, insurance companies, regulators, and others could all have an interest in accessing non-personal sensor data, yet the question of access would fall outside of the scope of data protection law.

On a similar note, data does not need to be linked to an identifiable or identified individual to impact his or her life. Schreurs et al. give the example of a shopping cart that can suggest certain products based only on the products that it senses are put in the basket and the speed at which the cart is pushed.[370] In this case, the customer does not need to be identified for choices to be tailored to his or her perceived preferences or needs.

To prevent data harms (e.g., discrimination) and bypass the murky issue of what constitutes personal data, it has been suggested that the "personal data" classification is

---

[367] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 5.

[368] *See* Korff, *supra* note 323, at 46.

[369] *See, e.g.*, Ohm, *supra* note 298, at 1752; Nadezhda Purtova, *Do Property Rights in Personal Data Make Sense after the Big Data Turn?: Individual Control and Transparency*, 10 J.L. & ECON. REG. 64, 74 (2017); Latanya Sweeney, *Only You, Your Doctor, and Many Others May Know*, TECH. SCI. (Sept. 29, 2015), https://techscience.org/a/2015092903 [https://perma.cc/38L5-ATQ8]; Vijay Pandurangan, *On Taxis and Rainbows*, MEDIUM (June 21, 2014), https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1 [https://perma.cc/HW7B-C6UW].

[370] Schreurs et al., *supra* note 359, at 246.

fundamentally broken and should be abandoned.[371] Abandoning this distinction would, of course, leave a gap in data protection law requiring some other classification of data to be introduced to constrain the scope of application of the law. Without a new classification, all data relating to people would effectively become personal data, greatly expanding the scope of coverage of data protection law.[372] While such a move to treat all data as personal data has its merits, such as eliminating overlapping boundaries between personal and non-personal data, such a radical step is not strictly necessary to resolve the specific weaknesses of data protection law concerning inferences. Of course, (sensitive) personal data should never be collected without the explicit consent of the user. But the problem does not lie so much with data collection, but rather with what can be read from the data and the decisions that are based on this knowledge.

Therefore, this Article suggests that continuing to rely on sensitivity and identifiability, or on the blurry distinction among personal data, sensitive data, non-personal, and anonymized data as metrics for the level of protection to grant to data is misguided. This approach fails to protect privacy in the broader sense described above from the novel risks of Big Data analytics and automated decision-making. Rather, greater emphasis should be placed on managing the outputs of data processing, understood here as inferences or decisions, regardless of the type of data informing them. This would reconfigure privacy as a holistic concept, and be more in line with the ECHR,[373] the Council of Europe's "Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data"[374] and their guidelines on AI

371 *See* Purtova, note 86, at 58–59; Wachter, *supra* note 7, at 443.

372 *See* Purtova, *supra* note 86.

373 For an overview of ECHR jurisprudence on privacy to 2017, see Council of Europe, *supra* note 60.

374 Comm. of Ministers, *Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data*, 128th Sess., CM/Inf(2018)15-final (2018), https://search.coe.int/cm/Pages/result_details.

and data protection,[375] and the European Parliament's resolution on a comprehensive European industrial policy on artificial intelligence and robotics.[376] One could also argue for a mediated application of privacy as a human right, and advocate for a "positive obligation" of states to implement laws.

However, the immediate political appeal of such a move is doubtful, given a recent proposal in the EU to facilitate exchange of non-personal data.[377] Unfortunately, the proposal lacks serious consideration of the privacy risks of non-personal data, along the lines outlined above. To make this proposal work, the ECJ would need to redefine the remit of data protection law as a tool to ensure accurate and fair data driven decision-making.

Given these challenges, in order to fully meet the aims of data protection law in the age of Big Data, a "right to reasonable inferences" must be introduced. In response to the novel threats posed by "high-risk inferences," a right to reasonable inferences can be derived from the right to privacy when viewed as a mechanism intended to protect identity, reputation, and capacities for self-presentation. This right

---

aspx?ObjectId=09000016807c65bf [https://perma.cc/RD3F-MVVS?type=image].

375 Directorate Gen. of Human Rights and Rule of Law, *Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108): Guidelines on Artificial Intelligence and Data Protection*, T-PD(2019)01 (Jan. 25, 2019), https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8 [https://perma.cc/H563-X673].

376 A Comprehensive European Industrial Policy on Artificial Intelligence and Robotics, EUR. PARL. DOC. P8_TA-PROV(2019)0081 (2019 [hereinafter Artificial Intelligence and Robotics]), http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2019-0081+0+DOC+PDF+V0//EN [https://perma.cc/UD8Y-ZF6E].

377 *Proposal for a Regulation of the European Parliament and of the Council on a Framework for the Free Flow of Non-Personal Data in the European Union*, at 2, COM (2017) 495 final (Sept. 13, 2017), https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-framework-free-flow-non-personal-data [https://perma.cc/A63H-HHJS].

would offer data subjects additional protections against inferences drawn through Big Data analytics that (1) are predicted or shown to cause reputational damage or invade one's privacy, and (2) have low verifiability in the sense of being predictive or opinion-based while being used to make important decisions.

To make such a right feasible, the ECJ should broaden its interpretation of data protection law regarding an individual's rights over inferred and derived data, profiling, and automated decision-making involving such information. The following Section sketches the scope of this right. To implement a "right to reasonable inferences," new policy mechanisms are needed focusing on ex-ante justification and *ex-post* contestation of unreasonable inferences, which can likewise support challenges to subsequent decisions. Justification would be established by providing evidence of the normative acceptability, relevance and reliability of inferences and the methods used to draw them. If the right were implemented, high-risk inferences would receive comparable levels of protection to automated individual decision-making.[378]

A. Justification to Establish Acceptability, Relevance and Reliability

The ex-ante component of the right to reasonable inferences would thus require data controllers to proactively establish whether an inference is reasonable. Data controllers would need to explain (1) why certain data are a normatively acceptable basis to draw inferences; (2) why these inferences are normatively acceptable and relevant for the chosen processing purpose or type of automated decision; and (3) whether the data and methods used to draw the inferences are accurate and statistically reliable.[379] These requirements

[378] *See generally* Wachter, Mittelstadt & Floridi, *supra* note 11.

[379] On why the immutable attributes rationale for prohibiting discrimination on suspect grounds (e.g., ethnicity) is unhelpful because talent and intelligence cannot be changed either but are treated as a

should be enacted through the introduction of legally binding verification and notification requirements to be met by data controllers prior to deploying high-risk inferential analytics at scale.[380]

The current rules in Article 5 around fairness, purpose limitation, accuracy, and data minimization (including relevance for the pursued purpose) look promising at first glance, but seem to be insufficient. Eskens convincingly argues that "fairness" as used relates to transparency and requires that the user be informed about data processing and their respective rights.[381] The fact that "fairness" is not defined in the GDPR and only appears in relation to lawfulness or transparency makes it questionable that "the fairness principle has any independent meaning at all," and because "fair processing" is never mentioned,[382] it seems unlikely that the GDPR is intended to govern it.

The European Data Protection Board (EDPB) also recently discussed fairness in relation to purpose limitation and legitimate interests of data controllers.[383] The EDPB stated fairness relates to reasonable expectations (e.g. Recital 47 and

legitimate basis for decision-making, see Janneke Gerards, *The Discrimination Grounds of Article 14 of the European Convention on Human Rights*, 13 HUM. RTS. L. REV. 99, 114–115, 115 n.70 (2013).

380 The caveat "at scale" is included to ensure that data controllers can carry out the initial processing necessary to demonstrate normative acceptability, relevance, and reliability. Without this condition, data controllers would be unable to engage in exploratory analysis or develop new methods and types of inferences. The intention is to introduce justificatory requirements to be met prior to widespread deployment, not to prevent development and deployment themselves.

381 Sarah Johanna Eskens, *Profiling the European Consumer in the Internet of Things: How Will the General Data Protection Regulation Apply to This Form of Personal Data Processing, and How Should it?* 27 (Feb. 29, 2016) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm ?abstract_id=2752010 [https://perma.cc/W78T-TQZ2]. For a different view, see Lee A. Bygrave, *Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision Making*, *in* ALGORITHMIC REGULATION (Karen Yeung & Martin Lodge eds.) (forthcoming 2019).

382 Eskens, *supra* note 381, at 27 n.125.

383 European Data Prot. Bd., *supra* note 2, at 5, 9.

50) for data subjects in relation to potential harms and consequences. However, even if this view is followed, "user expectation" is not a democratic or normative justification. The fact that something has become "normal" or commonplace does not necessarily mean it is justifiable or socially desirable.

Similarly, problems arise with purpose and data minimization (including relevance). In the past these provisions have not proven effective owing to the fact that very vague and broad purposes are named in terms and conditions governing data collection and processing. Recent instructive examples are the complaints filed relating to forced consent, as Article7(4) of the GDPR clarifies that consent can only be considered freely given if the data requested is limited to that which is necessary for the provision of a service.[384] If, as a prerequisite of using a service, consent must be given for the collection and processing data beyond that which is strictly necessary for service provision, the consent cannot be considered freely given. Critically, "purpose limitation," "accuracy" and "data minimization" (including relevance for the pursued purpose) seem to only apply to input data.

In general, Article 5 is seen as a transparency tool, not a justification mechanism. One of the problems is that it is the data controllers who define the purpose and relevance of the collected data. A right to reasonable inferences, on the other hand, would open up a dialogue with individual data subjects and society to discuss whether processing practices are normatively acceptable. Finally, a right to reasonable inferences would apply equally to inferences drawn by the data controller and those received from a third party which can subsequently be re-purposed.

In the first instance, the right should apply only to "high-risk inferences" drawn through Big Data analytics which (1) are privacy-invasive or damaging to reputation, or have a high likelihood of being so in the future, or (2) have low verifiability in the sense of being predictive or opinion-based while being used for important decisions. The first condition effectively

[384] *See GDPR: Noyb.eu Filed Four Complaints over "Forced Consent" Against Google, Instagram, WhatsApp and Facebook*, NOYB (May 25, 2018), https://noyb.eu/4complaints/ [https://perma.cc/6KXJ-P6ZZ].

sets a proportionality test for normative acceptability, according to which the damage to privacy or reputation caused by using a particular data source to draw an inference must be proportional to its predicted benefit or utility. Assessments of proportionality and the potential invasiveness of a data source and processing purpose should not be performed by data controllers in isolation.[385] Concerning the second condition, the right in effect applies to both verifiable and non-verifiable inferences in different ways, but is most immediately concerned with mitigating the potential harms of non-verifiable inferences.[386]

These two conditions are proposed as a starting point for application of the right to reasonable inferences. Inferences meeting either condition would meet the threshold for a "right to reasonable inferences" to be exercised. Alternatively, the both conditions could be seen as necessary for the right to be exercised. However, requiring low verifiability in addition to damage to privacy or reputation establishes a threshold that is perhaps too high in practice given the novel risks of inferential analytics.[387] The necessity of each condition for the right to apply should remain open to debate to determine their impact and assess whether general or sector-specific thresholds are preferable.

Alternative grounds for application or additional conditions may also be feasible. For example, the right could alternatively be based solely on the notion of "legal or similarly significant effects" as prescribed in Article 22(1) GDPR.[388] In the conditions proposed here, "important

[385] This type of assessment could conceivably form part of a data protection impact assessment, provided for in Article 35 of the GDPR, if a sufficient level of external review or governance could be guaranteed.

[386] *See infra* Section VI.B.

[387] *See infra* Part II.

[388] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L119) art. 22(1). Specifically, "automated decision-making" is

decisions" are those which have such "legal or similarly significant effects." However, such effects are not limited to "solely automated" decisions as is the case in Article 22(1) because the risks to private life caused by using non-intuitive inferences are not dependent on the extent of automation in the decision-making process.

In any case, basing the right entirely on a threshold of "legal or similarly significant" effects would position it as a complementary protection for the right not to be subject to automated individual decision-making, found in Article 22, which may be desirable. The Article 29 Working Party has provided examples of such effects in relation to Article 22: differential pricing and targeted advertisements that affect vulnerable groups, such as children playing online games being profiled as susceptible to advertisements or adults experiencing financial difficulties.[389] The precise scope of "legal or similarly significant effects" remains unclear in practice, although it will be clarified as the GDPR matures via legal commentary, national implementation, and jurisprudence.

These proposals are not arbitrarily chosen; rather, they reflect current trends in recent EU policy and offer a solution to the worrying weaknesses in data protection law described above. With regards to relevance, the Article 29 Working Party, for example, argues that disclosures providing "meaningful information about the logic involved" in automated decision-making, as required by Articles 13–15, should include "details of the main characteristics considered in reaching the decision, the source of this information and the relevance."[390] The Working Party explicitly warns that data controllers should prevent "any over-reliance on

defined as "a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." *Id.*

389 Article 29 Data Prot. Working Party, *supra* note 17, at 22, 29.

390 *Id*. at 25–26.

correlations,"[391] and explain why a "profile is relevant to the automated decision-making process."[392]

The second component of justification—reliability—requires data controllers to demonstrate that the analytical methods and data used to draw inferences (and potentially make automated decisions) are reliable, for example via statistical verification techniques.[393] The need to demonstrate reliability aligns with the GDPR's Recital 71, which suggests that in order to ensure fair and transparent processing, data controllers are directed to verify the statistical accuracy of their systems, ensure that inaccuracies in personal data can be corrected, and prevent discriminatory effects of automated decision-making.[394] Similarly, the Article 29 Working Party explicitly calls for "algorithmic auditing" to be implemented to assess "the accuracy and relevance of automated decision-

[391] *Id*. at 28.

[392] *Id*. at 31.

[393] Schreurs et al., *supra* note 359, at 253 ("Another matter of concern is the fact that group profiles may incorporate falsified presumptions, such as statistics that wrongly presume that mobile phones will cause cancer or information that people from a certain area have for instance been exposed to radioactive radiation. Knowledge of the logic involved could support an objection to the use of such profiles, even if no personal data of an identifiable person are collected to construct the profile.").

[394] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 14 ("In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.").

making including profiling[.]"[395] Controllers have a similar responsibility for input data, which must be shown to not be "inaccurate or irrelevant, or taken out of context,"[396] and to not violate "the reasonable expectations of the data subjects"[397] in relation to the purpose for which the data was collected.[398] The right to reasonable inferences would apply similar conditions to inferences, understood as a type of output data.

The obligation to demonstrate the reliability of input data and methods aligns with the Council of Europe's views on automated data processing and profiling. The Council has acknowledged a "lack of transparency, or even 'invisibility,' of profiling and the lack of accuracy that may derive from the automatic application of pre-established rules of inference [which] can pose significant risks for the individual's rights and freedoms."[399] It recommends that data controllers "should periodically and within a reasonable time reevaluate the quality of the data and of the statistical inferences used."[400]

Acceptability, relevance and reliability requirements for inferences are not without precedent in European data protection law and policy. Similar requirements for credit scoring have existed since 2010 in Germany's data protection law, although it is worth noting that this law is no longer in force. Specifically, Section 28b required data controllers making predictions or predictive inferences to establish that:

395 Article 29 Data Prot. Working Party, *supra* note 17, at 28, 32.

396 *Id*. at 17.

397 *Id*. at 11.

398 *Id.*; Bart Custers, Simone van der Hof, Bart Schermer, Sandra Appleby-Arnold & Noellie Brockdorff, *Informed Consent in Social Media Use – The Gap Between User Expectations and EU Personal Data Protection Law*, 10 SCRIPTED 435, 445–46 (2013).

399 *The Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling*, at 6, CM/Rec (2010)13 (Nov. 23 2010), https://rm.coe.int/16807096c3 [https://perma.cc/DDX4-HLKQ].

400 *Id.* at 11.

> 1. The methods being used are sound according to the state of the art in science, mathematics, or statistics, and that the data being used is relevant to the type of prediction being made.
> 2. Only legally obtained data is used.
> 3. Predictions regarding the probability of an event happening are not based solely on a data subject's physical address (e.g., post code).
> 4. If physical addresses are used, the data subject is informed of this fact, and it has been documented that the data subject has been so informed.[401]

These requirements closely align with this Article's proposal for data controllers to establish the normative acceptability, relevance and reliability of proposed methods and data sources for drawing inferences. In particular, requiring data subjects to be notified when known proxies for sensitive attributes are used is crucial.

If legally binding requirements are created along these lines, a balance must be struck between data subject and controller interests. At a minimum, data controllers should be obligated to provide information regarding the intended content or purpose of the inferences being drawn, the extent to which these inferences rely on proxies for sensitive attributes, and counterintuitive relationships between input data and the target inference (e.g., basing creditworthiness on clicking behavior). This type of information is intended to be the starting point of a dialogue between data subjects and data controllers regarding the justifiability of particular inferences. One of the greatest risks of inferential Big Data analytics and automated decision-making is the loss of control over how individuals are perceived, and the predictability or intuitive link between actions and the perceptions of others. The proposed notification requirements are intended to make the process of evaluating the data subject more open,

[401] The authors translated this from German. *See* Gesetz zur Änderung des Bundesdatenschutzgesetzes [Law Amending the Federal Data Protection Act], Jul. 29, 2009, BGBL I at 2254, § 28b (Ger.).

inclusive, and discursive, and to provide a new channel of remedies for data subjects who believe that unreasonable inferences have been drawn.

B. Contestation of Unreasonable Inferences

To complement ex-ante notification requirements, the second half of a "right to reasonable inferences" should provide an effective ex-post accountability mechanism for the data subject. The ex-ante justification is bolstered by an additional ex-post mechanism enabling unreasonable inferences to be challenged.[402] This right would allow data subjects to contest inferences themselves (e.g., credit score), which complements the existing right to contest automated decisions found in Article 22(3).[403] With the considerations of justification in Section VI.A in mind, the right to contest would be transformed from a mere procedural tool[404] to a remedy that allows assessment of the content behind a decision.

In practice, contesting would amount to raising an objection with the data controller if an inference drawn is found by the data subject to be inaccurate or unreasonable (e.g., if based on non-intuitive, unreliable, or invasive features or source data), and to offering supplementary information that could lead to an alternative preferred outcome. Contesting as imagined here encourages dialogue between the data subject and the controller if the accuracy or reasonableness of an inference is questioned.

The ex-post component of the right to reasonable inferences is not, however, intended to shift decision-making autonomy from private actors to data subjects. Contesting an inference and offering supplementary information does not guarantee that the inference in question (or subsequent

---

[402] In favor of such a solution, see Mireille Hildebrandt & Bert-Jaap Koops, *The Challenges of Ambient Law and Legal Protection in the Profiling Era*, 73 MOD. L. REV. 428, 448–49 (2010). On the need to remedy unjust judgments based on inferences, see Leenes, *supra* note 73, at 298.

[403] Mendoza & Bygrave, *supra* note 332, at 6, 14.

[404] *See supra* Section V.E.

decisions challenged under Article 22(3) of the GDPR) will also be modified. Data controllers have private autonomy in the ways they evaluate data subjects and make decisions about them. The right to reasonable inferences is not intended to violate this autonomy, but rather to provide the data subject with a way to learn more about the data controller's perceptions and decision-making processes, and to potentially convince the controller that one or both is wrong.

For verifiable inferences (e.g., Jessie is a homeowner), it is reasonable to assume that offering supplementary information demonstrating the original inference is inaccurate would lead to rectification of the inference, as accurate data is in the interests of both parties. This type of right is nothing new, as data subjects can already rectify data in this way under Article 16 of the GDPR.[405] This proposal only suggests broadening the scope of Article 16 from merely input data to also output data, which is in line with the Article 29 Working Party's view.[406]

For non-verifiable or predictive inferences (e.g., Jade will default on a loan in the next five years), data subjects arguably do not have an equivalent form of rectification. Non-verifiable inferences cannot be rectified as such due to their inherent uncertainty or subjectivity.[407] The data subject may nonetheless disagree with the controller's views or assessment if, for example, it does not align with their self-perception, the source data is perceived as irrelevant, or the scope of data considered was incomplete or insufficient. Contesting the normative acceptability, relevance or reliability of an inference on any of these grounds is distinct from rectifying a provably inaccurate inference.

The right to rectification in Article 16 may arguably already offer a remedy for non-verifiable inferences. Whether

---

[405] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 16.

[406] *See supra* Part III.

[407] *See supra* Section III.B, Section V.B.

this is the case depends upon one's view of the necessity of verifiability in classifying inferences as personal data[408] and its impact on subsequent application of data protection rights. The ECJ, for example, argues that the right to rectification is not intended to apply to the content of subjective (and thus non-verifiable) opinions and assessments.[409] In contrast, the Article 29 Working Party believes predictive inferences can also be "rectified" by providing supplementary information that would alter the assessment, meaning that verifiability is not necessary to exercise the right of rectification.[410]

The proposal for an ex-ante right to contest inferences made here may thus not represent a radical departure from existing law. Rather, if adopted, the right to reasonable inferences would effectively enshrine an answer to the verifiability question in law, and thus strengthen data protection rights over inferences regardless of their verifiability. This sort of strengthening is essential if the interests of data controllers are to form less of a barrier to exercising individual data protection rights against inferences than is currently the case.[411] In conjunction with the ex-ante notification requirements, the data subject's chances of successfully contesting inferences (and automated decision-making based upon them) would likewise improve, as the subject could draw on the justification disclosure made by the controller prior to an inference being drawn.

VII. BARRIERS TO A RIGHT TO REASONABLE INFERENCES: IP LAW AND TRADE SECRETS

As shown in Parts III and IV, the first hurdles to the implementation of a right to reasonable inferences lies with determining the legal status of inferences. Once consensus has been reached on whether inferences are personal data, the rights granted in the GDPR very often need to be

408 *See supra* Section III.B, Section V.B.

409 *See supra* Part IV, Section V.B.

410 *See supra* Part III.

411 *See supra* Part V.

counterbalanced with the legitimate interests of data controllers concerning, for example, trade secrets, intellectual property, or third-party privacy.[412]

The easiest legal solution to prevent unreasonable inferences from being drawn would be to allow data subjects to prevent models from being built in the first place, or to grant them control over the models used in inferential analytics, and how they are applied. Such a solution is of course not to be recommended, as it fails to respect the substantial public and commercial interests advanced by analytics and technological development more broadly. With regard to the mechanisms recommended in the preceding Section, a more reasonable approach would be to require controllers to justify to regulators or data subjects their design, choice, and usage of models and particular data types to draw inferences about individuals. However, there are an alarming number of provisions in the GDPR and other (proposed) regulations that could seriously hinder the protection afforded to data subjects against inferences.

In short, the GDPR, new and old IP laws, and the new European directive on trade secrets do much to facilitate Big Data analytics and the construction of machine learning models. This Part considers models to be the outputs of data processing involving inferential analytics that uses an individual's personal data. In other words, personal data is used to draw inferences which lead to a model, which can then be applied to other people, cases, or data to make decisions. Under the GDPR and the new Copyright Directive,[413] data subjects' rights are restricted for the purpose of constructing models. For construction that does not meet the requirements of the statistical purpose exemptions, data subjects would retain these rights. However, once an output (the model) has been produced, new regulations dealing with copyright and trade secrets would give the individual little say in how the model is used, and little to no share in the benefits it produces.

---

412 *See supra* Part V.

413 *Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, COM(2016) 593 final (Sept. 14, 2016).

A. Algorithmic Models and Statistical Purposes in the GDPR

The GDPR may facilitate inferential analytics by granting a number of privileges to processing for statistical purposes.[414] After data is collected based on one of the legal bases in Article 6, the strict "purpose limitation" in Article 5 no longer applies.[415] Article 5(1)(b) states that "further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes[.]"[416] The same privilege applies to the strict principle of storage limitation in Article 5(1)(e)[417], and thus the data does not need to be deleted after it is no longer necessary for the original processing purpose. This means as long as data is collected in a lawful manner following Article 6, and in accordance with "appropriate safeguards" pursuant to Article 83 (e.g., pseudonymization) are in place, the subsequent use for statistical purposes is lawful and does not require any additional legal basis for processing (e.g., consent) to be established.

Mayer-Schönberger and Padova believe that Big Data analytics can be considered "processing for statistical purposes," as they are strongly based on statistical

414 Mayer-Schönberger & Padova, *supra* note 50, at 326–27. *But see* Bertram Raum, *Verarbeitung zu Archivzwecken, Forschungszwecken*, *in* DATENSCHUTZ-GRUNDVERORDNUNG 31–32 (Eugen Ehmann & Martin Selmayr eds., 2017) (expressing uncertainty over whether the exemptions apply to Big Data).

415 Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 5(1)(b).

416 *Id.*

417 Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 5(1)(e).

methods.[418] Relatedly, Zarsky argues that Big Data would face significant difficulty to fall within this exemption.[419] If the exemption is applied, Member State law can grant controllers numerous privileges and exemptions from other rights and duties in the GDPR, as described in Article 89(2). These include exemptions from Articles 14(5)(b), 15, 16, 17(3)(d), 18 and 21, as well as the strict limitations on the use of sensitive data in Article 9(2)(j) and Recital 52.[420]

These exemptions have two implications for the diffusion of inferential analytics. First, they encourage the creation of new statistical models and profiles by lowering data protection requirements for such processing. Second, following from this relaxation of the law, when personal data is used for statistical purposes data subjects are unable to exercise the majority of their rights, and thus cannot prevent statistical uses. Similarly, data subjects lack any claim or rights over the resulting models or profiles (i.e., "statistical results" in Article 89(1)), despite having been built with their personal data.

It is important to note a further restriction on the Article 89 privileges. Recital 162 clarifies that statistical results generated under the statistical purposes exemption (which are aggregate data, not personal data), as well as the input personal data, cannot be used "in support of measures or decisions regarding any particular natural person."[421] It is

---

[418] Mayer-Schönberger & Padova, *Regime Change*, *supra* note 50, at 330.

[419] *See* Zarsky, *supra* note 317, at 1007–08.

[420] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 1, 10; *see also id.* at arts. 9(2)(j), 14(5)(b), 17(3)(d), 89(2).

[421] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) 30; *see also* Zarsky, *supra* note 317, at 1008; Schreurs et al., *supra* 359, at 248. Both Zarsky and Schruers et al. are silent on the view of applying profiles after creation but hint that the law

difficult to imagine how compliance and enforcement of this restriction will be handled (i.e., how to ensure that the model is not applied or intended to be applied to a natural person), or how to manage the sale of models generated under Article 89 exemptions to third parties. Presumably, if the results (which must not be personal data, per Recital 162) are then used to make decisions about individuals, the privileges granted by the statistical purposes exemption are no longer applicable, meaning normal data processing rules, such as Articles 6 and 22, will apply.[422]

An important point of contention regarding these exemptions is whether they apply to commercial data controllers, or only to public and research entities, such as government bodies and universities. Mayer-Schönberger and Padova argue that these privileges apply to "private companies for commercial gain" as well.[423] A similar view comes from Richter, who argues that the statistical purposes exemption can be used to pursue commercial interests as long as the results are not applied to individuals.[424] In contrast,

might prohibit this, albeit without any clear supporting evidence. This view translates to the GDPR because the DPD had a similar provision in Recital 29. For a view that the later application should be covered by Article 6, see Richter, *supra* note 328, at 585 who also warns that this can never be sufficiently regulated as there is no way of assessing how the models are subsequently used for other processing or by other data controllers.

[422] Article 29 Data Prot. Working Party, *supra* note 17, at 7 ("For instance, a business may wish to classify its customers according to their age or gender for statistical purposes and to acquire an aggregated overview of its clients without making any predictions or drawing any conclusion about an individual. In this case, the purpose is not assessing individual characteristics and is therefore not profiling."; Raum, *supra* note 414, at 41(explaining how further usage of statistical results is no longer covered by the privileges, but can be used if the normal requirements for data processing in the GDPR are met). For example, to assess individuals with a model built under the statistical purposes exemption, a further legitimate basis for processing would need to be established, such as consent.

[423] Mayer-Schönberger & Padova, *supra* note 50, at 326.

[424] *See* Richter, *supra* note 328, at 585 (arguing later application should not be lawful even if it fulfills Article 6 requirements due to the possible risks). Richter does not, however, offer a legal argument to justify this claim.

Raum suggests that the exemptions cannot be used for commercial interests, and that any subsequent usage of statistical results generated under these exemptions for commercial interests would require justification according to the GDPR's standard data processing requirements.[425] This suggestion is, however, not supported with any further legal argumentation.

Once the model is applied to a person, regardless of whether it was built under the statistical purposes exemption, the outcome of this application (i.e., an inference or decision) becomes the personal data of the person being assessed and the restrictions detailed in Part V apply. Members of the training set also retain rights over any of their personally identifiable data contained within the model, unless statistical purposes exemptions apply. However, while the model is admittedly applied to a data subject for the purpose of assessment, this does not mean the model will be considered the personal data of the person being assessed or the data subjects represented in the training data. Further, neither party will have rights over the model. To understand why this is the case, it is necessary to return to the judgements discussed in Part IV.

In *Nowak*, the ECJ made clear that the exam questions are not the candidate's personal data,[426] even if used to assess him. The exam questions are comparable with the model that is used to assess an individual. The same holds in the case of *YS and M and S*, where immigration law is comparable to a statistical model. The fact that immigration law was applied to the applicant to make a decision on residency does not mean the law itself became the applicant's personal data.[427] The

He further warns that the GDPR legalizes many applications that would have been illegal in Germany (e.g., private sector uses). *See id.*

[425] *See* Raum, *supra* note 414, at 41–42.

[426] Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶ 58.

[427] Not even the legal analysis (as an abstract application of the law) is personal data, but rather only the personal data undergoing processing is. Joined Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie,

data subject thus cannot rectify or delete the law. Therefore, neither the exam questions nor the applicable law are subject to the rights granted in the GDPR.

The request to access the "legal analysis" in *YS and M and S* further clarifies the distinction between a model and application of the model. As already discussed, immigration law provides the background framework, or model, in which residency applications are assessed. The application of this law to the particulars of an applicant's case, or the "legal analysis," can be considered equivalent to the application of a statistical model (i.e., the analysis or reasoning) to a data subject to make a decision. This relationship between a model and analysis can be equally applied to algorithmic decision-making models. For example, a decision tree used to make a decision on the basis of personal data can be considered a model. The analysis in this context would constitute the specific path, or branch, followed in the decision tree to reach an output or decision. So, in other words, a specific path in the decision tree relevant to deciding a specific case constitutes "analysis," whereas the entire tree constitutes a "model."

Even if models (e.g., immigration law or exam questions) were treated as personal data, the rights in the GDPR must be interpreted teleologically to avoid nonsensical results.[428] In *Nowak*, this was clearly seen in the determination that allowing the candidate to rectify answers on an exam would be nonsensical as it would undermine the original processing purpose of evaluating the candidate's performance, despite being the candidate's personal data. The same applies to rectification of the exam questions, which are not considered personal data. In the case of statistical or algorithmic decision-making models, rectification of the model itself would often be equally nonsensical, or at least not constitute a fair balance of subject and controller interests, due to its potential

Integratie en Asiel, 2014 E.C.R. I-2081, ¶¶ 48–49, 59. Therefore, the law will also not be seen as personal data.

[428] *See supra* Part IV.

impact on application of the model to other cases, or research and business interests more broadly.

Finally, the remit of data protection law does not include assessment of the accuracy or justifiability of decisions (and underlying opinions or evaluations),[429] and does not allow individuals to decide which models (e.g., exam questions, laws) are used to assess them.[430] Rather, these choices fall within the data controller's private decision-making autonomy.

An example may help to illustrate why models cannot be considered personal data. If a doctor asks about a patients height, and she replies 166 centimeters, such an utterance is her personal data. This data falls under the GDPR and can be rectified, deleted, etc. However, the fact that her height is expressed in centimeters does not mean that the metric system (i.e., the model used to assess her height) becomes her personal data, meaning that she would have rights over it. By having her height measured, she will not gain the right to rectify or delete the metric system. Similarly, she would not have a right to require that a different measuring system or

[429] Joined Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2014 E.C.R. I-2081, ¶¶ 32, 46–48; Case C–434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, ¶¶ 52–54. With regards to the ECJ's judgment in *Nowak*, the examples of cases in which exam answers or the examiner's comments could be considered "inaccurate" deal with cases in which the input data for a decision is somehow incomplete or corrupted (e.g., pages of answers were missing from the script assessed by the examiner). A clear distinction is drawn in paragraph 54 between the examiner's comments, and the examiner's evaluation of the candidate's performance, with the former being treated as "recording" the examiner's evaluation. *See id.* ¶ 54. The ECJ is thus indicating that the examiner's comments, which can themselves be considered inferences or subjective statements of opinion, can be rectified if they have been recorded on the basis of incomplete or corrupted input data. The candidate is not granted the right to rectify the opinion, analysis, or evaluation criteria of the examiner.

[430] *See supra* Part IV.

model be used, for example the imperial system, because she prefers the imperial system or finds it more accurate.[431]

One could argue that this example is not equivalent to trained algorithmic models, as personal data was not used to construct the metric system. So, while the model would not constitute personal data of the individual being assessed, it may still conceivably be the personal data of the individuals whose data was used to construct it. To address this alternative, consider instead a marking rubric as was presumably used in *Nowak*. In addition to the exam questions, such a rubric would constitute a model used to make a decision about the performance of the candidate. The rubric is arguably constructed from personal data, insofar as it is derived from the past experiences and opinions of the assessor or course leader with other exams, and perhaps specific answers provided by candidates in prior years. The rubric could even go so far as to include personal data, if for example a prior candidate's answer was copied into the rubric as an example response to a question.

In this case, it would be equally nonsensical to assume that the prior candidate whose personal data is contained in the rubric would have data protection rights over the rubric as a whole. Rather, in line with the ECJ's stance on the right to erasure in relation to exam answers, the prior candidate would retain rights over the extract of her responses contained in the rubric (assuming she was still identifiable, for example if the author of the rubric recalled who provided the example in question). In line with her data protection rights over personally identifiable data, the candidate could justifiably request access or deletion of the extracted

[431] For a view that trained models might be personal data, meaning the data subject would have rights over the model in its entirety, see Michael Veale, Reuben Binns & Lilian Edwards, *Algorithms That Remember: Model Inversion Attacks and Data Protection Law*, 376 PHIL. TRANSACTIONS ROYAL SOC'Y A 1, 1 (2018). This view, however, misinterprets the standing jurisprudence of the ECJ addressed here and does not take the remit of data protection law and the need to balance individual rights with trade secrets and IP law into account.

response.[432] In the context of a trained algorithmic model, the right to erasure could be interpreted as requiring the data to be removed from the training set, thus requiring the model to be re-trained.[433]

Regardless of whether the prior candidate's requests would be successful in the real world, they demonstrate why personal data being contained in a model should not be thought to automatically grant individual rights over the model itself. Rather, the data subject's rights apply only to the specific personally identifiable data contained within the model. This approach aligns with the teleological interpretation of individual rights described by the ECJ and AG in *Nowak*.[434] The purpose of a model is to assess individuals; it would be nonsensical to assume that individuals whose data was used to train the model would be able to modify or delete the model entirely, and thus have an unjustifiably significant impact on the individuals being assessed by it. The scope of data protection rights must be appropriately applied and constrained to reflect the relationship between the data subject and the model, and the relevant processing purposes. In other words, the mere presence of personal data in a model in no way equates to the full, unbounded exercise of rights over it.

Finally, law and policy on IP, copyright, and trade secrets also apply to the model which may prevent the exercise of individual data protection rights. In particular, these are likely to prevent requests to "delete" personal data from a model by re-training it from being successful, if doing so requires significant effort or is disruptive to business practice. The impact of these conflicts between frameworks are explored in the next three Sections.

B. Algorithmic Models and the EU's Copyright

[432] *See supra* Section IV.B.1 (discussing erasure of examination answers), Part V (outlining necessary conditions).

[433] On the challenges of implementing the right to be forgotten for AI systems, see Villaronga, Kieseberg & Li, *supra* note 260.

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

Directive

The previous Section shows that the GDPR facilitates the creation of profiles and models, either built from inferences (among other data) or is capable of producing them when applied to individuals. When the statistical purpose exemption applies, the individual cannot object to its construction and has no rights over it, even if the model is built using personal data. Further, even if the model is applied to a natural person (meaning the statistical purposes exemptions no longer apply), no control or rights over the model are likely to be granted if the jurisprudence of the ECJ is maintained. Similarly, members of the training data set will retain data protection rights over any personal data contained in the model and may be able to exercise rights in relation to it (unless statistical purposes exemptions apply), but this will not equate to any control or rights over the model as a whole.

The facilitation of model constructions and lack of individual rights seen in the GDPR can also be seen in IP and copyright law. Current discussion of machine learning and inferential analytics in the context of IP law focuses broadly on two issues: (1) whether the training data used to construct a model (e.g., content uploaded or created by their users) is protected by IP laws; and (2) whether the outcome of the algorithmic process can be protected under IP law.[435]

A new EU Copyright Directive[436] is currently under debate, which will complement the existing legal framework

[435] *See generally* Daniel Schönberger, Deep Copyright: Up- and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML), 10 INTELL. PROP. J. 35 (2018); Annemarie Bridy, *The Evolution of Authorship: Work Made by Code*, 39 COLUM. J.L. & ARTS 395 (2016).

[436] *Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, COM (2016) 593 final (Sept. 14, 2016).

on copyright[437] and will, among other things, govern the legal status of data mining.[438]

The Directive is among other things concerned with research organizations such as universities and research institutes (including public-private partnerships[439]) that use new technologies that "enable the automated computational analysis of information in digital form, such as text, sounds, images or data, generally known as text and data mining. Those technologies allow researchers to process large amounts of information to gain new knowledge and discover

[437] In order of enactment, see generally Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases, 1996 O.J. (L 77) 20–28; Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society, 2001 O.J. (L 167) 10–19) (implementing the "WIPO Copyright Treaty"); Directive 2006/115/EC of the European Parliament and of the Council of 12 December 2006 on Rental Right and Lending Right and on Certain Rights to Copyright in the Field of Intellectual Property , 2006 O.J. (L 376) 28–35; Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the Legal Protection of Computer Games, 2009 O.J. (L 111) 16–22; Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on Certain Permitted Uses of Orphan Works, 2012 O.J. (L 299) 5–12; and Council Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on Collective Management of Copyright and Related Rights and Multi-Territorial Licensing of Rights in Musical Works for Online Use in the Internal Market, 2014 O.J. (L 84) 72–98. Other frameworks are relevant but go beyond the scope of this paper. *See, e.g.*, Marrakesh Agreement Establishing the World Trade Organization, Apr. 15, 1994, 1869 U.N.T.S. 299, Annex 1C, Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS); For more, see *Patents for Software?: European Law and Practice*, EUROPEAN PATENT OFFICE, https://www.epo.org/news-issues/issues/software.html [https://web.archive.org/web/20180613235106/http://www.epo.org/news-issues/issues/software.html].

[438] *See* Amendments by the European Parliament to the Commission Proposal Directive (EU) 2019/…of the European Parliament and of the Council of on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC, A800245/271, art. 4, http://www.europarl.europa.eu/doceo/document/A-8-2018-0245-AM-271-271_EN.pdf?redirect [https://perma.cc/KC48-HY8M].

[439] *Id.* at 10–11.

new trends."[440] For text and data mining activities in such research environments, the Directive pushes for exceptions to the copyright regime (e.g., foregoing a need for license agreements[441] or remuneration[442]), as well as for exemptions from the Database Directive[443] to uses of data to monitor trends.[444]

These exemptions are concerning when considered alongside the GDPR's exemptions in Articles 85[445] and 89, which already grant exemptions from most of the rights granted in the GDPR (e.g., Articles 14, 15, 16, 18, 17(3)(d) and 21) for data controllers "processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes."[446] Recital 159 of the GDPR explains that "scientific research purposes" should be interpreted

[440] *Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, COM (2016) 593 final (Sept. 14, 2016) at 14.

[441] *See id*. at Recitals 8–9 and Article 3. For arguments in favor of license fees and access to data for AI training, see generally Schönberger, *supra* note 435.

[442] *Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, at art. 15, COM (2016) 593 final (Sept. 14, 2016).

[443] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases, 1996 O.J. (L 77) 20-28.

[444] *Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, COM (2016) 593 final (Sept. 14, 2016) at art. 3(1) (providing exemptions for text and data mining). Article 2(2) of the Proposal defines "text and data mining" as "any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations." *Id.* at art. 2(2).

[445] Article 85 addresses the inclusion of journalistic purposes in these exemptions.

[446] Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 89.

broadly to include "privately funded research."[447] Universities and research institutes covered by the new Copyright Directive will therefore receive substantial exemptions to data protection and IP requirements when constructing algorithmic models.

In fact, the draft of the Copyright Directive recently passed by the EP goes even further and states in Recital 11 that:

> Union research policy, which encourages universities and research institutes to collaborate with the private sector, research organisations should also benefit from such an exception when their research activities are carried out in the framework of public-private partnerships. While research organisations and cultural heritage institutions should continue to be the beneficiaries of that exception, they should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools.[448]

These privileges cover "access to content that is freely available online"[449] and there is no longer any clear storage limitation.[450] At the same time the private partner must not have "decisive influence" and research carried-out on a not-

---

[447] *Id.* at 30. For a discussion on the legal problems associated with for-profit research, see Tal Z. Zarsky, *Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society*, 56 ME. L, REV. 14 (2004). For a focus on the GDPR, see Gabe Maldoff, *How GDPR Changes the Rules for Research*, INT'L ASS'N OF PRIVACY PROFS. (Apr. 19, 2016), https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/ [https://perma.cc/PV6J-VGBS].

[448] Amendments by the European Parliament to the Commission Proposal Directive (EU) 2019/…of the European Parliament and of the Council of on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC, A800245/271, at 11, http://www.europarl.europa.eu/doceo/document/A-8-2018-0245-AM-271-271_EN.pdf?redirect [https://perma.cc/KC48-HY8M].

[449] *See id.* at 14.

[450] *See id.* at arts. 3(3), 4(3).

for-profit basis or without public-interest mission does not enjoy these privileges.[451]

The likely impact of the draft Copyright Directive appears to be to exempt research institutions, including public-private partnerships from the copyright regime for data mining. Users will thus have no control over how their data is used to build models under the GDPR's statistical exemptions, and under the Copyright Directive's research exemptions.

It must be noted that at the moment the actual impact of the Directive on inferential analytics and algorithmic models remains unclear and the framework is still in Trilogue negotiations.[452]

C. Algorithmic Models and Outcomes and Intellectual Property Law

Thus far, this Article has determined that data subjects are unlikely to have data protection rights over statistical models (e.g. those produced by machine learning) applied to them or built from their personal data under the GDPR. With regard to the EU Copyright Directive, if an algorithm is trained in a research environment via data mining, consent, license agreements, and remuneration are not required to use data as inputs to train the model. Therefore, these regulations could also form a new barrier to control over inferences.

In addition to the legal status of training data addressed thus far, there is growing debate on whether the data generated or creative "work" performed by algorithms should fall under intellectual property law. If IP law is applicable, business interests will be pitted against data subjects'

451 *See id.* at 12.

452 For all draft reports of the European Parliament, see *Draft Reports*, EUROPEAN PARLIAMENT COMMITTEES, http://www.europarl.europa.eu/committees/en/juri/draft-reports.html?ufolderComCode=JURI&ufolderId=07947&urefProcCode=&linkedDocument=true&ufolderLegId=8&urefProcYear=&urefProcNum [https://perma.cc/T7EH-Z5U7]. For further legal and ethical discussion, see Bart W. Schermer, *The Limits of Privacy in Automated Profiling and Data Mining*, 27 COMPUTER L. & SECURITY REV. 45 (2011); Zarsky, supra 65.

rights.[453] This means that the new EU Copyright Directive or the InfoSoc Directive 2001/29/EC[454] could apply to work generated by algorithms, in addition to training data.[455]

In any case, Directive 2009/24/EC on the protection of computer programs applies to software. Here, software is interpreted broadly, as Art 1(2) states that the "Directive shall apply to the expression in any form of a computer program."[456] In the ECJ's judgment in *SAS Institute Inc. v. World Programming Ltd*, this has been interpreted as applying to at least preparatory design material, machine code, source code,

---

[453] Madeleine de Cock Buning, *Is the EU Exposed on the Copyright of Robot Creations?*, 1 ROBOTICS L. J. 8, 8 (2015) ("It can either be the creator of the software who is deemed the owner of the rights; or it could be the owner of the software; or it could be both. It can also be the entity or person who invested financially in the software."); *see also* CHRISTOPHE LEROUX ET AL., SUGGESTION FOR A GREEN PAPER ON LEGAL ISSUES IN ROBOTICS: CONTRIBUTION TO DELIVERABLE D3.2.1 ON ELS ISSUES IN ROBOTICS (2012), https://www.researchgate.net/publication/310167745_A_green_paper_on_legal_issues_in_robotics [https://perma.cc/8752-NYDD]; Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2013 (INL)), EUR. PARL. DOC. A8-0005/2017 (Jan. 27, 2017), http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0005+0+DOC+PDF+V0//EN [https://perma.cc/P5BB-TFTC]; Malgieri, *supra* note 302; *Autonomous Creation – Creation by Robots: Who Owns the IP Rights?*, IPKM BLOG (Mar. 5, 2015), https://law.maastrichtuniversity.nl/ipkm/autonomous-creation-creation-by-robots-who-owns-the-ip-rights/ [https://perma.cc/85MC-3L2C].

[454] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society, 2001 O.J. (L 167) 10.

[455] Note there is also a discussion on whether algorithms should be equipped with personhood to be able to hold copyright, or alternatively whether copyrights should be transferred to the user or coder of the system. For discussion, see Bridy, *supra* note 435; James Grimmelmann, *There's No Such Thing as a Computer-Authored Work - And It's a Good Thing, Too*, 39 COLUM. J.L. & ARTS 403 (2016); Schönberger, *supra* note 435 (exploring the idea that the AI creation and the copyright should be in the hands of the public domain).

[456] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the Legal Protection of Computer Programs (Codified Version) (Text with EEA Relevance), 2009 O.J. (L 111) 16.

and object code, but not the functionality of the computer program or the format of the data files.[457] Following this judgment, while it remains unclear whether the output of software (here, a model or an inference) is protected under Directive 2009/24/EC, information about how the output was produced will be protected. IP law can thus form an additional barrier to accessing the reasoning or analysis that has led to a model or inference.

D. Algorithmic Models and Outcomes and Trade Secrets

The final framework to discuss as a potential barrier to the right to reasonable inferences is a "catch all" framework that may pose a substantial barrier to learning the justification behind inferences. Even if the aforementioned frameworks were not to apply to inferential analytics, the new EU Trade Secrets Directive[458] is likely to substantially limit controllers' transparency obligations.[459] The framework, which came into effect on June 9, 2018, may result in the creation of new data being classified as a trade secret. Article 2 of the Directive defines a trade secret as any information that is not "generally known," has commercial value due to this secrecy, and has been subject to reasonable steps to ensure it remains a

457 *See* Case C-406/10, SAS Inst. Inc. v. World Programming Ltd., 2012 E.C.R. I-259.

458 Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure, 2016 O.J. (L 157) 1.

459 *See* Rembert Niebel, Lorenzo de Martinis & Birgit Clark, *The EU Trade Secrets Directive: All Change for Trade Secret Protection in Europe?*, 13 J. INTELL. PROP. L. & PRAC. 445, 448–49 (2018).

secret.[460] Recital 1 further adds "valuable know-how and business information" to the definition.[461]

The definition of a trade secret is so broad as to include nearly any data handled by a commercial entity. For example, trade secrets could include "shopping habits and history of customers,"[462] "customer lists and profiles,"[463] "algorithms,"[464] and "[information about a] customer's behavior (creditworthiness, lifestyle, reliability, etc.), personalized marketing plans (e.g. pricing), or forecasts about [a] customer's future life based on probabilistic studies (life expectancy, estimated advancements in career, etc.)."[465]

---

[460] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure, 2016 O.J. (L 157) 1, art. 2. Article 39(2) of the TRIPS agreement has a similar definition of a trade secret. Marrakesh Agreement Establishing the World Trade Organization, Apr. 15, 1994, 1869 U.N.T.S. 299, Annex 1C, Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) art. 39(2). For discussion of trade secrets as a hindrance to due process and algorithmic accountability, see generally PASQUALE, *supra* note 11; Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018); Amy J. Schmitz, *Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots"*, 2014 MICH. ST. L. REV. 1411 (2014); Brenda Reddix-Smalls, *Credit Scoring and Trade Secrecy: An Algorithmic Quagmire or How the Lack of Transparency in Complex Financial Models Scuttled the Finance Market*, 12 U.C. DAVIS BUS. L.J. 87 (2011).

[461] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure, 2016 O.J. (L 157) 1, 1.

[462] Graef, Husovec & Purtova, *supra* note 302, at 1381.

[463] Purtova, *supra* note 369, at 71.

[464] Guido Noto La Diega, *Against the Dehumanisation of Decision-Making: Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, 9 J. INTELL. PROP. INFO. TECH. & ELECTRONIC COM. L. 3, 12 (2018).

[465] Malgieri, *supra* note 273, at 113–14 (internal citations omitted). According to Malgieri, disclosing, rectifying, or erasing any of these data "can probably adversely affect the 'dynamic' trade secret interest of business people and of employees." *Id.* at 114.

An EDPS document commenting on an early draft of the Directive[466] and a European Commission impact assessment accompanying the proposal for the Directive[467] further clarify the scope of trade secrets. According to these sources, trade secrets can consist of "data such as information on customers and suppliers, business plans or market research and strategies,"[468] "list[s] of clients/ customers; internal datasets containing research data,"[469] "private collations of individual items of publicly available information,"[470] as well as "data on customers and their behaviour and on the ability to collect and to monetise those data." The inclusion of customer data shows that personal data, subject to data protection law, can nonetheless constitute trade secrets.[471] Tension between individual privacy interests and business interests, or data protection and trade secrets laws, is thus inevitable.

The EDPS foresaw these possible tensions, urging "greater precision on the concept of trade secrets and clearer safeguards . . . to address adequately the potential effects of the proposal on the rights to privacy and to the protection of personal data."[472] The EDPS also recommended amending Article 4 of the Trade Secrets Directive to ensure that the data

---

[466] European Data Prot. Supervisor, *Opinion of the European Data Protection Supervisor on the Proposal for a Directive of the European Parliament and of the Council on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure* (Mar. 12, 2014), https://edps.europa.eu/sites/edp/files/publication/14-03-12_trade_secrets_en.pdf [https://perma.cc/7UE9-8WB6].

[467] *Commission Staff Working Document: Impact Assessment: Accompanying the Document Proposal for a Directive of the European Parliament and of the Council on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure*, at 107–18, 248–62, COM (2013) 813 final (Nov. 28, 2013).

[468] European Data Prot. Supervisor, *supra* note 466, at 3.

[469] *Id.*

[470] *Id.*

[471] *Id.*

[472] *Id.* at 2.

subject's "right to access the data being processed and to obtain rectification, erasure or blocking of the data where it is incomplete or inaccurate"[473] is guaranteed, referring to a case involving Facebook[474] where requests were denied. This suggestion was not adopted but rather moved to Recital 35. The final Directive in Article 9(4) only requires that "any processing of personal data pursuant to paragraphs 1, 2 or 3 shall be carried out in accordance with Directive 95/46/EC,"[475] without any clarification as to resolving the tension between trade secrets and data protection law. It is thus unclear how these clashes will play out, although Member States may implement new rules.

In any case, given the broad definition of trade secrets and the clear inclusion of personal data in its scope, it is safe to assume that derived and inferred data will be covered by the Trade Secrets Directive.[476] Even with this outlook, a fair balance between the right of privacy, IP laws, and the rights to conduct a business and freedom of expression will be necessary; the ECJ's jurisprudence has long reflected this position.[477]

---

[473] *Id.* at 5.

[474] Letter from Facebook User Operations—Data Access Request Team, to Max Schrems, (Sept. 28, 2011), http://www.europe-v-facebook.org/FB_E-Mails_28_9_11.pdf [https://perma.cc/B3TZ-UK4R].

[475] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure, art. 9(4), 2016 O.J. (L 157).

[476] For an overview of the definition of trade secrets according to the ECJ and its constituent courts, see Case T-353/94, Postbank NV v. Comm'n of the European Cmtys., 1996 E.C.R. II-921, and Case T-198/03, Bank Austria Creditanstalt AG v. Comm'n of the European Cmtys., 2006 E.C.R. II-1429.

[477] *See* Case C-275/06, Productores de Música de España (Promusicae) v. Telefónica de España SAU, 2008 E.C.R. I-271; Case C-70/10, Scarlet Extended SA v. Société Belge des Auteurs, Compositeurs et Éditeurs SCRL (SABAM), 2011 E.C.R. I-00000; Case C-557/07, LSG-Gesellschaft zur Wahrnehmung von Leistungsschutzrechten GmbH v. Tele2 Telecommunication GmbH, 2009 E.C.R. I-1227; Case C-461/10, Bonnier Audio AB, Earbrooks AB, Norstedts Förlagsgrupp AB, Piratförlaget AB & Storyside AB v. Perfect Commc'n Sweden AB, ECLI:EU:C:2012:219.

Taking into account the novel risks of inferential analytics and trends in the European legal landscape that appear to place greater emphasis on commercial and research interests, implementation of a right to reasonable inferences takes on renewed importance to ensure that the level of protection against inferences increases to reasonable standards. Data subjects require a new right addressing the riskiest type of personal data that, ironically, currently receives the least protection.

VIII. CONCLUSION AND RECOMMENDATIONS

Calls for accountability in Big Data analytics and algorithmic decision-making systems are motivated by a common concern: Assessments and inferences drawn from disparate, often non-intuitive features and data sources increasingly drive decision-making about people. These inferences are based not only on data individuals have provided or has been observed, but also on information derived or inferred from it, as well as from anonymous or third-party data. Similarly, inferential analytics can be used to infer our preferences, weaknesses, sensitive attributes (e.g. race or sexual orientation), and opinions (e.g. political stances). These can form the basis for micro-targeting, nudging, and manipulation, as seen in online advertisement[478] or the recent Cambridge Analytica scandal. Too much emphasis is placed on governing the collection of these types of data, while too little is paid to how it is evaluated.[479]

To illustrate, even if a bank can explain which data and variables have been used to make a decision (e.g. banking records, income, post code), the decision turns on inferences drawn from these sources; for example, that the applicant is not a reliable borrower. This is an assumption or prediction about future behavior that cannot be verified or refuted at the time of decision-making. Thus, the actual risks posed by Big

[478] Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014).

[479] *See* Wachter, *supra* note 316.

Data analytics and AI are the underpinning inferences that determine how we, as data subjects, are being viewed and evaluated by third parties.

This Article has considered whether inferences or derived data constitute personal data according to the Article 29 Working Party's three-step model and jurisprudence of the European Court of Justice. If inferences are seen as personal data, the rights in the GDPR could apply and allow data subjects to know about (Articles 13–14), access (Article 15), rectify (Article 16), delete (Article 17), and object to them (Article 21). Further, profiling and automated decision-making, which may include inferences, can already be contested (Article 22). The Article 29 Working Party sees verifiable and unverifiable inferences as personal data (e.g. results of a medical analysis), but leaves open whether the reasoning and process behind that inference is seen as personal data. The ECJ is still finding its voice on this topic, as its current jurisprudence is inconsistent. Future jurisprudence will continue to define the scope of personal data and the protection afforded to it. It is crucial to note that the question of whether inferences are personal data, is not the most important one. The underlying problem goes much deeper and relates to the tension of whether individuals have rights, control, or recourse over how they are seen by others.

Some scholars are worried that broad interpretation of personal data turns data protection law into the "law of everything."[480] However, as shown in Section V, inferences are treated as "economy class" personal data that are afforded little meaningful protection, and certainly less than personal data provided by the data subject or sensitive personal data. In part, third parties may have an interest in inferences and derived data and the techniques used to create it (e.g. trade secrets) due to their value or the costs involved.

The GDPR, the draft e-Privacy regulation, the Digital Content Directive, and legal scholars attribute only limited rights over inferences to data subjects. At the same time, new frameworks such as the EU Copyright Directive and

[480] Purtova, *supra* note 86.

provisions in the GDPR push to facilitate data mining, knowledge discovery, and big data analytics by limiting data subjects' rights over their data. The new Trade Secrets Directive also poses a barrier to accountability, as models, algorithms, and inferences may very well fall under this framework.

Even if the ECJ decides to consistently classify inferences as personal data, current jurisprudence is a strong indicator that the court will offer insufficient protection against unreasonable inferences under data protection law. The core problem stems from how the ECJ interprets the remit of data protection law. In standing jurisprudence, the ECJ (in *Bavarian Lager*,[481] *YS and M and S*,[482] and *Nowak*[483]) and Advocate General (in *YS and M and S*[484] and *Nowak*[485]) have consistently explained that the remit of data protection law is not to assess whether inferences and decisions based upon them are accurate or justified. Rather, individuals need to consult sectoral laws and governing bodies applicable to their specific case to seek possible recourse. More generally, the ECJ views data protection law as a tool for data subjects to assess whether the (input) data undergoing processing was legally obtained, and whether the purpose for processing is lawful. To ensure this, data protection law grants various rights to individuals, for example the rights of access, rectification, and deletion.[486] Of course this can change in the future, as the definition of personal data and the associated rights depend on the purpose for which it was collected. As the rights in the GDPR must be interpreted teleologically, it is not

[481] Case C-28/08 P, European Comm'n v. Bavarian Lager, 2010 E.C.R. I-6055.

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

[483] Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994.

[484] Joined Cases C-141/12 & C-372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2013 E.C.R. I-838.

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

[486] Case C-553/07, College van Burgemeester en Wethouders van Rotterdam v. M. E. E. Rijkeboer, 2009 E.C.R. I-3889.

unthinkable that future jurisprudence could apply these rights to the content of assessments and inferences. A change is, however, unlikely in inherently antagonistic situations that pose data subjects' rights to privacy, identity, reputation against companies' rights to freedom of contract and free speech. Dialogue is needed to determine the point at which the right to privacy must take precedence over the private autonomy of decision-makers.

This situation is ironic, as data subjects are most in need of protection from the risks posed by inferences and derived data. To close these accountability gaps and promote justification of inferences, this Article proposes a new "right to reasonable inferences" applicable to "high risk" inferences that cause damage to privacy or reputation, or have low verifiability in the sense of being predictive or opinion-based while being used for important decisions. This right would require *ex-ante* justification to be given by the data controller to establish whether an inference is reasonable. This disclosure would address (1) why certain data are normatively acceptable bases to draw inferences; (2) why these inferences are normatively acceptable and relevant for the chosen processing purpose or type of automated decision; and (3) whether the data and methods used to draw the inferences are accurate and statistically reliable. An ex-post mechanism would allow data subjects to challenge unreasonable inferences, which can support challenges against automated decisions exercised under Article 22(3) of the GDPR.

Of course, a solution outside of data protection law may be possible.[487] However, few standards exist, especially in the private sector, that govern how decisions are made. A right to reasonable inferences is an essential response to the novel risks introduced by inferential analytics. It is both the essence and the extension of data protection law.

[487] *See generally* Bert-Jaap Koops, *The Trouble with European Data Protection Law*, 4 INT'L DATA PRIVACY L. 250 (2014).

In the same way it was necessary to create a "right to be forgotten" in a Big Data world,[488] it is now necessary to create a "right on how to be seen." The proposed re-imagining of the purpose of data protection law would be more in line with the original remit proposed in the ECHR,[489] as well as the Council of Europe's Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data[490] and its guidelines on AI,[491] and the European Parliament's resolution on a comprehensive European industrial policy on artificial intelligence and robotics.[492] It would reconfigure privacy as a holistic concept with a stronger focus on adaptable identity, self-presentation, and reputation. One could also argue for a mediated application of the human right of privacy, and advocate for a "positive obligation" of states to implement laws to protect citizens from privacy invasion by the public and private sectors.[493]

Based on the preceding analysis of the legal status and protection of inferences, the following recommendations can be made for European policy:

A. Re-Define the Remit of Data Protection Law

In order to ensure data protection law protects against the novel risks introduced by Big Data analytics and algorithmic decision-making, the ECJ should re-define the law's remit to include assessment of the reasonableness of inferential analytics and accuracy of decision-making processes. However, it has to be noted that the court's limitation of Article 16 made sense in this regard in the discussed case law. It would be an odd situation where data protection authorities are competent to rule on the accuracy of immigration cases or

488 *See generally* MAYER-SCHÖNBERGER, *supra* note 10; VAN HOBOKEN, *supra* note 8.

489 See also ECHR jurisprudence on privacy until 2017, reviewed in Council of Europe, *supra* note 60.

490 Comm. of Ministers, *supra* note 374.

491 Directorate Gen. of Human Rights and Rule of Law, *supra* note 375.

492 *See* Artificial Intelligence and Robotics, *supra* note 376.

493 De Hert & Gutwirth, *supra* note 73, at 61; Wachter, *supra* note 54.

examination disputes. In these cases, procedures are in place to deal with complaints. However, the same cannot always be said for inferences that the private sector draws. It is often left to the private autonomy of industry to assess and evaluate people. Companies are relatively free in how they assess people, except where laws exist (e.g. anti-discrimination law) that limit this freedom. As discussed in Part II, due to the widespread implementation of inferential analytics by companies for profiling, nudging, manipulation, or automated decision-making, these "private" decisions can to a large extent impact the privacy of individuals. Thus, dialogue is needed to determine the point at which the right to privacy must be given greater weight than the private autonomy of decision-makers. In effect, individuals should have a right to be reasonably assessed.

Data protection is only one component of the right to privacy, which also includes a right to identity, reputation, self-presentation, and autonomy. Big Data analytics produces privacy-invasive, unpredictable, and counterintuitive inferences that threaten these components of privacy. In response, data subjects require greater control over when, how, and under what conditions they are being assessed by automated systems.

B. Focus on How Data is Evaluated, Not Just Collected

The categories of personal, sensitive, anonymous and non-personal data reflect characteristics of data when it is collected, and they determine the level of protection granted to input data. These characteristics can, however, change over time, as data is used for different purposes. The German Supreme Court has previously argued that there is no such thing as "irrelevant data" when it comes to data protection law, as informational technologies might use it for purposes that affect the data subject. Seemingly neutral data can be

turned into data that affects the right to privacy,[494] or offers grounds for discrimination and other harms.

Basing protections on these distinctions is thus ineffective.[495] The damage that can be done by data does not depend on any of these categories, but rather on how it is used. Inferences or profiles drawn from any of these sources can be applied to and harm an individual or group. The belief that certain categories of data are fundamentally less harmful or risky than others is undermined by Big Data analytics.[496]

This Article recommends adopting the position taken by the Article 29 Working Party concerning the transformation of categories of data based upon processing purposes and impact.[497] In future European policy-making and jurisprudence, levels of protection should be granted to data based primarily on its usage and impact, and secondarily on its source.

C. Do Not Focus Only on the Identifiability of Data Subjects

In order for data protection rights to apply, data must be suitable to identify the individual. This is misguided,[498] because the identifiability of data is fluid and can change over time[499] depending on linkage, re-identification attacks, and other technological progress.[500]

---

494 See Ernst, *supra* note 91 (citing a judgment of the German Constitutional Court).

495 Schreurs and others argue that anonymous data can still impact data subjects, despite being outside the scope of data protection law. *See* Schreurs et al., *supra* note 359, at 248–49. Zarsky explains that any data could potentially become sensitive data, rendering the classification meaningless. *See* Zarsky, *supra* note 317, at 1013.

496 *See* Wachter, *supra* note 316, at 6.

497 Article 29 Data Prot. Working Party, *supra* note 81, at 8, 10, 11.

498 *See* Wachter, *supra* note 316,

499 Wachter, *supra* note 316.

500 *See* Korff, *supra* note 323, at 46.

Companies can use anonymization[501] techniques to avoid many obligations under data protection law. Similarly, pseudonymization techniques[502] can potentially minimize the requirements to respect individual rights. In such cases, data controllers are not required to comply with requests from data subjects under Articles 15–20 if they are "not in a position to identify" him or her, unless the data subject can provide additional information that allows the data to be re-identified.[503] Together, these provisions could create an incentive to de-identify data in order to avoid compliance with individual rights, which has happened in the past.[504]

As argued above, inferences drawn from anonymous and non-personal data still pose risks for data subjects.[505] As a result, identifiability as a prerequisite to exercise individual rights creates a gap in the protection afforded to data subjects against inferential analytics. The potential and actual harm of inferential analytics should be reflected in future European policy-making and jurisprudence, regardless of whether the affected parties can be identified.[506] This is not to suggest that data subjects should be granted rights over personal and anonymous data which has not been applied to them. Rather, improved channels of redress are required against models, profiles, and other background knowledge built from third-party and anonymous data and subsequently applied to identifiable individuals.

---

501 *See* Schreurs et al., *supra* note 359, at 248–49.

502 *See* Frederik J. Zuiderveen Borgesius, *Singling Out People Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation*, 32 COMPUTER L. & SECURITY REV. 256 (2016).

503 Se Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 11(2).

504 For example, there are similar prior experiences where companies have claimed "disproportionate effort" to avoid compliance with access requests. *See, e.g.*, Letter from Facebook User Operations, *supra* note 474.

505 *See supra* Section II.A. and Part VI.

506 Mittelstadt, *supra* note 25; Mantelero, *supra* note 25; Bygrave, *supra* note 349, at 283–95.

D. Justify Data Sources and Intended Inferences Prior to Deployment of Inferential Analytics at Scale

Following the recommendation to implement a right to reasonable inferences, data controllers should proactively justify their design choices for high-risk inferential analytics prior to widespread deployment. Inspiration can be drawn from the German data protection law's provisions on predictive assessments, such as credit scoring. Controllers should pay increased attention to addressing the following aspects of the source data and outputs of inferential analytics in addressing justification:

- The privacy invasiveness and the counter-intuitiveness of the data sources used to draw inferences, for example clicking behavior, browsing behavior,[507] or mouse tracking.[508]
- The aim of the inference to be drawn should justify the means or sources of data being used in terms of invasiveness. Inferring gambling or alcohol addiction to drive targeted advertising, for example, may actively harm the data subject.
- The usage of known proxy features (e.g. post code) or other (potentially) discriminatory data, or the intention to infer sensitive attributes (e.g. political views[509]) from non-sensitive data.
- The normative acceptability and relevance of the source data and inference to a particular processing purpose, such as, the relevance of Facebook profiles and friend networks to loan decisions.[510]
- The statistical reliability of the methods used to draw inferences.

This is a preliminary list of potential topics and information types to be included in justification disclosures under the right to reasonable inferences. Extensive debate and further research is required to determine which

507 *See* Allerhand et al., *supra* note 41.

508 *See* Chen et al., *supra* note 46.

509 *See* Coutts, *supra* note 34.

510 *See* Taylor & Sadowski, *supra* note 33.

information should be included in different sectors. The myriad applications of inferential analytics demand a sectoral approach.

E. Give Data Subjects the Ability to Challenge Unreasonable Inferences

In line with the implementation of a right to reasonable inferences, European policymakers should grant data subjects a new right to challenge unreasonable high-risk inferences, which can also support challenges to subsequent decisions.[511] Data subjects can raise an objection with the data controller on the grounds that the inference or its source data is irrelevant or unreliable.[512] For verifiable inferences, the data subject can provide supplementary information to rectify the inaccurate inference. For non-verifiable and subjective inferences, supplementary information can also be provided to attempt to convince the data controller to change its assessment.

The right to rectification (Article 16 of the GDPR) may arguably already offer a remedy for non-verifiable and subjective inferences and opinions, depending upon one's view of the necessity of verifiability in classifying inferences as personal data.[513] Taking this view, the right to reasonable inferences would embed an answer to the verifiability question in law, and thus strengthen data protection rights over inferences regardless of their verifiability and subjectivity. Similarly, it would complement the existing right to contest solely automated decisions[514] and profiling[515] with legal and significant effects, and potentially transform it from

---

511 For a favorable view of such a solution, see Hildebrandt & Koops, *supra* note 402, at 449. On the need to remedy unjust judgments based on inferences, see Leenes, Hildebrandt & Gutwirth, *supra* note 73, at 295.

512 *See supra* Section VI.B.

513 *See supra* Section III.B, Part IV, Section V.B.

514 For a discussion on this legal loophole, see Wachter, Mittelstadt & Floridi, *supra* note 11; Bygrave, *supra* note 349.

515 Mendoza & Bygrave, *supra* note 332.

a merely procedural tool to a meaningful accountability mechanism.[516]

The intention of an ex-post right to contest unreasonable inferences is, however, not to guarantee that a data controller must change its inference or assessment at the data subject's request. Rather, it aims to establish a dialogue between data controllers and subjects in which the former share details and justifications for the proposed inferential processing that are open to comments and interrogation by the latter.[517] This will be fruitful for both sides, as accurate assessment is in the interests of both parties. To achieve this, it will be necessary to redefine the purpose of data protection law (as suggested above) to include justification of assessments. Strengthening the position of the data subject in relation to controllers is necessary to sufficiently mitigate the novel risks of inferential analytics.[518]

Given the novel risks of Big Data analytics and algorithmic decision-making, inferences cannot justifiably remain "economy class" personal data. Data subjects' privacy interests require renewed protection to restore the fair balance between individual, public, and commercial interests that inspires data protection law. The current remit of data protection law works well to govern input data, but fails to provide meaningful control over how personal data is evaluated. A right to reasonable inferences is a first step to correct this imbalance.

---

[516] *See supra* Section V.E.

[517] Article 22(3) of the GDPR also allows data subjects to express their views and human intervention (in addition to the right to contestation) if a solely automated decision has been made. Regulation (EU) 2016/679 of the European Parliament and of the Council 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, 2016 O.J. (L 119) art. 22(3).

[518] *See supra* Section II.A.