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

Sandra Wachter[1] *& Brent Mittelstadt*[2]

INTRODUCTION[3]

Big Data analytics and artificial intelligence (AI) draw non-intuitive and unverifiable inferences and predictions about the behaviours, 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.[4] The intuitive link between actions and perceptions is being eroded, leading to a loss of control over identity and how we 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 about us that we cannot predict, understand, or refute.

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 broad concept of personal data in Europe could be interpreted to include inferences, predictions, and assumptions that refer to or impact on an individual. If seen as personal data, individuals are 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[5] and the European Court of Justice.

---

[1] 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.

[2] 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..

[3] 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. We would also like to thank Dr. Alessandro Spina and Prof. Manfred Stelzer for their insightful and considerate comments from which this paper benefitted greatly.

[4] Brent Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society <http://bds.sagepub.com/lookup/doi/10.1177/2053951716679679> accessed 15 December 2016.

[5] It is worth noting that as of the GDPR's implementation on May 25, 2018, the Article 29 Working Party has ceased to exist and been succeeded by the European Data Protection

As we will show in this paper, 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 controller's interests (e.g. trade secrets, 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; Bavarian Lager,[6] YS. and M. and S.,[7] and Nowak[8]) and the Advocate General (AG; YS. and M. and S.[9] and Nowak[10]) have 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.[11] Critically, the ECJ has likewise made clear that data protection law is not intended to ensure the accuracy of decisions and

Board. One of the first acts of the Board 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: European Data Protection Board, 'Endorsement 01/2018' (2018) 01/2018 <https://edpb.europa.eu/sites/edpb/files/files/news/endorsement_of_wp29_documents.pdf> accessed 12 September 2018. Documents pertaining to the 1995 Data Protection Directive have not yet been addressed, although the Board has met only twice as of September 2018. In this time, the Board has also produced two sets of guidelines addressing Articles 42 and 43 of the GDPR, neither of which are relevant for our purposes here. Throughout the paper we therefore continue 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 that the European Data Protection Board 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.

[6] *Commission v Bavarian Lager - Case C‑28/08 P* (European Court of Justice (Grand Chamber)) para 49.

[7] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (European Court of Justice (Third Chamber)) para 45-47.

[8] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (European Court of Justice (Second Chamber)) para 54-55.

[9] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (Advocate General Sharpston) para 32 and 60.

[10] *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (Advocate General Kokott) para 54-55.

[11] *College van burgemeester en wethouders van Rotterdam v M E E Rijkeboer C-553/07* (Judgment of the Court (Third Chamber)).

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 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.[12]

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 Trades Secrets Directive provides extensive protection of commercial interests attached to the outputs of these processes (e.g. models, algorithms and inferences).

In this paper we argue that a new data protection right, the 'right to reasonable inferences', is needed to help close the accountability gap currently posed 'high risk inferences' , meaning inferences that are privacy invasive or reputation damaging and have low verifiability in the sense of being predictive or opinion-based.[13] In cases where algorithms draw 'high risk inferences' about individuals, 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 is a relevant basis to draw inferences; (2) why these inferences are 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. The *ex-ante* justification is 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 IP and trade secrets law as well as freedom of expression[14] and Article 16 of the EU Charter of Fundamental Rights: the freedom to conduct a business.

---

[12] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 45-47; *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 54-55; *Commission v Bavarian Lager - Case C‑28/08 P* (n 6) para 49.

[13] Sandra Wachter, 'Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR' (2018) 34 Computer Law & Security Review 436; Sandra Wachter, 'GDPR and the Internet of Things: Guidelines to Protect Users' Identity and Privacy' <https://papers.ssrn.com/abstract=3130392> accessed 5 August 2018.

[14] Joris Van Hoboken, *Search Engine Freedom: On the Implications of the Right to Freedom of Expression for the Legal Governance of Web Search Engines* (Kluwer Law International

In Section I, we first examine 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. Section II reviews the position of the Article 29 Working Party on the legal status of inferences. Section III 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. Section IV 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, Section V then describes the aims and scope of the proposed 'right to reasonable inferences'. Section VI then examines barriers likely to be encountered in the implementation of the proposed right, drawing from data protection law, as well as intellectual property (IP) law and the new EU Trade Secrets Directive. 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 it as it was necessary create a "right to be forgotten" in a Big Data world,[15] we think is it now necessary to create a "right on how to be seen."

I. 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,[16] ranging

Den Haag 2012); Joris van Hoboken, 'The Proposed Right to Be Forgotten Seen from the Perspective of Our Right to Remember' [2013] Freedom of Expression Safeguards in a Converging Information Environment, Prepared for the European Commission, Amsterdam.

[15] Viktor Mayer-Schönberger, *Delete: The Virtue of Forgetting in the Digital Age* (Princeton University Press 2011); van Hoboken (n 14).

[16] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why There Is No Right to Explanation in the General Data Protection Regulation' [2017] International Data Privacy Law <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469>; Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' [2017] arXiv preprint arXiv:1711.00399; Finale Doshi-Velez and others, 'Accountability of AI Under the Law: The Role of Explanation' [2017] arXiv preprint arXiv:1711.01134; Joshua A Kroll and others, 'Accountable Algorithms' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2765268 <http://papers.ssrn.com/abstract=2765268> accessed 29 April 2016; Jenna Burrell, 'How the Machine "Thinks:" Understanding Opacity in Machine Learning Algorithms' [2016] Big Data & Society; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015); Viktor Mayer-Schönberger and Thomas Ramge, *Reinventing Capitalism in the Age of Big Data* (Basic Books 2018); SC Olhede and PJ Wolfe, 'The Growing Ubiquity of Algorithms in Society: Implications, Impacts and Innovations' (2018) 376 Phil. Trans. R. Soc. A 20170364.

from calls for regulatory instruments and the development of technical methods for providing explanations, to standards setting and mapping of their importance for accountability in public and private institutions.[17] These diverse streams of work have made much progress in legal, ethical, policy, and technical terms. Yet each is united by a common blind spot: a legal or ethical basis is required to justify demands for explanations, and determine their required content.[18] 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.

Unfortunately, there is little reason to assume that organisations will voluntarily offer full explanations covering the process, justification, and accuracy of the decision-making process 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 likewise imposes additional costs and risks for the organisation.

Where a general legal or ethical justification for explanations of algorithmic decisions does not exist,[19] requests will require alternative

[17] See for example: Executive Office of the President National Science and Technology Council Committee on Technology, 'Preparing for the Future of Artificial Intelligence' (Executive Office of the President 2016) <https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf> accessed 11 November 2016; House of Commons Science and Technology Committee, 'Algorithms in Decision-Making' (2018) HC 351; European Parliament, 'Civil Law Rules on Robotics - European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))' (European Parliament 2017) P8_TA-PROV(2017)00 51 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN> accessed 14 April 2017; Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' [2017] arXiv preprint arXiv:1706.07269; Corinne Cath and others, 'Artificial Intelligence and the "Good Society": The US, EU, and UK Approach' [2017] Science and Engineering Ethics 1.

[18] See Doshi-Velez and others (n 16) who share this view.

[19] 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. For discussion, see: Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469>; Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2018) 17/EN WP 251rev.01 <http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826>.

grounds to be successful.[20] We refer to these potential grounds to demand 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 that certain standards be followed within that procedure, such as not basing the hiring decision on a protected attribute (e.g. ethnicity) because doing so would constitute discrimination.

Nonetheless, granting explanations is only one possible way forward to make 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.[21] An explanation might inform the individual about the outcome or decision and about underlying assumptions, predictions or inferences that led to it. It does not, however, ensure that the decision, assumption, prediction or inference are justified.[22] 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 grant data subjects the ability to manage how privacy-invasive inferences are drawn and seek redress once they are created and used for decision-making with significant consequences.

A. The novel risks of inferential analytics and a right to reasonable inferences

In the following sections, we explain how European law is not equipped to protect individuals against the novel risks brought upon by automated decision-making driven by inferential analytics. We argue that a new right –

[20] Doshi-Velez and others (n 16) for example suggest that explanations are justified if accompanied by recourse for harm suffered..

[21] Wachter, Mittelstadt and Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (n 19).

[22] Miller (n 17); Mireille Hildebrandt, 'Primitives of Legal Protection in the Era of Data-Driven Platforms'.

a 'right to reasonable inferences' – might help to close the accountability gap currently posed by these technologies in Europe.[23]

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 discriminatory, biased, and invasive decision-making based on inferential analytics.[24] Modern data analytics has access to unprecedented volumes and varieties of linked up data to assess the behaviours, preferences, and private lives of individuals.[25] 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,[26] and constructing group profiles from personal, third party, and anonymous data.[27]

Numerous applications of 'Big Data analytics' to draw potentially troubling inferences about individuals and groups have emerged in recent years.[28] Major internet platforms are behind many of the highest profile examples: Facebook may be able to infer sexual orientation[29] (which may be

[23] Wachter, 'Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR' (n 13); Sandra Wachter, 'GDPR and the Internet of Things: Guidelines to Protect Users' Identity and Privacy'.

[24] Mittelstadt and others (n 4).

[25] Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data : A Revolution That Will Transform How We Live, Work and Think* (John Murray 2013); Brent Mittelstadt and Luciano Floridi, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2016) 22 Science and Engineering Ethics 303; Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review.

[26] Danielle Keats Citron and Frank A Pasquale, 'The Scored Society: Due Process for Automated Predictions' (Social Science Research Network 2014) SSRN Scholarly Paper ID 2376209 <https://papers.ssrn.com/abstract=2376209> accessed 4 March 2017; Tal Z Zarsky, 'Understanding Discrimination in the Scored Society' (2014) 89 Wash. L. Rev. 1375; Peter Grindrod, *Mathematical Underpinnings of Analytics: Theory and Applications* (OUP Oxford 2014); Peter Grindrod, 'Beyond Privacy and Exposure: Ethical Issues within Citizen-Facing Analytics' (2016) 374 Phil. Trans. R. Soc. A 20160132.

[27] Brent Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' [2017] Philosophy & Technology <http://link.springer.com/10.1007/s13347-017-0253-7> accessed 3 July 2017; Alessandro Mantelero, 'From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era', *Group Privacy* (Springer 2017).

[28] Christopher Kuner and others, *The Challenge of 'Big Data'for Data Protection* (Oxford University Press 2012).

[29] José González Cabañas, Ángel Cuevas and Rubén Cuevas, 'Facebook Use of Sensitive Data for Advertising in Europe' [2018] arXiv preprint arXiv:1802.05030.

illegal), and other protected attributes (e.g. race),[30] political opinions[31], and imminent suicide attempts,[32] while third parties have used Facebook to decide on the eligibility for loans[33] and infer political stances on abortion[34]. Tendencies to depression can be inferred though usage of Facebook[35] and Twitter[36]; Google has attempted to predict flu outbreaks[37] as well as other diseases and their outcomes[38]; and Microsoft can likewise predict Parkinson's disease[39] and Alzheimer's disease[40] from search engine interactions. Other recent potentially invasive applications include prediction of pregnancy in customers by Target,[41] researchers inferring levels of user

[30] 'Facebook Is (Still) Letting Housing Advertisers Exclude Users By Race - Digg' <http://digg.com/2017/facebook-housing-discrimination> accessed 31 July 2018.

[31] Owen Bowcott and Alex Hern, 'Facebook and Cambridge Analytica Face Class Action Lawsuit' *The Guardian* (10 April 2018) <http://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit> accessed 31 July 2018.

[32] 'Facebook Rolls out AI to Detect Suicidal Posts before They're Reported | TechCrunch' <https://techcrunch.com/2017/11/27/facebook-ai-suicide-prevention/?guccounter=1> accessed 29 July 2018.

[33] Astra Taylor and Jathan Sadowski, 'How Companies Turn Your Facebook Activity Into a Credit Score' [2015] *The Nation* <https://www.thenation.com/article/how-companies-turn-your-facebook-activity-credit-score/> accessed 31 July 2018.

[34] 'Anti-Choice Groups Use Smartphone Surveillance to Target "Abortion-Minded Women" During Clinic Visits' (*Rewire.News*) <https://rewire.news/article/2016/05/25/anti-choice-groups-deploy-smartphone-surveillance-target-abortion-minded-women-clinic-visits/> accessed 31 July 2018.

[35] Megan A Moreno and others, 'Feeling Bad on Facebook: Depression Disclosures by College Students on a Social Networking Site' (2011) 28 Depression and anxiety 447.

[36] Moin Nadeem, 'Identifying Depression on Twitter' [2016] arXiv preprint arXiv:1607.07384.

[37] Donald R Olson and others, 'Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales' (2013) 9 PLoS computational biology e1003256.

[38] 'Google AI Can Predict When People Will Die with "95 per Cent Accuracy" | The Independent' <https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-ai-predict-when-die-death-date-medical-brain-deepmind-a8405826.html> accessed 29 July 2018; Ryen W White, P Murali Doraiswamy and Eric Horvitz, 'Detecting Neurodegenerative Disorders from Web Search Signals' (2018) 1 npj Digital Medicine 8; Alvin Rajkomar and others, 'Scalable and Accurate Deep Learning with Electronic Health Records' (2018) 1 npj Digital Medicine 18.

[39] Liron Allerhand and others, 'Detecting Parkinson's Disease from Interactions with a Search Engine: Is Expert Knowledge Sufficient?' [2018] arXiv preprint arXiv:1805.01138.

[40] White, Doraiswamy and Horvitz (n 38).

[41] Charles Duhigg, 'How Companies Learn Your Secrets' *The New York Times* (16 February 2012) <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html> accessed 29 July 2018; Mayer-Schönberger and Cukier (n 25).

satisfaction with search results using mouse tracking,[42] and finally China's far reaching Social Credit Scoring system.[43]

None of these applications can claim to generate inferences or predictions with absolute certainty, and in several cases have suffered highly visible failures (e.g. Google Flu Trends). Many are likewise used solely for targeted advertising. Justification for these invasive uses of personal data is crucial from an ethical[44] as well as legal[45] 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,[46] thus posing risks to privacy[47] and identity,[48] data protection, reputation,[49] and informational self-

[42] Ye Chen and others, 'User Satisfaction Prediction with Mouse Movement Information in Heterogeneous Search Environment' (2017) 29 IEEE Transactions on Knowledge and Data Engineering 2470.

[43] Scott R Peppet, 'Regulating the Internet of Things: First Steps toward Managing Discrimination, Privacy, Security and Consent' (2014) 93 Tex. L. Rev. 85; https://www.facebook.com/simon.denyer?fref=ts, 'China's Plan to Organize Its Society Relies on "Big Data" to Rate Everyone' (*Washington Post*) <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> accessed 31 July 2018.

[44] For an ethical approach of AI accountability and justification see Reuben Binns, 'Algorithmic Accountability and Public Reason' [2017] Philosophy & Technology <http://link.springer.com/10.1007/s13347-017-0263-5> accessed 24 April 2018; Hildebrandt, 'Primitives of Legal Protection in the Era of Data-Driven Platforms' (n 22).,

[45] Viktor Mayer-Schönberger and Yann Padova, 'Regime Change: Enabling Big Data through Europe's New Data Protection Regulation' (2015) 17 Colum. Sci. & Tech. L. Rev. 315, 332 favours moving away from consent for data uses to governance of fair and ethical uses of data; similar see Alessandro Mantelero, 'The Future of Consumer Data Protection in the EU Re-Thinking the "Notice and Consent" Paradigm in the New Era of Predictive Analytics' (2014) 30 Computer Law & Security Review 643.

[46] Brent Daniel Mittelstadt and Luciano Floridi, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2016) 22 Science and Engineering Ethics 303.

[47] Paul Ohm, 'The Fourth Amendment in a World without Privacy' (2011) 81 Miss. LJ 1309; Pauline T Kim, 'Data-Driven Discrimination at Work' 58 81.

[48] Mittelstadt (n 27); Luciano Floridi, 'The Informational Nature of Personal Identity' (2011) 21 Minds and Machines 549.

[49] Sandra Wachter, 'Privacy: Primus Inter Pares — Privacy as a Precondition for Self-Development, Personal Fulfilment and the Free Enjoyment of Fundamental Human Rights'

determination.[50] As Tene and Polonetsky argue, "In 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."[51] The Article 29 Working Party has recognised a similar challenge, arguing that "More 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."[52] The European Data Protection Supervisor (EDPS) has likewise expressed concern over the privacy risks of inferences and the need for governance.[53]

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), the usage 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,[54] and self-determination of individuals.[55] The ECHR[56] has a long-

(Social Science Research Network 2017) SSRN Scholarly Paper ID 2903514 <https://papers.ssrn.com/abstract=2903514> accessed 27 December 2017.

[50] Urteil des Ersten Senats vom BVerfG, '15. Dezember 1983, 1 BvR 209/83: Volkszählungsurteil' URL: http://dejure. org/dienste/vernetzung/rechtsprechung. 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' [2006] Human Rights Quarterly 186.

[51] Omer Tene and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2012) 11 Nw. J. Tech. & Intell. Prop. xxvii, 270.

[52] Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation, 00569/13/EN WP 203, Adopted on 2 April 2013' (2013) 47 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> accessed 31 July 2018.

[53] The European Data Protection Supervisor (EDPS), 'Opinion 3/2018 on Online Manipulation and Personal Data' 3/2018 8–16 <https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf> accessed 30 July 2018.

[54] For an in-depth discussion on identity and profiling, see: Mireille Hildebrandt and Serge Gutwirth, *Profiling the European Citizen* (Springer 2008); Antoinette Rouvroy, 'Privacy, Data Protection, and the Unprecedented Challenges of Ambient Inteligence' (Social Science Research Network 2007) SSRN Scholarly Paper ID 1013984 <https://papers.ssrn.com/abstract=1013984> accessed 27 December 2017.

[55] Wachter, 'Privacy' (n 49); Nora Ni Loideain, 'Surveillance of Communications Data and Article 8 of the European Convention on Human Rights', *Reloading Data Protection* (Springer 2014).

[56] 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' (2017) T-PD(2017)23 <https://rm.coe.int/case-law-on-data-protection/1680766992> accessed 30 July 2018.

standing tradition of linking the right to personality to the right of privacy.[57] This link suggests that, to remain in control of their identity in the face of uncertainty, data subjects may alter their behaviour (e.g. self-censorship) when using digital technologies.[58] 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 Zarsky puts it – as "an autonomy trap."[59] Therefore, there is also a public and collective interest in the protection of privacy.[60]

The tendency in mature information societies[61] to create, share (or 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, behaviours and preferences that would otherwise remain private.[62] 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 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."[63] 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

[57] Wachter, 'Privacy' (n 49). 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' (2017) 33 Computer Law & Security Review 584.

[58] Jon Penney, 'Chilling Effects: Online Surveillance and Wikipedia Use' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2769645 <https://papers.ssrn.com/abstract=2769645> accessed 27 December 2017; PEN America, 'Chilling Effects: NSA Surveillance Drives US Writers to Self-Censor' [2013] New York: PEN American Center; Sauvik Das and Adam DI Kramer, 'Self-Censorship on Facebook.', *ICWSM* (2013).

[59] 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' (2002) 5 Yale JL & Tech. 1, 35.

[60] Priscilla M Regan, 'Privacy as a Common Good in the Digital World' (2002) 5 Information, Communication & Society 382.

[61] Luciano Floridi, 'Mature Information Societies—a Matter of Expectations' (2016) 29 Philosophy & Technology 1.

[62] Mittelstadt and Floridi (n 46).

[63] Mittelstadt (n 27) 482.

forever."[64] Inferential analytics thus poses substantial and novel risks not only to identity, but reputation and the choices offered to an individual by data-driven services.

While the potential harms of inferences have been recognised by European legal scholars and policy-makers, data protection law has 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 (see: Section III).[65] Rather, the default 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 (see: Sections III and IV), 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. At the moment individuals are not guaranteed to be aware 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.[66] We argue that the introduction of a right to reasonable inferences is precisely the type of reaction required.

II. 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

[64] Luciano Floridi, 'Four Challenges for a Theory of Informational Privacy' (2006) 8 Ethics and Information Technology 109, 112.

[65] Tene and Polonetsky (n 51) who argue that decison-making criteria of companies should be disclosed.

[66] Serge Gutwirth and Paul De Hert, 'Regulating Profiling in a Democratic Constitutional State', *Profiling the European citizen* (Springer 2008); similar on the need for transparent decsion-making processes RE Leenes, M Hildebrandt and S Gutwirth, 'Addressing the Obscurity of Data Clouds' [2008] Profiling the European citizen 341, 298.

data. Here, we define 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 we are interested in 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, and have low verifiability in the sense of being predictive or opinion-based (see: Sections I.A and V). Several distinctions between 'types' of personal data relevant to the legal status of inferences can be seen in the GDPR itself as well as guidance issued by the Article 29 Working Party. Art 4 GDPR defines personal data as "any information relating to an identified or identifiable natural person." Art 9(1) 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." Sensitive personal data' incurs additional restrictions on processing (Art 9(2) to Art 9(4)). 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.[67] Provided data includes any data that the data subject has directly provided to the data controller, for example name or email address. Observed data is also "provided by" the data subject, but indirectly or passively, including things such as location data, clicking activity, or unique aspects of a person's behaviour such as handwriting, keystrokes, or a particular way of walking or speaking.[68] 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 personalisation 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

[67] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 8; Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (2017) 16/EN WP 242 rev.01 10 <https://ec.europa.eu/newsroom/document.cfm?doc_id=44099> accessed 10 October 2017.

[68] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (20 June 2007) 8 <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf> accessed 1 April 2018.

some cases, other background data.[69] The Article 29 Working Party's guidelines on data portability[70] provide examples of personal data derived from untraditional sources, such as data produced "from the observation of his activity," for example clicking or browsing behaviour and the inferences drawn from it.[71] 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."[72] Clearly, if inferences can be considered personal data, they are of the latter type: derived or inferred.

A. Three-step model

To determine whether data is 'personal data', the Article 29 Working Party[73] has proposed a three-step model. According to this model, the content, purpose, or result[74] of the data (processing) must relate to an identifiable person either directly or indirectly.[75] This approach allows for non-personal data to be transformed into personal data through linkage to an identified individual.[76] For example, the value of a house can become personal data

[69] Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (2016) 16/EN WP 242 10–11 <http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf> accessed 10 October 2017.

[70] Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (n 67) 9, 10, Fn. 21.

[71] Even though inferences are not covered by Art 20, but rather by Art 15. See: ibid Fn. 20.

[72] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 9; Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (n 67) 9–10. which refers to "observed data" such as "activity logs, history of website usage or search activities."

[73] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68); 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' (2018) 10 Law, Innovation and Technology <https://papers.ssrn.com/abstract=3036355> accessed 2 May 2018.

[74] Results here means "to evaluate, treat in a certain way or influence the status or behaviour of an individual." Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 10.

[75] ibid 11.

[76] For an excellent overview of the concept of personal data, see: Douwe Korff, 'New Challenges to Data Protection Study - Working Paper No. 2: Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments' (European Commission DG Justice, Freedom and Security 2010) <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1638949> accessed 8 December 2016.

used to assess an individual, such as the amount of tax to be paid.[77] Due to technical affordances, some commentators have argued that it is difficult to locate data that cannot potentially be transformed into personal data.[78]

The third step of the model, 'result', is key to the legal status of inferences.[79] The Article 29 Working Party argues that data being "likely to have an impact on a certain person's rights and interests"[80] 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"[81] 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'). 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 the GDPR's definition (Art 4(1)), which the Article 29 Working Party takes as evidence of legislators' intent to have a very wide definition of 'personal data'. They argue that personal data includes 'subjective' "information, opinions, or assessments"[82] 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."[83] 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,"[84] for example by providing additional comments.[85] 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.[86]

---

[77] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 9.

[78] Stefan Ernst, 'DS-GVO Art. 4 Begriffsbestimmungen' in Boris Paal and Daniel A Pauly (eds), *Datenschutz-Grundverordnung* (1st edn, beck-online 2018) Rn. 8-13.

[79] See: Korff, 'New Challenges to Data Protection Study-Working Paper No. 2' (n 76) 52–53 who argues that profiles, understood as a bundle of interences and assumptions, should be treated as pesonal data.

[80] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 11.

[81] ibid 10.

[82] ibid 6.

[83] ibid.

[84] ibid.

[85] ibid Footnote 5.

[86] For example, see: Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 17–18. which clarifies that the rights to rectification, erasure, and restriction of processing

B. Subjectivity and verifiability

Inferences are often precisely this type of subjective and non-verifiable "information, opinions, or assessments"[87] 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")."[88] Such subjective third party assessments can be considered a type of inference, as the assessment involves inferring a non-observed characteristic or subjective opinion of the subject from data already held[89].

Concerning non-verifiability, a second example is provided of a child's drawing depicting her family and her mood towards them. Such a drawing, although created by the child, can allow for information about the behaviours of the child's parents to be inferred. The drawing itself, and any information about her parents' behaviour inferred from it, are classified as the parents' personal data as a result. Such inferences are not necessarily verifiable, and

---

apply to inferred and derived data. See also Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (n 69) 11. Here, following the text of Article 20(1) 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 and others, 'The Right to Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services' [2017] Computer Law & Security Review.

[87] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 6.

[88] ibid.

[89] For a discussion of opinions and assessments being classified as personal data under EU data protection law, see: Korff, 'New Challenges to Data Protection Study-Working Paper No. 2' (n 76).

are subjective due to interpretation being required to derive information about the parent's behaviours.[90]

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.[91] By definition, predictions cannot be verified at the time they are made, but can nonetheless describe an identified or identifiable person. Klabunde[92] similarly believes that assumptions and assessments ("Einschätzungen und Urteile") are also personal data, irrespective of whether they are accurate or verifiable.

III. JURISPRUDENCE OF THE EUROPEAN COURT OF JUSTICE

While the legally non-binding guidelines of the Article 29 Working Party clearly endorses 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 'personal data' and applicable rights.[93] Two recent cases (YS. and M. and S.[94], and Nowak[95]) 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

[90] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 8. As such, the child's parents can exercise their right of access in relation to the drawing.

[91] Ernst (n 78) Rn. 14-18.

[92] Achim Klabunde, 'DS-GVO Art. 4 Begriffsbestimmungen' in Eugen Ehmann and Martin Selmayr (eds), *Datenschutz-Grundverordnung* (1st edn, CHBeck 2017) Rn. 7-8.

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

[94] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7).

[95] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8).

YS. and M. and S. addressed whether an applicant has a right to access the legal analysis (or "information about the assessment and application") underlying a decision of legal residency. The ECJ's judgement[96] and the associated opinion of the Advocate General[97] in this case suggest a troubling direction of travel for protection of data subjects for three reasons: (1) the limited scope of personal data, (2) the limited right of access, 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."[98] This ruling indicates that only the personal data contained or used within the legal analysis, but the not the analysis itself, is personal data subject to protection under the 1995 Data Protection Directive.[99] Specifically, the ECJ noted that only the "applicant's name, date of birth, nationality, gender, ethnicity, religion and language,"[100] or only data that is 'about' the data subject are personal data.[101]

---

[96] For an in-depth analysis of the judgment, see: 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' (2015) 31 Computer Law & Security Review 112; Evelien Brouwer and 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 (C-141/12 and C-372/12)' (2015) 17 European Journal of Migration and Law 259; Purtova (n 73).

[97] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9).

[98] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) 70 (1).

[99] See: *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) footnote 38 for other cases where access to documents was sought for example, "point 24 of Advocate General Poiares Maduro's Opinion in Joined Cases C-39/05 P and C-52/05 P Sweden and Turco v Council [2008] ECR I-4723, endorsing the General Court's observation at first instance that 'the reference to "legal advice" [in Article 4(2) of Regulation No 1049/2001] does not raise any difficulty of interpretation'. Other cases where access was sought to legal opinions of the legal services of the EU institutions or to legal documents submitted to the ECJ include, for example, Joined Cases C-514/07 P, C-528/07 P and C-523/07 P Sweden and Others v API and Commission [2010] ECR I-8533.".

[100] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) 38.

[101] Purtova (n 73) 28.

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.[102] Examples of personal data named in prior judgements include "telephone number, and information about his/her working conditions or hobbies,"[103] "surname and given name of certain natural persons whose income exceeds certain thresholds," as well as "their earned and unearned income,"[104] "static IP addresses,"[105] "dynamic IP addresses,"[106] "fingerprints,"[107] "record of working time and rest periods,"[108] "data collected by private detectives,"[109] "image of a person recorded by a camera,"[110] "tax data,"[111] and "press releases."[112]

In contrast, in YS. and M. and S. the ECJ addresses whether legal analysis can be considered personal data. This determination is very relevant for the legal status of inferences. A legal analysis is comparable to analysis of personal data from which new data is derived or inferred. A legal analysis can include interim inferences (i.e. assessment of how the law applies to a case) on which final inferences (i.e. not meeting the required standards for

[102] Of course it has to be borne in mind that the Court can only rule on the cases referred to it, thus the Court has no power to take views that fall outside the cases it considers.

[103] *Criminal proceedings against Bodil Lindqvist C-101/01* (Judgment of the Court of Justice).

[104] *Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy and Satamedia Oy, Case C-73/07* (Judgment of the Court (Grand Chamber)).

[105] *Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM), Case C-70/10* (Judgment of the Court (Third Chamber)).

[106] *Patrick Breyer v Bundesrepublik DeutschlandCase C‑582/14* (JUDGMENT OF THE COURT (Second Chamber)) stating that "all the information enabling the identification" does not need to be in the 'hands of one person'.

[107] *Michael Schwarz v Stadt Bochum, C-291/12* (Judgment of the Court (Fourth Chamber)).

[108] *Worten – Equipamentos para o Lar SA v Autoridade para as Condições de Trabalho (ACT), C-342/12* (Judgment of the Court (Third Chamber)).

[109] *Institut professionnel des agents immobiliers (IPI) v Geoffrey Englebert, C‑473/12* (JUDGMENT OF THE COURT (Third Chamber)).

[110] *František Ryneš v Úřad pro ochranu osobních údajů, C‑212/13* (JUDGMENT OF THE COURT (Fourth Chamber)).

[111] *Smaranda Bara and Others v Președintele Casei Naționale de Asigurări de Sănătate, C‑201/14* (Judgment of the Court (Third Chamber)).

[112] *Kalliopi Nikolaou, ancien membre de la Cour des comptes des Communautés européennes, demeurant à Athènes (Grèce), représentée par Mes V Christianos et V Vlassi, avocats, partie requérante, contre Commission des Communautés européennes, représentée par Mme M Condou-Durande et M C Ladenburger, en qualité d'agents, T-259/03* (JUDGMENT OF THE Court (Second Chamber)). For a general overview of the jurisprudence frim 2000-2015, see: Laudati, *Summaries of Eu Court Decisions Relating to Data Protection 2000-2015* (January 2016) <https://ec.europa.eu/anti-fraud/sites/antifraud/files/caselaw_2001_2015_en.pdf>.

residency) are built that lead to a decision (i.e. denial of residency).[113] Three issues arise here: (1) is the legal analysis (including interim inferences) itself personal data, (2) are the final inferences produced by the analytic process personal data and (3) is the decision based on it personal data? The ECJ's judgement makes clear that the first question must be answered in the negative, meaning the analysis itself is not considered personal data, and provides no answer to the second and the third questions.

In this regard the judgement followed the opinion of the Advocate General (AG).[114] 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,"[115] or "the reasoning underlying the resolution of a question of law."[116] 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,"[117] and thus a "legal analysis is not itself personal data."[118]

To unpack the distinction between facts (as personal data) and analysis, the AG uses 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,[119] she argues 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."[120] As a result, legal analysis, and more broadly "the steps of reasoning by which [a] conclusion is reached"[121] about an individual, cannot be considered personal data.

The distinction made here between describing a person as 'underweight' or 'obese' and "the steps of reasoning by which the conclusion is reached"[122] is important for answering the second question. Elsewhere in the opinion, the AG suggests that it is unhelpful "to distinguish between 'objective' facts and

---

[113] An alternative view could be that the analysis is not equivalent to interim inferences, but rather the reasoning or logic that leads to the inference. However, it is more likely that the reasoning is a cognitive process and the analysis is the output of the reasoning. Even if one wishes to argue that this is not the case, the outcome of our argument would not change as the problems remain the same. Regardless of how broadly one defines personal data, the rights granted are very limited as will be shown in Section III.A.2, III.B.2, and Section IV.

[114] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9).

[115] ibid para 54.

[116] ibid para 59.

[117] ibid para 56.

[118] ibid para 61.

[119] On objective and subjective communication of facts see: ibid para 57-58.

[120] ibid para 58.

[121] ibid.

[122] ibid.

'subjective' analysis,"[123] as "facts can be expressed in different forms, some of which will result from assessing whatever is identifiable."[124] 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."[125] 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').[126]

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 (See: Section II). 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 of 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,"[127] but leaves open how the analysis or reasoning leading to the assessment should be classified. Interestingly enough, the AG also leaves open how results of the analysis (our second question) should be classified, even though – as discussed above (see: Section III.C) – 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," 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.

---

[123] ibid 57.
[124] ibid.
[125] ibid para 57.
[126] "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" ibid para 58. "The explanation itself is not information relating to an identified or identifiable person" see ibid para 59."
[127] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) footnote 40.

Unfortunately, this is not the first time the AG and ECJ have expressed the view that a legal analysis is not personal data[128] and denied access to it.

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 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.[129]

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),[130] and the right of access in particular, is not designed to provide

---

[128] ibid para-48' This is not the first time that a question about access to legal analysis or advice has been put to the Court. See, for example, point 24 of Advocate General Poiares Maduro's Opinion in Joined Cases C-39/05 P and C-52/05 P Sweden and Turco v Council [2008] ECR I-4723, endorsing the General Court's observation at first instance that "the reference to 'legal advice' [in Article 4(2) of Regulation No 1049/2001] does not raise any difficulty of interpretation". Other cases where access was sought to legal opinions of the legal services of the EU institutions or to legal documents submitted to the Court include, for example, Joined Cases C-514/07 P, C-528/07 P and C-523/07 P Sweden and Others v API and Commission [2010] ECR I-8533. See also points 13 and 14 above. In those cases, however, it would appear that access was sought on other bases. In particular, Article 4(2), second indent, of Regulation No 1049/2001. See point 14 above. The Court was not required to examine whether and why a document containing legal analysis or advice is different from one having a different content.'

[129] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 45-46.

[130] *Commission v Bavarian Lager - Case C‑28/08 P* (n 6).

access to or facilitate assessments of the accuracy for decision-making processes. In 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.[131]

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

> 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.[132]

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."

These particular limits on the right of access are not one-off. In 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, or object to processing.[133] 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. Rather, as the ECJ argued 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."[134]

---

[131] ibid para 49.
[132] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 47.
[133] *College van burgemeester en wethouders van Rotterdam v M. E. E. Rijkeboer C-553/07* (n 11) para 51-52.
[134] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) Para 70(2).

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 privacy,"[135] 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."[136] She does admit that data subjects have a valid interest in "knowing exactly what circumstances were relevant to the decision taken,"[137] 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.[138] Instead, review of "the decision for which…legal analysis was prepared"[139] should be left to a relevant "independent judicial authority."[140] 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 (see Section I). 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,[141] unless relevant sectoral laws allow them to do so.

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

---

[135] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) para 60.
[136] ibid.
[137] ibid para 36.
[138] ibid para 32.
[139] ibid para 60.
[140] ibid.
[141] See also Douwe Korff who has expressed similar concerns in the past Douwe Korff, 'EU Law Analysis: The Proposed General Data Protection Regulation: Suggested Amendments to the Definition of Personal Data' (*EU Law Analysis*, 15 October 2014) <http://eulawanalysis.blogspot.com/2014/10/the-proposed-general-data-protection.html> accessed 30 July 2018.; Steve Peers, 'EU Law Analysis: Data Protection Rights and Administrative Proceedings' (*EU Law Analysis*, 17 July 2014) <http://eulawanalysis.blogspot.com/2014/07/data-protection-rights-and.html> accessed 30 July 2018. Robert Madge, 'Five Loopholes in the GDPR' (*MyData*, 27 August 2017) <https://medium.com/mydata/five-loopholes-in-the-gdpr-367443c4248b> accessed 30 March 2018.

sector based on democratic legitimacy,[142] 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 on Nowak[143] 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.[144] As with YS. and M. and S., the case centred 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."[145] Specifically, the Court determined that an opinion or assessment that is "linked to a particular person" by "reason of its content, purpose or

[142] Paul De Hert and Serge Gutwirth, 'Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power' [2006] Privacy and the criminal law 61.

[143] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8); see also Purtova (n 73).

[144] *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (n 10) para 13.

[145] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 34.

effect"[146] counts as personal data. 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.[147] 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. It is worth noting, however, that exam questions were not considered the candidate's personal data.[148]

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 scored for certain parts of an examination. That marking merely summarises the examination performance, which is recorded in detail in the examination script itself.[149]

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.[150] Further, both the ECJ and AG argued that the fact that "unintended" or "undesired" 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.[151] 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.[152] In other words, the scope of data protection rights

---

[146] ibid para 35.
[147] ibid para 42 and 44.
[148] ibid para 58.
[149] *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (n 10) para 27.
[150] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 44.
[151] ibid para 46.
[152] ibid para 51-56.

must be interpreted contextually, or with reference to the specific purposes for which data was collected, and the broader aims of data protection law.

For an exam script, the rights of access and rectification should not result in the candidate being allowed to correct answers a posteriori.[153] 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.[154]

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.[155]

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 fact would be nonsensical, as the purpose for which the data was collected was to evaluate the candidate's performance. 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 [..] the script of another examination candidate had been ascribed to

[153] ibid para 51-52.

[154] ibid para 54.

[155] Note ibid para 56. This reads: "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)," which could give the impression that the assessment also falls under the right of rectification. However, considering the examples provided for a sensible use of rectification (see also para 45 of the Advocate General's opinion for Nowak), 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.

the data subject, which could be shown by means of, inter alia, the handwriting, or if parts of the script had been lost."[156]

While the AG acknowledged that assessments (i.e. the assessor's comments) can be personal data,[157] 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."[158] 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,"[159] as "It is precisely that opinion that the comments are meant to record."[160] Rectification would therefore be inappropriate, as "such comments would not be wrong or in need of correction even if the evaluation recorded in them were not objectively justified."[161] 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."[162] 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. 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.[163]

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.[164] 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 comments after an appropriate

---

[156] *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (n 10) para 35-36.

[157] 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: ibid 58–59.

[158] ibid para 54.

[159] ibid.

[160] ibid.

[161] ibid.

[162] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 56.

[163] ibid para 60-61.

[164] *Commission v Bavarian Lager - Case C‑28/08 P* (n 6) para 49. as well as YS. and M. and S and *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) 45–47.

period of time.[165] 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."[166]

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.[167] Rather other applicable laws and remedies need to be consulted, for example through examination procedures.[168] 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.[169]

C. Lessons from jurisprudence of the ECJ

These two cases 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.[170] The AG even went so far as to argue that only (verifiable) facts constitute personal data.[171] 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.[172]

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, for example the

---

[165] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 55.
[166] ibid para 50.
[167] ibid para 51-53, 54-55, 58.
[168] ibid para 54-55.
[169] *College van burgemeester en wethouders van Rotterdam v M. E. E. Rijkeboer C-553/07* (n 11); *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 57.
[170] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 38-39.
[171] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) para 56.
[172] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 34.

decision not to grant residency or to fail someone at an exam, is not personal data. This seems to be especially true since the ECJ in Nowak[173] uses the same terminology as the Article 29 Working Party's' three-step definition of personal data, which includes the output data.[174]

Despite seemingly 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.[175] So, for example, the right of access might conflict with the right to privacy of the assessor,[176] 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 cannot be corrected.[177] The same holds true for the comments and assessment of the examiner.[178]

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 Art 15-18 and Art 21 of the GDPR.[179] Specifically, the Working Party appears to fully extend certain individual rights of the GDPR to derived and inferred data, including non-verifiable predictions.[180] This much is explicitly stated in relation to the right to rectification (Art 16), which is said to apply "to the 'input personal data' (the personal data used to create the profile) and to the 'output data' (the profile itself or 'score' assigned to the person, which is personal data relating to the person concerned)." The rights of access (Art 15), erasure (Art 17), restriction of processing (Art 18), and to object to processing (Art 21) are also said to apply. Art 18 is explicitly said to apply to any stage of the profiling process.[181]

The Article 29 Working Party's position appears to assume that data protection law aims to ensure accurate decision-making, which would allow

---

[173] ibid para 35, 42, 44.

[174] This was even recognised by the AG in YS. and M. and S by referring to the view of the Article 29 Working Party that the results of a medical analysis (regardless of verifiability) are personal data. See: *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) Footnote 40.

[175] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 53.

[176] ibid para 44, 59 and 61.

[177] ibid para 51-53.

[178] ibid para 54-55.

[179] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 17–18.

[180] ibid 18. An example is used of a profiling containing the prediction that a patient will suffer from heart disease.

[181] ibid 17–18.

inferences to be fully accessed, corrected, and erased (for example, if thought to be irrelevant). However, this view runs against the ECJ in Bavarian Lager,[182] YS. and M. and S.[183] and Nowak[184], and the AG in YS. and M. and S.[185] and Nowak[186] (see: Section III). Rulings and opinions in these cases clarify that the remit of data protection law is not to assess the accuracy of the reasoning behind decision 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,[187] YS. and M. and S.[188] and the AG in YS. and M. and S.[189] and Nowak[190] made it very clear that data protection law does not guarantee lawful decision-making (e.g. a right to good administration or correct marking).[191] The ECJ in Nowak did not disagree even though reference was made to all these views. The limited way to which the right to rectification applies to the comments of the assessor was even mentioned by the ECJ in its judgement.[192] Based on these considerations and the examples of rectification provided,[193] agreement is implicit. In general the ECJ in Nowak even noted how the GDPR allows broader

[182] *Commission v Bavarian Lager - Case C‑28/08 P* (n 6) para 49.
[183] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 45-47.
[184] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 54-55.
[185] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) para 32 and 60.
[186] *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (n 10) para 54-55.
[187] *Commission v Bavarian Lager - Case C‑28/08 P* (n 6) para 49.
[188] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 45-47.
[189] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) para 32 and 60.
[190] *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (n 10) para 54-55.
[191] The ECJ and AG's view 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 judgements reviewed here were made prior 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: *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) 44, 59, and 61–62.
[192] ibid para 54.
[193] ibid.

exemptions to the right of access.[194] Art 16 only aims to verify that the data undergoing processing is complete and accurate.[195]

At first glance, the broadening of the scope of personal data in Nowak compared to preceding jurisprudence seems to move toward higher protection standards for inferences.[196] However, if the rights in the GDPR (e.g. Art 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.[197] 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 to determine 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 section.

IV. 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. GDPR, Directive on the supply of digital content[198] and the ePrivacy Regulation[199]) are also relevant to determine the full legal status of inferences in data protection law. This section reviews the rights available to data subjects to manage how inferences are drawn and used

---

[194] ibid para 61.

[195] ibid para 54.

[196] See Purtova (n 73) who argues that the data protection law becomes the law of everything due to the scope of 'personal data'.

[197] For a general discussion see Hildebrandt (n 90) who is concerned that data subjects have no control over inferences.

[198] European Parliament Committee on the Internal Market and Consumer Protection and European Parliament Committee on Legal Affairs, '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' (European Parliament 2017) A8-0375/2017 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0375+0+DOC+XML+V0//EN> accessed 5 August 2018.

[199] European Commission, '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)' (European Commission 2017) COM/2017/010 final-2017/03 (COD) <https://publications.europa.eu/en/publication-detail/-/publication/b54bd5d0-d809-11e6-ad7c-01aa75ed71a1/language-en> accessed 5 August 2018. For an assessment of the proposal see Frederik Zuiderveen Borgesius and others, 'An Assessment of the Commission's Proposal on Privacy and Electronic Communications' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2982290>.

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 incur multiple transparency rights (Art 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 which the data subject has not directly provided.[200] This type of oversight is an essential prerequisite to exercise other rights granted by the GDPR. Unfortunately, the GDPR's notification duties (Art 13-14) are unlikely to fulfil this aim.

Art 13 describes numerous notification requirements for data controllers when personal data is collected 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. Given this timeline, Art 13 by definition covers only data provided by the data subject, including observed data.[201] 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, Art 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),[202] and "from which source the personal data originate, and if applicable, whether it came from publicly accessible sources" (Art 14 GDPR). In practice, a data 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 Art 14 applies, the data controller only needs to inform about the categories of data involved. "Categories of personal data" is not defined in the GDPR, but suggests that data controllers do not need to reveal details of

[200] Concerning Article 15, see: Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (n 67) 10 Fn 20.
[201] ibid 10.
[202] It 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.

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.[203] 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 (Art 14(5)a).[204] 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" (Art 14(5)b). 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 (Recital 62). Each of these gaps indicate that the data subject will not necessarily be informed when and what kind of inferred or derived data have been obtained from a third party.

Art 13-14 leave open one final very significant loophole that can result in the data subject 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 (Art 13) or a third party (Art 14). Controllers can thus avoid notification duties by drawing inferences themselves.

The right of access (Art 15) 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[205] and guidelines on profiling[206], the right of access (Art 15) applies to inferred and derived data, including profiles built from such data by the data controller. Art 15 allows the data subject to request – at any time – information about the purposes of the processing,[207] categories of personal data held, recipients or categories of recipients, and the source of the data obtained.

---

[203] Rainer Knyrim, 'DS-GVO Art. 14 Informationspflicht bei Erhebung von Daten' in Eugen Ehmann and Martin Selmayr (eds), *Datenschutz-Grundverordnung* (1st edn, CHBeck 2017) Rn. 26-27.

[204] ibid Rn. 6-7.

[205] Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (n 69) 10 footnote 20.

[206] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 17.

[207] 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.

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 undergoing processing (Art 15(3)). 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 the jurisprudence of the European Court of Justice (see: Sections II and III).

While promising, Art 15(3) is not an absolute right. Obtaining such a copy must "not adversely affect the rights and freedoms of others," which according to Recital 63 includes "trade secrets or intellectual property and in particular the copyright protecting the software" (see Section VI). 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. The view of the limited scope of Art 15 is supported – even if not related to trade secrets - by the ECJ's judgement in YS. and M. and S, which confirm that only a summary of personal data undergoing processing needs to be provided.[208] 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 has confirmed as much in the Nowak ruling, stating that the GDPR has more generous clauses to restrict the right of access using Art 15(4) and Art 23[209] to protect the privacy of others (in Nowak, the examiner)[210] and other public interests.

The ECJ has thus revealed through these judgements that the right of access, particularly when addressing inferred and derived data, requires a balance of the interests of the data 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 cannot 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 that "online marketplaces will have to inform the consumers about the main parameters determining the ranking of the results."[211] Such disclosures may need to include information

[208] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 70(2).

[209] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 59 and 61.

[210] ibid para 44.

[211] European Commission, 'Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee - A New Deal for Consumers' (2018) COM(2018) 183 final para 2 <https://eur-lex.europa.eu/legal-

about inferences drawn about the user that underlie the rankings. While promising, it is still very early days 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 (Art 16), erasure (Art 17), objection to processing (Art 21), and contesting automated decision-making including profiling (Art22(3)). 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.

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" Rectification implicitly relies upon the notion of verification, meaning 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 drawn has a factual basis or in other words is 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 (see: Section II.B). 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. It has been argued that only data that can be verified count as personal data and thus fall within the

---

content/EN/TXT/?uri=COM%3A2018%3A183%3AFIN> accessed 31 July 2018; for a discussion of the relationship between consumer and data protection rights, see: Natali Helberger, Frederik J Zuiderveen Borgesius and Agustin Reyna, 'The Perfect Match? A Closer Look at the Relationship between EU Consumer Law and Data Protection Law' (2017) 54 Common Market Law Review.

scope of the right to rectification, excluding unverifiable inferred data.[212] In contrast, Kamann and Braun suggest that 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.[213]

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.[214] 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.[215] 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."[216]

The European Court of Justice has similarly (but not consistently) argued that opinions and assessments can constitute personal data (see: Section III.B). 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 rectified under data protection law, and could only be contested if there is a procedure in place to contest the evaluation.[217]

---

[212] "This, however, is a restrictive view on such data: they can constitute a complex list of information units (number 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." Quoted from Gianclaudio Malgieri, 'Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data' (2016) 4 Privacy in Germany - PinG 133 ff, 144.

[213] Hans-Georg Kamann and Martin Braun, 'DS-GVO Art. 16 Recht auf Berichtigung' in Eugen Ehmann and Martin Selmayr (eds), *Datenschutz-Grundverordnung* (1st edn, CHBeck 2017) Rn. 20-21.

[214] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 6; Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 18.

[215] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 18.

[216] Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation, 00569/13/EN WP 203, Adopted on 2 April 2013' (n 52) 47.

[217] "Any objections to the comments would consequently have to be dealt with as part of a challenge to the evaluation of the script." Quoted from: *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (n 10) para 55.

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.[218] According to Art 17(1) 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;[219] or (3) an objection to processing is entered that is not trumped by compelling legitimate grounds of the data controller.[220]

Concerning point (2), from the controller's perspective, one potential source is found in Article 16 of the EU Charter of Fundamental Rights: the freedom to conduct a business.[221] The GDPR does not prescribe a specific balance between data subject's right to erasure and the legitimate interest of controllers. The Article 29 Working Party in relation to the DPD 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)"[222] as areas where the data subject's interests may not prevail.

Concerning point (3), Art 21 grants data subjects the right to object to or stop data processing if the processing is based on Art 6(1)e (public interest or official authority) or 6(1)f (legitimate interests), which includes inferring or deriving new data from existing records. In the case of profiling for direct marketing purposes (Art 21(2)), an objection is guaranteed to be successful, meaning new inferences cannot be generated, unless data controllers can claim an alternative legitimate basis that is not direct marketing. Any other

---

[218] For a discussion why the right to be forgotten is essential in the connected world, see: Mayer-Schönberger (n 15).

[219] Article 29 Data Protection 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, Adopted on 9 April 2014' (2014).

[220] On the challenges of implementing the right to be forgotten for AI systems, see: Eduard Fosch Villaronga, Peter Kieseberg and Tiffany Li, 'Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten' (2018) 34 Computer Law & Security Review 304.

[221] Norbert Nolte and Christoph Werkmeister, 'DS-GVO Art. 17 Recht auf Löschung ("Recht auf Vergessenwerden")' in Peter Gola (ed), *Datenschutz-Grundverordnung VO (EU) 2016/679* (1st edn, CHBeck 2017) Rn. 47-48.

[222] See Article 29 Data Protection 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, Adopted on 9 April 2014' (n 219) 25 where some of the most common legitimate interests are listed.

purpose than direct marketing[223] must be weighed against the "compelling legitimate grounds" of the data controller.[224] Again, it remains unclear what such "compelling legitimate grounds" would look like.[225] However, if it is determined that the data subject has a stronger interest that allows processing to be stopped, Art 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'. Requests for deletion will thus only be successful if the data controller no longer requires the data.[226]

Concerning inferences specifically, some commentators have cast doubt on the applicability of the right to erasure to inferences. Some scholars appear to suggest that Art 17 will not apply to inferences altogether,[227] while others

[223] According to 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 (Art 4). The question remains whether personalised 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 behavioural advertisements as the current draft seems too narrow, see: Article 29 Data Protection Working Party, 'Opinion 01/2017 on the Proposed Regulation for the EPrivacy Regulation (2002/58/EC),17/EN WP 247, Adopted on 4 April 2017' (2014) 21.

[224] Very often data controllers use consent for data processing as lawfulness is easier to prove using Art 7. However, after withdrawing consent, the controller can continue processing if the same purpose is also covered under Art 6(1)(f) GDPR (legitimate interest). See also Nolte and Werkmeister (n 221) Rn. 13-15.

[225] Pointing at this loophole, see: Wachter, 'Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR' (n 13); Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 18 states that 'compelling legitimate grounds' are not defined.

[226] Nolte and Werkmeister (n 221) Rn. 17-18.

[227] Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a 'Right to Explanation'Is Probably Not the Remedy You Are Looking For' [2017] Duke Law & Technology Review 13, 68–69. Edwards and Veale believe that Art 17 is not designed to cover observed data, but do not offer support. They also assume a complementary relationship between Art 17 and Art 20, which implies that inferred data would not be covered under Art 17 (as is the case with Art 20). The existence of such a relationship is however highly doubtful as Art 20 states: 'The exercise of the right referred to in paragraph 1 of this Article shall be without prejudice to Article 17.' .

argue that the financial expenditure of a data controller to create inferences will trump the data subject's request for deletion.[228]

These positions stand 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,[229] and potentially even results[230] (i.e. provided and inferred data as well as the reasons for the inferences), although the right must be counterbalanced against other laws[231] (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 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.[232]

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. Art 19 requires disclosure "to each recipient to whom the personal data have been disclosed, unless this proves impossible or involves disproportionate effort,"[233] 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 (Art 17(2)). The latter case is unlikely to apply to inferences, as profiles and inferences are not routinely made public. However, in both cases (Art 19 and Art 17(2)), even if inferences are deleted by one data controller,

---

[228] "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." Gianclaudio Malgieri, 'Trade Secrets v Personal Data: A Possible Solution for Balancing Rights' (2016) 6 International Data Privacy Law 102, 115.

[229] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 55.

[230] 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 to not prohibit this. See: ibid 55.

[231] ibid para 55, 60.

[232] ibid para 60. Regarding the view that privacy and business interests must be balanced, see: *Postbank NV v Commission of the European Communities Case T-353/94* (Judgment of the Court of First Instance) para 87; and *Bank Austria Creditanstalt AG v Commission of the European Communities Case T-198/03* (Judgment of the Court of First Instance (Second Chamber)) para 71.

[233] If the data subject requests it, data controllers must disclose with whom the data was shared with, provided that Art 17 was successfully levied.

these deleted inferences can 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 (Art 13-14) and right of access (Art 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 (see Section IV.A).

Finally, even though "disproportionate effort" cannot be invoked by data controllers to deny a deletion request (Art 17),[234] Art 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." 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.[235] The framework states that "listening, tapping, storing, monitoring, scanning or other kinds of interception" (including monitoring of browsing behaviours; Recital 15) shall not be allowed unless explicitly permitted under the regulation (Art 5 EPR). Consent to processing of content data (Art 6(3)a-b) EPR) or metadata (Art 6(2)c) EPR) for one or more specified purposes is valid under this regime following the requirements for consent under Art 7 GDPR (Art 9(3) EPR).[236]

According to Art 7 of the EPR, metadata and communication data must be erased or anonymised after the "receipt of electronic communication content," or if the metadata "is no longer needed for the purpose of the transmission of a communication." 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, Art 7 EPR only refers to provided data (or content data, e.g. text, voice, videos, images, and sound)

---

[234] Nolte and Werkmeister (n 221) Rn. 31-33.

[235] European Commission, '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)' (n 199); For commentary on the current draft of the ePrivacy Regulation, see: Frederik Zuiderveen Borgesius and Wilfred Steenbruggen, 'The Right to Communications Confidentiality in Europe: Protecting Trust, Privacy, and Freedom of Expression' (Social Science Research Network 2018) SSRN Scholarly Paper ID 3152014 <https://papers.ssrn.com/abstract=3152014> accessed 13 September 2018.

[236] The consent requirement covering both content data and metadata is also coupled with the "necessity requirement." This means that the evaluation of content and metadata is only valid if necessary for the purpose and cannot be fulfilled using anonymous data.

and observed data (metadata), but not derived or inferred data. It will thus be unlikely to provide alternative means for data subjects to delete inferences drawn about them.

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)[237] 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,"[238] excluding healthcare, gambling, and financial services. Art 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" (Art 3 DCD).[239] 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."[240]

The interesting segment of the framework concerns actions to be taken after a contract is terminated. Following termination of a long-term contract (Art 16(4)b DCD), or due to a lack of conformity with the contract (Art 13(2)c 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,[241]

---

[237] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content' (European Commission 2015) COM/2015/0634 final-2015/0287 (COD) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2015:0634:FIN>; For an overview of the drafting process and the current views of the European Parliament and Council, see: European Parliament, 'Briefing Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content, EU Legislation in Progress October 2017' (2017) <http://www.europarl.europa.eu/RegData/etudes/BRIE/2017/608748/EPRS_BRI%282017%29608748_EN.pdf> accessed 31 July 2018.

[238] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content' (n 237) Art 2.

[239] ibid Art 3.

[240] European Commission, 'Press Release - A New Deal for Consumers: Commission Strengthens EU Consumer Rights and Enforcement' (2018) 1 <http://europa.eu/rapid/press-release_IP-18-3041_en.pdf>.

[241] "…with the exception of the content which has been generated jointly by the consumer and others who continue to make use of the content." European Commission, 'Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content' (n 237) Art 13(2)b.

and to render it anonymous (Art 13 DCD).[242] This right covers deletion of user-generated data as described in Recital 15, including "digital images, video and audio files, blogs, discussion forums, text-based collaboration formats, posts, chats, tweets, logs, podcasting, content created on mobile devices, content created in the context of online virtual environments, ratings and collections of links referring to online content."

Concerning 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 with legal and contractual requirements, including for example geolocation data for mobile applications, tracking cookies, and automatically generated data (e.g. IP addresses). 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).[243] 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.[244] As a result, users that have 'paid' for content or a 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 a '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 (Art 9 GDPR). 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 balance to be struck between the interests of data

[242] On the potential impossibility of anonymising data, see: Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006> accessed 28 June 2017.

[243] 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, 'Briefing Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content, EU Legislation in Progress October 2017' (n 237) 11–12.

[244] For a favourable view, see also: Inge Graef, Martin Husovec and Nadezhda Purtova, 'Data Portability and Data Control: Lessons for an Emerging Concept in EU Law' (Social Science Research Network 2017) SSRN Scholarly Paper ID 3071875 23 <https://papers.ssrn.com/abstract=3071875> accessed 11 September 2018; Gianclaudio Malgieri, ''User-Provided Personal Content'in the EU: Digital Currency between Data Protection and Intellectual Property' (2018) 32 International Review of Law, Computers & Technology 118.

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 consent and objecting to processing if consent is withdrawn. When a data subject withdraws consent, the data controller is required to cease processing unless an alternative lawful basis can be established. The existence of overriding 'legitimate interests' of the controller is one such basis. However, this is not the case for sensitive data processing based upon explicit consent (which is very common).[245] Unlike non-sensitive data, if explicit consent is withdrawn, the 'legitimate interests' of the data controller cannot serve as a lawful basis for further processing. Other potential lawful bases for processing sensitive data or drawing sensitive inferences of course remain, but compared to non-sensitive data, one less route is available to controllers seeking to continue processing against the data subject's wishes.

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. Art 9 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." It is important to note that gender, age, information about a person's financial situation, and personal profiles are not considered sensitive data under Art 9, despite often serving as grounds for discrimination.[246] 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."

Concerns about inferences are implicit in the definition of 'special categories of personal data'. The phrase "personal data revealing" suggests

[245] See Art 9 GDPR, which describes further exceptions. Most are coupled with some kind of public interest, or require that the data was made public by the data subject.

[246] The sensitive nature 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 Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' (2011) Ares(2011)444105-20/04/2011 10 <http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf> accessed 1 October 2017.

that the definition is intended to cover data that both directly discloses and indirectly reveals protected attributes.[247] 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 from which sensitive information with regard to an individual can be concluded."[248] 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 other data which is not special category data in its own right but becomes so when combined with other data [italics added]."[249] 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.

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 age range 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 our purposes. Second, when personal data can be shown to allow for sensitive attributes to be inferred (i.e. 'indirectly revealed'), the source data from which

---

[247] Sebastian Schulz, 'DS-GVO Art. 9 Verarbeitung besonderer Kategorien personenbezogener Daten' in Peter Gola (ed), *Datenschutz-Grundverordnung VO (EU) 2016/679* (1st edn, CHBeck 2017) Rn. 11-12, who refers to: Article 29 Data Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' (2011) Ares(2011)444105-20/04/2011 6 <http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf> accessed 22 October 2017, 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 and Veale (n 227) 37 who are unsure 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 Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' (n 246); Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68); Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19).

[248] Article 29 Data Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' (n 246) 6.

[249] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 15.

sensitive inferences can be drawn can also be treated as sensitive data (e.g. last name or location of birth to infer race).

This fluidity of the categorisation 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. 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.[250]

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 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. Scholars believe that 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.[251] Schulz gives the example of a pizzeria delivering to customers in a drug abuse centre. Transaction records would not be considered sensitive data unless the pizzeria intended to infer information about the health status of their customers.[252] Schiff argues similarly 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.[253] Nguyen goes so far as to argue that

[250] Tal Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 Seton Hall Law Review 1013 <https://papers.ssrn.com/abstract=3022646> accessed 26 February 2018; Mayer-Schönberger and Cukier (n 25); Mayer-Schönberger and Ramge (n 16).

[251] Alexander Nguyen, 'Videoüberwachung Insensitiven Bereichen' (2011) 35 Datenschutz und Datensicherheit 715; Alexander Schiff, 'DS-GVO Art. 9 Besonderer Kategorien personenbezogener Daten' in Eugen Ehmann and Martin Selmayr (eds), *Datenschutz-Grundverordnung* (1st edn, CHBeck 2017) Rn. 20-21.

[252] Schulz (n 247) Rn. 11-14.

[253] Schiff (n 251) Rn. 14-15.

sensitive attributes coincidentally revealed by non-sensitive data do not require a re-classification of the source data as sensitive. An example is given of a CCTV image depicting a person wearing religious attire, which was not captured to assess the individual's religious beliefs.[254] The same holds true for photos that reveal disabilities or wedding photos at the church from which religion or sexual orientation can be inferred.

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.[255] 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.[256]

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.[257] Schulz provides two examples: attendance records at union events and online browsing behaviour 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.[258] The ECJ has similarly affirmed that data must reliably reveal sensitive information to be considered 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 ECJ used the example of knowing that an individual works as an assistant to a MEP. This relationship was not taken as sufficient to infer the individual's political beliefs, suggesting that reliability is an essential attribute of 'sensitive data'.[259]

---

[254] Nguyen (n 251) even though this does not refer to 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 (n 247) Rn. 11-12.

[255] Article 29 Data Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' (n 246) 8.

[256] Douwe Korff, 'Comparative Study on Different Approaches to New Privacy Challenges, in Particular in the Light of Technological Developments-Contract No' (European Commission DG Justice, Freedom and Security 2010) JLS/2008/C4/011–30-CE-0219363/00-28, Working paper No. 2: Data protection laws in the EU: The difficulties in meeting the challenges posed by global social and technical developments 41.

[257] Nguyen (n 251); Schulz (n 247) Rn. 13-14.

[258] Schulz (n 247) Rn. 12-13.

[259] *Kathleen Egan and Margaret Hackett v European ParliamentCase T‑190/10* (JUDGMENT OF THE GENERAL COURT (Fifth Chamber)).

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.[260]

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 on 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 constrain its potential impact on the data subject's life. In effect, if this approach was adopted, the burden of data protection would shift to the data subject to object to further processing of inaccurate inferences, or to rectify or delete them. 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 invoke legitimate interests as a basis for further processing.

To summarise, 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,[261] both,[262] or perhaps neither[263] condition must be met to re-classify non-sensitive source data as sensitive data capable of revealing sensitive information.

---

[260] This is one reason why Germany's data protection law (§ 28b Nr. 3 BDSG) prohibits credit scores solely based on postcodes or addresses. See: Philipp Richter, 'Big Data, Statistik Und Die Datenschutz-Grundverordnung' (2016) 40 Datenschutz und Datensicherheit-DuD 581, 583.

[261] Schiff (n 251) Rn. 26-27 argues that intention is required, but reliability is not. *Kathleen Egan and Margaret Hackett v European ParliamentCase T‑190/10* (n 259). Here, the ECJ affirms the necessity of reliability, but does not address intentionality.

[262] Schulz (n 247) Rn. 11-14; Nguyen (n 251).

[263] Article 29 Data Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' (n 246).

E. The right to contest decisions based on inferences

Although there is no consensus about the legal rights over inferences (see Section IV), there is an argument to be made that even the GDPR goes beyond data control and management (informational self-determination), and provides safeguards against inferences and decisions based on inferences with the right to contest in Art 22(3).[264] The following section will call into question that the right to contest can be meaningfully implemented without underlying decision-making standards.

Art 22(3) GDPR describes safeguards against decisions based solely on automated processing, including profiling, that produce legal or similarly significant effects for data subjects (Art 22(1)). 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 data control and management (or informational self-determination, e.g. Art 15) to allow data subjects to evaluate and challenge automated decisions and profiling which 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,[265] the right to explanation in Recital 71 similarly recognises data subjects' interests in how they are evaluated. This recognition of valid interests regarding the output of data processing distinguishes Art 22 from the majority of other 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 cannot be contested, or where relevant legal or ethical decision-making standards do not exist. But as shown in Section I, 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."[266]

---

[264] Isak Mendoza and Lee A Bygrave, 'The Right Not to Be Subject to Automated Decisions Based on Profiling' in Tatiani Synodinou and others (eds), *EU Internet Law: Regulation and Enforcement* (Springer 2017) <https://papers.ssrn.com/abstract=2964855> accessed 10 May 2017.

[265] Wachter, Mittelstadt and Floridi, 'Why There Is No Right to Explanation in the General Data Protection Regulation' (n 16).

[266] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 10.

At first glance the right to contest appears to strengthen the protection afforded to data subjects against all types of 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 (the ECJ in Bavarian Lager[267], YS. and M. and S[268] and Nowak[269]; the AG in YS. and M. and S[270] and Nowak[271]). The ECJ has argued that the remit of data protection law does not include assessment of the accuracy of decision-making processes. In Nowak the ECJ and AG explicitly took this position, referring to both the Data Protection Directive and GDPR (which was forthcoming at the time of the judgement). The ECJ denied data subjects an opportunity to assess the decision-making process themselves, explaining instead that this evaluation rests with competent sectoral authorities that handle complaints (e.g. an examination procedure[272]). 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

---

[267] *Commission v Bavarian Lager - Case C‑28/08 P* (n 6) para 49.
[268] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 45-47.
[269] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 54-55.
[270] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) para 32 and 60.
[271] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 54-55.
[272] ibid.

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 do not have to be justified to the data subject. The protection will likely be an empty shell, if this view is continued in the future.

V. 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 (see: Section III). 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 (see: Section III.C). 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 (see: Section I.A) suggest that the prescribed remit of data protection law may be too narrow to realise the law's original aims. In this section we argue that 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 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.[273] In this regard the term 'data protection' is misleading, as it suggests that the laws aim to protect the data, when in fact

[273] Wachter, 'Privacy' (n 49); Mireille Hildebrandt and Serge Gutwirth (n 54); Mayer-Schönberger (n 15); Mayer-Schönberger and Cukier (n 25). For an overview of EU discrimination laws and jurisprudence, see: Karoliina Ahtela, 'Cases, Materials and Text on National, Supranational and International Non-Discrimination Law–Edited by Dagmar Schiek' (2009) 15 European Law Journal 674.

it is intended to protect people.[274] Data can directly and indirectly reveal aspects about 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."[275]

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."[276] 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.[277] Protecting this right is a key aim of data protection law. Standing jurisprudence of the ECJ[278] and ECHR[279] has recognised 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 on 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.[280] Stemming from the idea that an individual should have the right to be left alone by the state, the right to privacy was

---

[274] Wachter, 'Privacy' (n 49); Mireille Hildebrandt, 'Profiling: From Data to Knowledge' (2006) 30 Datenschutz und Datensicherheit-DuD 548.

[275] Article 29 Data Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' (n 246) 4.

[276] Lee A Bygrave, *Data Protection Law: Approaching Its Rationale, Logic and Limits* (Kluwer Law Intl 2002) 128–129.

[277] See also: Daniel J Solove, 'I've Got Nothing to Hide and Other Misunderstandings of Privacy' (2007) 44 San Diego L. Rev. 745.

[278] *Criminal proceedings against Bodil Lindqvist C-101/01* (n 103); *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8); *Patrick Breyer v Bundesrepublik DeutschlandCase C‑582/14* (n 106).

[279]*Case of Amann v Switzerland, Application no 27798/95* (European Court of Human Rights)§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" (see the Niemietz v. Germany judgment of 16 December 1992, Series A no. 251-B, pp. 33-34, § 29, and the Halford judgment cited above, pp. 1015-16, § 42). That broad interpretation corresponds with that of the Council of Europe's Convention of 28 January 1981 [...]'; See also: ECHR jurisprudence on privacy until 2017 Council of Europe (n 56).

[280] Alessandro Mantelero and Giuseppe Vaciago, 'Data Protection in a Big Data Society. Ideas for a Future Regulation' (2015) 15 Digital Investigation 104; Alessandro Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection' (2016) 32 Computer Law & Security Review 238, 243.

originally proposed as a defence mechanism against governmental surveillance.[281] 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 Art 14 of the EU Convention of Human Rights.[282] Both aims are reflected in the 1995 Data Protection Directive and now the 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 (Art (9)). Sensitive or protected attributes are linked to observable variables that have historically proven discriminatory (e.g. ethnicity, religion).[283]

As the novel risks of automated decision-making and profiling suggest (see Section I), 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 are 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 (see: Section I.A).

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

[281] Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics' (n 280) 245; Christoph Grabenwarter, *The European Convention for the Protection of Human Rights and Fundamental Freedoms: A Commentary* (01 edition, Beck/Hart Publishing 2014).
[282] Grabenwarter (n 281).
[283] Mittelstadt (n 27).

source.[284] Third party personal data, anonymised data, and other forms of non-personal data can also be used to develop inferences and profiles. This background knowledge, built from anonymised, non-personal, or third party data, can then be applied to individual data subjects.[285] 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.[286] Ultimately, affected individuals are not (fully) able to exercise their data protection rights (e.g. access[287] or erasure[288]) until standalone inferences or profiles based on anonymised, non-personal, or third party data have been applied at an individual level.[289] By using data about people not linked to a particular individual, or by purposefully anonymising data prior to drawing inferences and constructing profiles,[290] 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

---

[284] Jeffery L Johnson, 'Privacy and the Judgment of Others' (1989) 23 The Journal of Value Inquiry 157.

[285] Wim Schreurs and others, 'Cogitas, Ergo Sum. The Role of Data Protection Law and Non-Discrimination Law in Group Profiling in the Private Sector', *Profiling the European citizen* (Springer 2008).

[286] ibid 248 takes the view that data subjects need to consent before data is anonymised.

[287] Hildebrandt explains "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." Hildebrandt, 'Profiling: From Data to Knowledge' (n 274) 550.

[288] Rubinstein for example doubts that the right to be forgotten would apply to profiles built from anonymised or aggregated data: "it is not even clear whether Article 17 [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." Ira Rubinstein, 'Big Data: The End of Privacy or a New Beginning?' (2013) 3 International Data Privacy Law 74, 80.

[289] See also: Hildebrandt, 'Profiling: From Data to Knowledge' (n 274) 550; on why exclusion of anonymous data from data protection law is a problem, see: Schreurs and others (n 285) 241.

[290] Schreurs and others (n 285) 248.

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.[291] 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 the name is not known, but other characteristics describe the person sufficiently.[292] The 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" (Recital 26 GDPR).

This can have major implications for assessing problematic behaviour of the car, such as a crash, not least because such a definition of 'identifiability' is fluent and changes with advances in technology.[293] Scholars have shown that 'anonymised' data can often be linked back to individuals.[294] 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 on his or her life. Schreus et al. give the example of a shopping trolley that only senses the products put in the basket and the speed at which the trolley is pushed in order to offer certain products.

---

[291] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Transparent, Explainable, and Accountable AI for Robotics' (2017) 2 Science Robotics.

[292] Korff, 'Comparative Study on Different Approaches to New Privacy Challenges, in Particular in the Light of Technological Developments-Contract No' (n 256) 45; on why the distinction between identifiable and non-identifiable uses is important in the Big Data era, see: Colin J Bennett and Robin M Bayley, '8 Privacy Protection in the Era of "Big Data": Regulatory Challenges and Social Assessments' [2016] Exploring the Boundaries of Big Data 205; Ira S Rubinstein and Woodrow Hartzog, 'Anonymization and Risk' (2016) 91 Wash. L. Rev. 703.

[293] Korff, 'Comparative Study on Different Approaches to New Privacy Challenges, in Particular in the Light of Technological Developments-Contract No' (n 256) 46.

[294] Latanya Sweeney, 'Only You, Your Doctor, and Many Others May Know' (2015) 2015092903 Technology Science; Ohm (n 242); Nadezhda Purtova, 'Do Property Rights in Personal Data Make Sense after the Big Data Turn?: Individual Control and Transparency' (Social Science Research Network 2017) SSRN Scholarly Paper ID 3070228 <https://papers.ssrn.com/abstract=3070228> accessed 11 September 2018; Vijay Pandurangan, 'On Taxis and Rainbows' (*Vijay Pandurangan*, 21 June 2014) <https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1> accessed 29 July 2018.

In this case, the customer does not need to be identified for choices to be tailored to his or her perceived preferences or needs.[295]

To prevent data harms (e.g. discrimination) and bypass the murky issue of what constitutes personal data, it has been suggested to treat all data (e.g. even weather data[296]) as personal data.[297] While such a bold proposal has its merits, such as eliminating overlapping boundaries between personal and non-personal data, such a radical step is unnecessary to address 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 that we face does not lie so much with data collection, but rather with what can be read from the data and the decision that are based on this knowledge.

Therefore, we suggest that continuing to rely on sensitivity and identifiability, or the blurry distinction between personal data, sensitive data, non-personal and anonymised data, as metrics for the level of protection to grant 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 be more in line with the ECHR[298] and the Council of Europe's "Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data,"[299] and would reconfigure privacy as a holistic concept. 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.

We are, however, doubtful of the immediate political appeal of such a move, given a recent proposal in the EU to facilitate exchange of non-

[295] Schreurs and others (n 285) 246ff.

[296] Purtova (n 73). It is worth noting that this example is intentionally hyperbolic.

[297] Sandra Wachter, 'Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR' (Social Science Research Network 2017) SSRN Scholarly Paper ID 3083554 <https://papers.ssrn.com/abstract=3083554> accessed 27 December 2017; Sandra Wachter, 'Privacy: Primus Inter Pares — Privacy as a Precondition for Self-Development, Personal Fulfilment and the Free Enjoyment of Fundamental Human Rights' <https://papers.ssrn.com/abstract=2903514> accessed 14 September 2017; Ohm (n 242); Peppet (n 43); Mireille Hildebrandt and Serge Gutwirth (n 54).

[298] For an overview of ECHR jurisprudence on privacy to 2017, see also: Council of Europe (n 56).

[299] 128th Session of the Committee of Ministers, Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data by the he Member States of the Council of Europe 2018 [CM/Inf(2018)15-final].

personal data.[300] Unfortunately, the proposal lacks serious consideration of the privacy risks of non-personal data, along the lines outlined above.[301] 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, we believe that 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 would offer data subjects additional protections against inferences (1) predicted or shown to cause reputational damage or invade one's privacy, including through application in an automated decision-making process; and (2) have low verifiability in the sense of being predictive or opinion-based.

To make such a right feasible, the European Court of Justice should broaden their interpretation of data protection law regarding 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 (Art 22(3) GDPR). Justification is established by providing evidence of the relevance and reliability of inferences and methods used to draw them. If implemented, high risk inferences would receive comparable levels of protection to automated individual decision-making.[302]

A. Justification to establish 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 is a relevant basis to draw inferences; (2) why these inferences are relevant for

[300] European Commission, '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' (2017) COM(2017)495 <https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-framework-free-flow-non-personal-data> accessed 31 July 2018.

[301] Graef, Husovec and Purtova (n 244).

[302] Wachter, Mittelstadt and Floridi, 'Why There Is No Right to Explanation in the General Data Protection Regulation' (n 16).

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.[303] These requirements 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 s at scale.[304] Finally, these requirements would apply equally to inferences drawn by the data controller and those received from a third party and re-purposed.

In the first instance, we suggest that the right should apply only to 'high-risk inferences' which are (1) privacy-invasive or damaging to reputation, or have a high likelihood of being so in the future, and (2) have low verifiability in the sense of being predictive or opinion-based. The first condition effectively sets a proportionality test, according to which the privacy invasion or reputational damage posed by using a particular data source to draw an inference must be proportional to the 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.[305] 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. (see: Section V.B).

This set of conditions is proposed as a starting point for application of the right to reasonable inferences. Requiring low verifiability in additional to damage to privacy or reputation may, for instance, prove too high a threshold in practice. The necessary conditions for the right to apply should remain open to debate to determine their impact and assess whether minor alterations are required for sectoral application.

Alternative grounds for application or additional conditions may also be feasible. For example, the right could alternatively be based on the notion of "legal or similarly significant effects" as prescribed in Art 22(1) GDPR. Adopting this threshold would position the right as a complementary

[303] On why the focus on attributes that one cannot change (e.g. ethnicity) is unhelpful, as talent and intelligence cannot be changed either but are treated as a legitimate basis for decision-making, see: Janneke Gerards, 'The Discrimination Grounds of Article 14 of the European Convention on Human Rights' (2013) 13 Human Rights Law Review 99, 115 and footnote 70.

[304] The caveat 'at scale' is included to ensure data controllers can carry out initial processing necessary to demonstrate relevance and reliability. Without this condition, data controllers would be unable to engage in exploratory analysis or develop new methods and types of inferences. Our intention is to introduce justificatory requirements to meet prior to widespread deployment, not to prevent development and deployment themselves.

[305] This type of assessment could conceivably form part of a data protection impact assessment (Art 35 GDPR) if a sufficient level of external review or governance can be guaranteed.

protection for the right not to be subject to automated individual decision-making (Art 22 GDPR), which may be desirable. The Article 29 Working Party has provided examples of such effects in relation to Art 22 GDPR: differential pricing and targeted advertisements that affect vulnerable groups (e.g. children playing online games profiled as susceptible to advertisements, or adults experiencing financial difficulties).[306] However, the precise scope of "legal or similarly significant effects" remains unclear in practice, but 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 (Art 13-15 GDPR) should include "details of the main characteristics considered in reaching the decision, the source of this information and the relevance."[307] The Working Party explicitly warn that data controllers should prevent "any over-reliance on correlations,"[308] and explain why a "profile is relevant to the automated decision-making process."[309]

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.[310] The need to demonstrate reliability aligns with Recital 71 GDPR, 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 to prevent discriminatory effects of automated decision-making.[311] Similarly, the Article 29 Working Party explicitly call

[306] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 22, 29.
[307] ibid 26.
[308] ibid 28.
[309] ibid 31.
[310]"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." Quoted from: Schreurs and others (n 285) 253."
[311] Recital 71 states: "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

for "auditing algorithms" to be implemented to assess "the accuracy and relevance of automated decision-making including profiling."[312] Controllers have a similar responsibility for input data, which must be shown to not be "inaccurate or irrelevant, or taken out of context,"[313] and to not violate "the reasonable expectations of the data subject"[314] in relation to the purpose for which the data was collected.[315] 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."[316] They recommend that data controllers "should periodically and within a reasonable time re-evaluate the quality of the data and of the statistical inferences used."[317]

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 (Bundesdatenschutzgesetz), although it is worth noting that this law is no longer in force.[318] Specifically, § 28b required data controllers making predictions or predictive inferences to establish that:

---

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."

[312] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 28.

[313] ibid 17.

[314] ibid 11.

[315] Bart Custers and others, 'Informed Consent in Social Media Use-The Gap between User Expectations and EU Personal Data Protection Law' (2013) 10 SCRIPTed 435.; Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 11.

[316] Council of Europe, 'Recommendation CM/Rec(2010)13 of the Committee of Ministers to Member States on the Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling' (23 November 2010) 1 <http://194.242.234.211/documents/10160/10704/Recommendation+2010+13+Profiling.pdf> accessed 30 March 2018.

[317] ibid 3.10.

[318] Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz 2010 - DSG 2010) 2010.

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 cannot be based solely on a data subject's physical address (e.g. post code).
4. If physical addresses are used, the data subject must be informed of this fact, and it must be documented that the data subject has been so informed.

These requirements closely align with our proposal for data controllers to establish the 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 behaviour). This type of information is intended to be the starting point of a dialogue between data subjects and data controller 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 we 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, inclusive, and discursive, and to provide a new channel of remedies for the data subjects that believe 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.[319] This right would allow data subjects to contest inferences

[319] In favour of such a solution, see: Mireille Hildebrandt and Bert-Jaap Koops, 'The Challenges of Ambient Law and Legal Protection in the Profiling Era' (2010) 73 The Modern Law Review 428, 449; On the need to remedy unjust judgements based on inferences, see: Leenes, Hildebrandt and Gutwirth (n 66) 295.

themselves (e.g. credit score), which complements the existing right to contest automated decisions[320] (Art 22(3) GDPR). With the considerations of justification in Section V.A in mind, the right to contest would be transformed from a mere procedural tool (see: Section IV.E) 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. based on non-intuitive, unreliable, or invasive features or source data), and to offer supplementary information that could lead to an alternative preferred outcome. Contesting as imagined here encourages dialogue between the data subject and 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 decisions challenged under Art 22(3)) 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 Art 16 GDPR. This proposal only suggests to broaden the scope of Art 16 from merely input data to also output data, which is in line with the Article 29 Working Party's view (see: Section II).

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 (see: Sections II.B and IV.B). The data subject may nonetheless disagree with the controller's views or assessment, for example because 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 relevance or reliability of an inference on any of these grounds is distinct from rectifying a provably inaccurate inference.

[320] Mendoza and Bygrave (n 264).

The right to rectification (Art 16 GDPR) may arguably already offer remedy for non-verifiable inferences. Whether this is the case depends upon one's view of the necessity of verifiability in classifying inferences as personal data (see: Sections II.B and IV.B) 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 (see: Sections III and IV.B). In contrast, the Article 29 Working Party believes predictive inferences can also be 'rectified' by providing supplementary information that would alter the assessment, meaning verifiability is not necessary to exercise right of rectification (see: Section II).

The proposal for an *ex-ante* right to contest inference 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 (see: Section IV). 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, due to being able to draw on the justification disclosure made by the controller prior to an inference being drawn.

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

As shown in Sections II and III, 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 counterbalanced with the legitimate interests of data controllers, concerning for example trade secrets, intellectual property, or third-party privacy (see: Section IV).

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. Here, we consider models to be the outputs of data processing involving inferential analytics which 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, data subjects' rights are restricted for the purpose of constructing of 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.

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.[321] After data is collected based on one of the legal bases in Art 6, the strict "purpose limitation" in Art 5 no longer applies. Art 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." The same privilege applies to the strict principle of storage limitation Article 5(e), 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 Art 6, and "appropriate safeguards" pursuant to Art 83 (e.g. pseudonymisation) 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 as well as Zarsky believe that Big Data analytics can be considered 'processing for statistical purposes', as they are

[321] Mayer-Schönberger and Padova (n 45); In contrast, Bertram Raum, 'DS-GVO Art. 89 Verarbeitung zu Archivzwecken, Forschungszwecken' in Eugen Ehmann and Martin Selmayr (eds), *Datenschutz-Grundverordnung* (1st edn, CHBeck 2017) Rn. 31-32 is unsure whether the exemptions apply to Big Data.

strongly based on statistical methods.[322] If this is the case, controllers will enjoy numerous privileges and exemptions from other rights and duties in the GDPR as described in Art 89(2). These include exemptions from Art 14(5)(b), 15, 16, 17(2)(d), 18 and 21, as well as the strict limitations on the use of sensitive data in Art 9(2)(j) and Recital 52.

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"; Art 89(1)), despite having been built with their personal data.

It is important to note a further restriction on the Art 89 privileges. Recital 162 clarifies that statistical results generated under the statistical purposes exemption, as well as the input personal data, cannot be used "in support of measures or decisions regarding any particular natural person."[323] It is 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 Art 89 exemptions to third parties. Presumably, if the results (which must not be personal data; 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 will apply (e.g. Art 6 and 22 GDPR).[324]

---

[322] Mayer-Schönberger and Padova (n 45) 329; Zarsky, 'Incompatible' (n 250).

[323] Zarsky, 'Incompatible' (n 250) 1008; Schreurs and others (n 285) 247–48 are silent on the view of applying profiles after creation but hint that the law 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 Art 6, see: Richter (n 260) 585 who also warns that this can never be sufficiently regulated as we have no way of assessing how the models are subsequently used for other processing or by other data controllers.

[324] "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." Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 7; Raum (n 321) Rn. 41 explains 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 (e.g. consent).

An important point of contention regarding these exemptions is whether they apply to commercial data controllers, or only public and research entities (e.g. government bodies, universities). Mayer-Schönberger and Padova argue that these privileges apply to "private companies for commercial gain."[325] 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.[326] In contrast, 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 GDPR's standard data processing requirements.[327] 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 Section IV apply. Members of the training set also retain rights over any of their personally identifiable data contained within the model. 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 Section III.

In Nowak, the ECJ made clear that the exam questions are not the candidate's personal data,[328] 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 becomes the applicant's personal data.[329] The data subject thus cannot rectify

---

[325] Mayer-Schönberger and Padova (n 45) 326; Richter (n 260) 585 thinks commercial purposes are covered and the later application should not be lawful even if it fulfils Art 6 GDPR requirements due to the possible risks. He does not, however, offer a legal argument to justify this claim. He further warns that the GDPR legalises many applications that would have been illegal in Germany (e.g. private sector uses).

[326] Richter (n 260) 585.

[327] Raum (n 321) Rn. 41-42.

[328] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 58.

[329] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 48, 59 states that not even the legal analysis (as an abstract application of the law) is personal data, only the personal data undergoing processing. We can conclude that the law will also not be seen as personal data.

or delete the law. Both the exam questions[330] and applicable law[331] are thus not personal data, and are therefore not 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 (see: Section III). In Nowak, this was clearly seen in the determination that allowing the candidate to rectify answers on an exam would be nonsensical due to undermining the original processing purpose (i.e. to evaluate 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 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 how decisions are made,[332] and does not allow individuals to decide which models (e.g. exam questions, laws) are used to assess them (see: Section III).[333] Rather, these choices fall within the data controller's private decision-making autonomy.

---

[330] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 58.

[331] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) 48 and 59.

[332] ibid para 46-48; *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) para 32; *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) in respect of comments made by the examiner (para 54) and the exam questions (para 58).

[333] At this point the authors would also like to thank Prof Viktor Mayer-Schönberger for his generous and insightful feedback and advice that strengthened our argument.

An example may help to illustrate why models cannot be considered personal data. If a doctor asks about my height, and I reply 166cm, this utterance is my personal data. This data falls under the GDPR and can be rectified, deleted, etc. However, the fact that my height is expressed in centimetres does not mean that the metric system (i.e. the model used to assess my height) becomes my personal data, meaning I have rights over it. By having my height measured, I will not gain the right to 'rectify' or 'delete' the metric system. Similarly, I would not have a right to require that a different measuring system (or model) be used, for example the imperial system, because I prefer the imperial system or find it more accurate.[334]

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 his responses contained in the rubric (assuming he was still identifiable, for example if the author of the rubric recalled who provided the example in question). In line with his data protection rights over personally identifiable data, the candidate could justifiably request access or deletion of the extracted response (under the conditions outlined in Section IV). In the context of a trained algorithmic

[334] 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 and Lilian Edwards, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' [2018] arXiv preprint arXiv:1807.04644 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.

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.[335]

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 (see: Section III). 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 full, unbounded exercise of rights over it.

Finally, law and policy on intellectual property, 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[336] 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 Copy Rights Directive

The previous section shows that the GDPR facilitates the creation of profiles and models, either built from inferences (among other data), or 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 build 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, but this will not equate to any control or rights over the model as a whole.

[335] On the challenges to implement the right to be forgotten for AI systems, see: Villaronga, Kieseberg and Li (n 220).
[336] ibid.

The facilitation of model constructions and lack of individual rights seen in the GDPR can also be seen in intellectual property (IP) and copyright law. Current discussion of machine learning and inferential analytics in the context of IP law focuses broadly on two issues: (1) is the training data used to construct a model (e.g. content uploaded or created by their users) protected by IP laws; and (2) can the outcome of the algorithmic process be protected under IP law?[337]

A new EU Copyright Directive[338] is currently under debate, which will complement the existing legal framework on copyright[339] and will, among other things, govern the legal status of training data and related uses of "content uploaded by their users"[340] by 'society service providers' (e.g. Internet platforms).

The Directive is among other things concerned with research organisations such as universities and research institutes (including public-private partnerships and commercial scientific research with some limitations[341]) that use new technologies that "enable the automated

[337] Daniel Schönberger, 'Deep Copyright: Up-and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)' (2018) 10 Zeitschrift fuer Geistiges Eigentum/Intellectual Property Journal 35; Annemarie Bridy, 'The Evolution of Authorship: Work Made by Code' (2015) 39 Colum. JL & Arts 395.

[338] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market' (2016) 2016/0280 (COD) / COM (2016) 593 final.

[339] Consisting of: Directive 96/9/EC (legal protection of databases), Directive 2001/29/EC (copyright implementing the "WIPO Copyright Treaty"), Directive 2006/115/EC (rental right and lending right copyright), Directive 2009/24/EC, Directive 2012/28/EU (legal protection of computer programs) and Directive 2014/26/EU (copyright for musical work). And see also other frameworks that are relevant but go beyond the scope of this paper: The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), Art 52 European Patent Convention (EPO) that allows for patenting of some "computer-implemented inventions"(its functionality) if they are "technical contribution in technical field" (e.g. T06/83, IBM Data processor network" and "T1002/92 PETTERSSO/ Queuing system) but not computer programs or mathematical methods; "The EPC as interpreted in the case law enables and obliges the EPO to grant patents for inventions in many fields of technology in which computer programs make a technical contribution. Such fields include medical devices, the automotive sector, aerospace, industrial control, communication/media technology such as automated natural language translation, voice recognition and video compression, and also the computer/processor itself". Quoted from: European Patent Office, 'Patents for Software?' <https://www.epo.org/news-issues/issues/software.html> accessed 1 August 2018.

[340] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market' (n 338) Recital 37.

[341] See: ibid 8, Recital 10, and Art 2(1)b which states "pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by the scientific research cannot be enjoyed on a preferential basis by an undertaking exercising a decisive influence upon such organisation."

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 new trends."[342] 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 licences[343] or remuneration[344]), as well as for exemptions from the Database Directive[345] to uses of data to monitor trends.

These exemptions are concerning when considered alongside the GDPR's exemptions in Articles 85[346] 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." Recital 159 GDPR explains that "scientific research purposes" should be interpreted broadly to include "privately funded research."[347] 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.

For all other 'society service providers', Recital 38 of the draft EU Copyright Directive explains that if providers "store and provide access to the public to copyright protected works or other subject-matter uploaded by their users, thereby going beyond the mere provision of physical facilities and performing an act of communication to the public, they are obliged to conclude licensing agreements with right holders, unless they are eligible for the liability exemption provided." In the associated staffing document,[348] it

[342] ibid Recital 8.

[343] See ibid 10, Recitals 5, 8–9, and Art 3; for arguments in favour of licence fees and access to data for AI training, see: Schönberger (n 337).

[344] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market' (n 338) Recital 13.

[345] Directive 96/9/EC of the European Parliament and of the Council of 11 march 1996 on the legal protection of databases 1996.

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

[347] 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' (2004) 56 Maine Law Review 47; and with a focus on GDPR, see: Gabe Maldoff, 'How GDPR Changes the Rules for Research' <https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/> accessed 1 August 2018.

[348] On the modernisation of EU copyright rules, see: European Commission, 'Commission Staff Working Document - Executive Summary of the Impact Assessment on the Modernisation of EU Copyright Rules' (2016) SWD(2016) 302 final <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016SC0302> accessed 1 August 2018.

is noted that the new provisions are "likely to encourage the conclusion of agreements for the use of content and to generate additional revenues for right holders," but it seems that mandatory licences or remuneration for users providing the data used to train models is not guaranteed.[349]

While the idea of fair remuneration contracts for authors and performers is introduced in Art 14 of the Directive, this can be circumvented if the administrative burden in comparison to the exploitation of the work is too high (Art 14(2)), or the "contribution of the author or performer is not significant having regard to the overall work or performance" (Art 14(3)). Presumably "normal" users of online platforms will not be considered "authors or performers," and thus not entitled to remuneration for contributing data to a platform or service. This is further reinforced by the staffing document which explains that revenues are hoped to be a result of the new regulation, but are not the primary goal.[350] No tendencies are seen in the draft report of the Committee on Culture and Education that would suggest otherwise.[351]

The likely impact of the draft Copyright Directive appears to be to exempt research institutions, including for-profit and commercial research, 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 no claim to remuneration or a licensing agreement under the Copyright Directive's research exemptions. In cases where users upload content to information society service providers not meeting these exemptions, licencing agreements are "encouraged," but remuneration is not guaranteed. Additionally, it remains unclear whether models built on data mining and inferences, as well as further uses of user generated content, are protected by IP laws or trade secrets.[352]

---

[349] The mandatory licence agreement is also only mentioned in Recital 38 of the proposed Copyright Directive, and not Art 13, suggesting the former may not be binding.

[350] The primary goal of Art 13 of the Copyright Directive is to introduce copyright infringement policing on platforms.

[351] Committee on Culture and Education and for the Committee on Legal Affairs, 'Draft Opinion on the Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD))' (2017) <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-%2F%2FEP%2F%2FNONSGML%2BCOMPARL%2BPE-595.591%2B01%2BDOC%2BPDF%2BV0%2F%2FEN> only discusses mandatory licence agreements between users and service providers for "user generated content," while renumeration is not mentioned. See also proposed amendments of Art 13 on p. 15 and of Recital 38 on p. 26-27 ; see also: Malgieri, ''User-Provided Personal Content'in the EU: Digital Currency between Data Protection and Intellectual Property' (n 244).

[352] For a general discussion of this topic, see: Natali Helberger and others, 'Legal Aspects of User Created Content'; for an analysis of how copyright law can mitigate biases in AI, see:

It must be noted that there are currently hundreds of proposed amendments to the EU's Copyright Directive.[353] Due to the number of amendments, at the moment the actual impact of the Directive on inferential analytics and algorithmic models remains unclear.[354]

C. Algorithmic models and outcomes and intellectual property law

Thus far we have 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 according to 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. For non-research environments, the Copyright Directive currently requires a 'society service provider' to have a license agreement in place with the user to process his or her data. The provider does not, however, need to remunerate the data subject even if they hold a copyright. 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 rights.[355]

Amanda Levendowski, 'How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem' (2018) 93 Wash. L. Rev. 579.

[353] For a recent list of amendments adopted by the European Parliament on 13 September 2018, see: European Parliament, 'Copyright in the Digital Single Market ***I - Amendments Adopted by the European Parliament on 12 September 2018 on the Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD))' (2018) P8_TA-PROV(2018)0337 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2018-0337+0+DOC+PDF+V0//EN> accessed 12 September 2018.

[354] For all draft reports of the European Parliament, see: European Parliament, 'Draft Reports | Documents | JURI | Committees | European Parliament' <http://www.europarl.europa.eu/committees/en/juri/draft-reports.html?ufolderComCode=JURI&ufolderId=07947&urefProcCode=&linkedDocument=true&ufolderLegId=8&urefProcYear=&urefProcNum> accessed 1 August 2018; for further legal and ethical discussion, see: Zarsky, 'Mine Your Own Business: Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion' (n 59); Bart W Schermer, 'The Limits of Privacy in Automated Profiling and Data Mining' (2011) 27 Computer Law & Security Review 45.

[355] Madeleine de Cock Buning, 'Is the EU Exposed on the Copyright of Robot Creations?' (2015) 3 Robotics Law Journal 8 <http://www.cier.nl/wp-content/uploads/2015/10/MdeCockBuning-Copyright-and-robotisc-robotics-law-journal_03_november_2015.pdf> states that "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

This means that the new EU Copyright Directive or the InfoSoc Directive2001/29/EC could apply to work generated by algorithms, in addition to training data.[356]

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." In the ECJ's judgement in SAS Institute Inc. v World Programming Ltd, this has been interpreted as applying to at least preparatory design material, machine code, source code, and object code, but not the functionality of the computer program or the format of the data files.[357] Following this judgement, 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

---

be both. It can also be the entity or person who invested financially in the software" ; Christophe Leroux and others, 'Suggestion for a Green Paper on Legal Issues in Robotics' (2012) 3 Contribution to Deliverable D; European Parliament Committee on Legal Affairs, 'Report with Recommendations to the Commission on Civil Law Rules on Robotics' (European Parliament 2017) 2015/2103(INL) 28 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0005+0+DOC+PDF+V0//EN> accessed 11 November 2016; Malgieri, 'Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data' (n 212); European Parliament, 'Civil Law Rules on Robotics - European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))' (n 17); Muzdalifah Faried Bakry and Zhilang He, 'Autonomous Creation – Creation by Robots: Who Owns the IP Rights?' (*IPKM Blog*, 2015) <https://law.maastrichtuniversity.nl/ipkm/autonomous-creation-creation-by-robots-who-owns-the-ip-rights/> accessed 5 August 2018.

[356] 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: James Grimmelmann, 'There's No Such Thing as a Computer-Authored Work-And It's a Good Thing, Too' (2015) 39 Colum. JL & Arts 403; Bridy (n 337); Schönberger (n 337) who explores the idea that the AI creation and the copyright should be in the hands of the public domain.

[357] *SAS Institute Inc v World Programming Ltd,Case C-406/10* (European Court of Justice (Grand Chamber)).

Secrets Directive[358] is likely to substantially limit controller's transparency obligations. The framework, which came into effect on 9 June 2018, may result in the creation of new data being classified as a trade secret. Art 2 of the Directive defines a trade secrets as any information that is not "generally known," has commercial value due to this secrecy, and has been subject to reasonable steps ensure it remains a secret.[359] Recital 1 further adds "valuable know-how and business information" to the definition.

The definition of a trade secret is so broad to as include nearly any data handled by a commercial entity. For example, trade secrets could include "shopping habits and history of customers,"[360] "customer lists and profiles,"[361] "algorithms,"[362] and "information about creditworthiness, lifestyle, reliability, etc., personalized marketing plans (eg pricing), or forecasts about customer's future life based on probabilistic studies (life expectancy, estimated advancements in career, etc.)."[363]

An EDPS document commenting on an early draft of the Directive[364] and a European Commission impact assessment accompanying the proposal for the Directive[365] further clarify the scope of trade secrets. According to these

[358] 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.

[359] Art 39(2) of the TRIPS agreement which has a similar definition of a trade secret. See also 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' (2011) 12 UC Davis Bus. LJ 87; Amy J Schmitz, 'Secret Consumer Scores and Segmentations: Separating Haves from Have-Nots' [2014] Mich. St. L. Rev. 1411; Pasquale (n 16); Rebecca Wexler, 'Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System' (2018) 70 Stan. L. Rev. 1343.

[360] Inge Graef, Martin Husovec and Nadezhda Purtova, 'Data Portability and Data Control: Lessons for an Emerging Concept in EU Law' 15.

[361] Purtova (n 294) 10.

[362] Guido Noto La Diega, 'Against the Dehumanisation of Decision-Making: Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information' [2018] Journal of Intellectual Property, Information Technology and Electronic Commerce Law RN 34.

[363] Malgieri, 'Trade Secrets v Personal Data: A Possible Solution for Balancing Rights' (n 228) 113–114. 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."

[364] European Data Protection 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' (2014) 3 <https://edps.europa.eu/sites/edp/files/publication/14-03-12_trade_secrets_en.pdf>.

[365] European Commission, '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

sources, trade secrets can consist of "data such as information on customers and suppliers, business plans or market research and strategies,"[366] "list of clients/ customers; internal datasets containing research data,"[367] "private collations of individual items of publicly available information,"[368] 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.[369] 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 are required to address adequately the potential effects of the proposal on the rights to privacy and to the protection of personal data."[370] The EDPS also recommended amending Art 4 of the Trade Secrets Directive to ensure that the data 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"[371] is guaranteed, referring to a case involving Facebook[372] 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," 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 the scope, it is safe to assume that derived and inferred data will be covered by the Trade Secrets Directive.[373] Even with this outlook, a fair balance between the right of privacy, IP laws, and the rights to

Secrets) against Their Unlawful Acquisition, Use and Disclosure"' (European Commission 2013) SWD(2013) 471 final Annex 4 and 21.

[366] European Data Protection Supervisor (n 364) 3.

[367] ibid.

[368] ibid.

[369] ibid.

[370] ibid 2.

[371] ibid 5.

[372] Facebook User Operations – Data Access Request Team, 'Access Request Response Max Schrems' <http://www.europe-v-facebook.org/FB_E-Mails_28_9_11.pdf> accessed 1 August 2018.

[373] For an overview of the definition of trade secrets according to the ECJ, see: ECJ Judgement of 18 September 1996, Case T-353/94 (Postbank v Commission), para 87, and ECJ Judgement of 30 May 2006, Case T-198/03 (Bank Austria Creditanstalt v Commission), para 71.

conduct a business and freedom of expression will be necessary; jurisprudence of the ECJ has long reflected this position.[374]

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.

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 we have provided and has been observed, but also information derived or inferred from it as well as from anonymous or third party data. Too much emphasis is placed on governing the collection of these types of data, while too little is paid to how it is evaluated.

For example, 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 decisions 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 behaviour that cannot be verified or refuted at the time of decision-making. Thus the actual risks posed by Big Data analytics and AI are the underpinning inferences that determine how we, as data subjects, are being viewed and evaluated by third parties.

In this paper, we have 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 (Art 13-14), access (Art 15), rectify (Art 16), delete (Art 17), and object to them (Art 21). Further, profiling and automated decision-making, which may include inferences, can already be contested (Art 22). The Article 29 Working Party sees verifiable and unverifiable inferences as personal data (e.g. results of a medical analysis), but leaves

[374] C-275/06 Promusicae; Case C-70/10, Scarlet Extended Sa V. Societe Belge Des Auteurs, Compositeurs Et Editeurs Scrl (Sabam), 24.11.2011 ("Scarlet"), C-557/07, Lsg-Gesellschaft Zur Wahrnehmung Von Leistungsschutzrechten Gmbh V. Tele2 Telecommunication Gmbh, 19.2.2009 ("Lsg"); Case C-461/10, Bonnier Audio Ab Et Al. V. Perfect Communication Sweden, 19.4.2012 ("Bonnier").

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 the current jurisprudence is inconsistent.

Some scholars are worried that broad interpretation of personal data turns data protection law into the "law of everything."[375] However, as we have shown in Section IV inferences are treated as 'economy class' personal data that is 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 provisions in the GDPR push to facilitate data mining, knowledge discovery and big data analytics by limiting data subjects' right 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.

However, 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 European Court of Justice (in Bavarian Lager,[376] YS. and M. and S.,[377] and Nowak[378]) and Advocate General (in YS. and M. and S.[379] and Nowak[380]) 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

---

[375] Purtova (n 73).

[376] *Commission v Bavarian Lager - Case C‑28/08 P* (n 6) para 49.

[377] *YS, M and S v Minister voor Immigratie, Integratie en AsielJoined Cases C-141/12 and C-372/12* (n 7) para 45-47.

[378] *Peter Nowak v Data Protection Commissioner Case C‑434/16* (n 8) para 54-55.

[379] *YS, M and S v Minister voor Immigratie, Integratie en Asiel - Joined Cases C-141/12 and C-372/12 - Opinion of Advocate General Sharpston* (n 9) para 32 and 60.

[380] *Peter Nowak v Data Protection Commissioner Case C-434/16 - Opinion of Advocate General Kokott* (n 10) para 54-55.

rights to individuals, for example the right of access, rectification and deletion.[381]

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, we propose a new "right to reasonable inferences" applicable to inferences based non-verifiable and counterintuitive predictions which invade an individual's privacy or damage reputation. 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 is a relevant basis to draw inferences; (2) why these inferences are 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 allows data subjects to challenge unreasonable inferences, which can support challenges against automated decisions exercised under Art 22(3) GDPR..

Of course, a solution outside of data protection law may be possible.[382] 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.

In the same way it was necessary create a "right to be forgotten" in a Big Data world,[383] we argue that 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 original remit proposed in the ECHR,[384] as well as the Council of Europe's "Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data."[385] 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

[381] *College van burgemeester en wethouders van Rotterdam v M. E. E. Rijkeboer C-553/07* (n 11).

[382] Bert-Jaap Koops, 'The Trouble with European Data Protection Law' (2014) 4 International Data Privacy Law 250.

[383] Mayer-Schönberger (n 15); van Hoboken (n 14).

[384] See also ECHR jurisprudence on privacy until 2017, reviewed in: Council of Europe (n 56).

[385] 128th Session of the Committee of Ministers Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data by the he Member States of the Council of Europe (n 299).

obligation" of states to implement laws to protect citizens from privacy invasion by the public and private sector.[386]

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

1. 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 accuracy of decision-making processes. 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 how they are being seen or assessed by automated systems.

2. 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 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,[387] or offers grounds for discrimination and other harms.

Basing protections on these distinctions is thus ineffective.[388] The damage that can be done by data does not depend on any of these categories, but rather how it is used. Inferences or profiles drawn from any of these sources can be

---

[386] Paul De Hert and Serge Gutwirth, 'Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power' [2006] Privacy and the criminal law 61, 128; Wachter, 'Privacy' (n 297).

[387] See Ernst (n 78) citing the judgement of the 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): "Dadurch kann ein für sich gesehen belangloses Datum einen neuen Stellenwert bekommen; insoweit gibt es unter den Bedingungen der automatischen Datenverarbeitung kein 'belangloses' Datum mehr.".

[388] Rubinstein (n 288) 7; Schreurs and others (n 285) both argue that anonymous data can still impact on data subjects, despite being outside the scope of data protection law. Zarsky, 'Incompatible' (n 250) 1013 explains that any data could potentially become sensitive data, rendering the classification meaningless.

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.

We recommend adopting the position taken by the Article 29 Working Party concerning the transformation of categories of data based upon processing purposes and impact.[389] In future European policy-making and jurisprudence, levels of protection should be granted to data based primarily on their usage and impact, and secondarily on their source.

3. 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, because the identifiability of data is fluid and can change over time, depending on linkage, re-identification attacks, and other technological progress.[390]

Companies can use anonymisation[391] techniques to avoid many obligations under data protection law. Similarly, pseudonymisation techniques[392] can potentially minimise the requirements to respect individual rights. In such cases, data controllers are not required to comply with requests from data subjects under Art 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 (Art 11(2)). 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.[393]

As we have argued above, inferences drawn from anonymous and non-personal data still pose risks for data subjects (see: Sections I.A and V). 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

[389] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 8.
[390] Korff, 'Comparative Study on Different Approaches to New Privacy Challenges, in Particular in the Light of Technological Developments-Contract No' (n 256) 46.
[391] Schreurs and others (n 285) 248.
[392] Frederik J Zuiderveen Borgesius, 'Singling out People without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation' (2016) 32 Computer Law & Security Review 256.
[393] See similar prior experiences where companies have claimed 'disproportionate effort' to avoid compliance with access requests. For example: Facebook User Operations – Data Access Request Team (n 372).

affected parties can be identified.[394] 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.

4. 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 (see: Section V.A). Controllers should pay increased attention to addressing the following aspects of the source data and outputs of inferential analytics in addressing justification (see: Section V.A):

- The privacy invasiveness and the counter-intuitiveness of the data sources used to draw inferences, for example clicking behaviour, browsing behaviour,[395] or mouse tracking.[396]
- 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 data (e.g. post code), or the intention to infer sensitive attributes (e.g. political views[397]) from non-sensitive data.
- The relevance of the source data and inference to a particular processing purpose. For example, the relevance of Facebook profiles and friend networks to loan decisions.[398]
- The statistical reliability of the methods used to draw inferences.

[394] Mittelstadt (n 27); Mantelero, 'From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era' (n 27); Bygrave (n 276) ch 15.
[395] Allerhand and others (n 39).
[396] Chen and others (n 42).
[397] 'Anti-Choice Groups Use Smartphone Surveillance to Target "Abortion-Minded Women" During Clinic Visits' (n 34).
[398] Taylor and Sadowski (n 33).

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

5. Give data subjects the ability to challenge unreasonable inferences

In line with the implementation of a right to reasonable inferences, European policy-makers should grant data subjects a new right to challenge unreasonable high-risk inferences, which can also support challenges to subsequent decisions.[399] Data subjects can raise an objection with the data controller on the grounds that the inference or its source data is irrelevant or unreliable (see: Section V.B). 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 (Art 16 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 (see: Sections II.B, III. and IV.B). 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[400] and profiling[401] with legal and significant effects (Art 22(3) GDPR), and potentially transform it from a merely procedural tool to a meaningful accountability mechanism (see section IV. E).

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 which are open to

[399] For a favourable view of such a solution, see: Hildebrandt and Koops (n 319) 449; on the need to remedy unjust judgements based on inferences, see: Leenes, Hildebrandt and Gutwirth (n 66) 295.
[400] For a discussion on this legal loophole see Wachter, Mittelstadt and Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (n 19); Bygrave (n 276).
[401] Mendoza and Bygrave (n 264).

comments and interrogation by the latter (Art 22(3)).[402] 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 (see: Section I.A).

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.

[402] 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, see Art 22(3) GDPR.