Dark Patterns in Personal Data Collection: Definition, Taxonomy and Lawfulness

Luiza Jarovsky

I. Introduction

Bob starts reading a newspaper article on a trending political topic, and when he wants to scroll down, the screen is blocked, and a banner appears asking him to create an account. The signup button says, 'Signup and never miss out on the next hot topic!' He feels that it is better to sign up and be up to date with political news than to be lazy. The signup screen asks for his name, email, telephone, and home address, and requests him to check boxes referring to his personal interests (so that the articles can be specially tailored for him) and to accept marketing offers by email and SMS. He would rather not receive any marketing offers, but when he unselects these options, a banner appears stating that 'some of the website's functionalities might not work well,' so he selects them back, feeling there is no other option.

Alice heard it is important to protect her privacy, so she opens her favourite social network and looks up the settings, to check that everything is ok. She has some difficulty with how to proceed, as some of the configurations are under 'settings and privacy,' others under 'visibility,' others under 'sign in & security,' and yet others under 'data privacy'. Moreover, after clicking each link there are additional texts with information, drop-down menus with multiple links, links that redirect her to external URLs and buttons whose meaning are unclear. It all seems too complicated, and she

gives up, feeling certain that her interests are not correctly reflected in the configurations.

Charlie recently started using a new social app focused on funny videos. He considers himself a shy person, so the possibility of recording short clips using filters, music and subtitles seems to be a good opportunity to surprise his closest friends. He takes care not to add anyone as a friend there, so that he can start recording 'in secret' and practice with the app tools before he shares with anyone. It is easy and immediate to record, so he spends hours creating funny dances animated by pop music. When he goes to sleep, strangely, his phone beeps non-stop. It turns out that the default setting of this app is to share all videos publicly and to allow others to download them. An influencer with a large following reposted one of his videos using an offensive hashtag and tagged him, and now there are hundreds of people adding their own offensive comments on his videos and re-uploading edited versions. He became a meme and feels helpless.

Susan works online all day and is really annoyed with the number of banners requesting consent for cookie collection. The two options that are usually offered are 'accept all' and 'more information'. She does not really know what cookies are, so once she clicked the 'more information' and dozens of green switches with unknown names appeared, and these companies seemingly were collecting her data. It would take too much time to turn off all the switches, and she cannot see very well, therefore since then she just clicks 'accept all' every time. She heard that the parliament in her

country enacted a Data Protection Law, therefore nothing bad could happen. Recently she started receiving various email offers from a business she never heard of. She has also been receiving SMS and calls from representatives; it is usually hard to distinguish what is legitimate and what is not, as the offers tend to be suitable to her. Last week a privacy organisation informed her that her credit card information was among the data leaked from a company she never heard about, so she should cancel her card. She feels helpless with the internet and wishes she could just live like in the old times.

The cases above are adaptations of real situations commonly experienced online. All of them contain one or more *dark patterns in personal data collection* (DP) and, unfortunately, they are common in today's online landscape, even among services run by publicly traded global companies.

In this Article, I argue that DP represent a new challenge to the emerging global privacy laws, especially to the European Union (EU) data protection framework and the General Data Protection Regulation (GDPR),[1] which is the main jurisdictional focus of the present work. DP involve the exploitation of cognitive biases and the deployment of malicious techniques to manipulate data subjects through the interface design of apps and websites. I contend that despite an apparent absence of legal tools to deal with DP, the current EU framework can be adapted and

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

amended to identify and curb them, especially through a refinement of the requisites for lawfulness of consent and the reformulation of the fairness principle in data protection.

I am especially motivated by the complexity of the study of dark patterns in the privacy context, which involves interdisciplinary analysis of design, human-computer interaction (HCI), computer science, cognitive psychology, behavioural economics, ethics, and law. Law has traditionally refrained from closely regulating design,[2] given the risk of hindering technological innovation or being too rigid to adapt to the varied and rapidly changing set of business practices. However, DP have the potential to create ruptures in the web of protections currently afforded to data subjects and lower the already volatile amount of trust that is awarded to service providers. Therefore, this Article aims to integrate these various fields of knowledge through the discussion of cognitive biases that underlie many DP, their manifestation within technological environments, the rationale behind their deployment, and their legal meaning.

Dark patterns do not target exclusively personal data; they consist of a broader phenomenon and can affect other areas such as one's finances, emotions, attention and so on. The term *dark patterns* was first coined in 2010 by Harry Brignull,[3] who launched a website called *darkpatterns.org*. Brignull defines dark patterns as 'tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something'.[4] The HCI, computer

[2] For the challenges of regulating design, see Woodrow Hartzog, *Privacy's Blueprint: The Battle to Control the Design of New Technologies* (Harvard University Press, 2018).

[3] Harry Brignull, 'Dark Patterns: Deception vs. Honesty in UI Design', A List Apart (November 1, 2011), at https://alistapart.com/article/dark-patterns-deception-vs-honesty-in-ui-design (last visited on 2 March 2022 – all cited websites were last visited on this date).

[4] Dark Patterns (homepage), at http://darkpatterns.org.

science and legal literature offer various descriptions for dark patterns,[5] reflecting different nuances that can be highlighted from the point of view of these fields.

When discussing the essence of DP, it is important to raise the topic of nudges. Juxtaposing them may clarify some of the boundaries and help understand how dark patterns differ from other practices. A nudge, a term coined by economist Richard Thaler and jurist Cass Sunstein, is 'any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting the fruit at eye level counts as a nudge. Banning junk food does not'.[6] Designers therefore can manipulate interface design elements to steer individuals to a certain direction, nudging them.

[5] Such as: 'instances where designers use their knowledge of human behaviour (*e.g.*, psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest' – see Colin Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt & Austin Toombs, 'The Dark (Patterns) Side of UX Design' (2018) CHI 2018. 'Intent on the part of the designer to deliberately sacrifice the user experience in an attempt to achieve the designer's goals ahead of those of the user' – see Gregory Conti & Edward Sobiesk, 'Malicious Interface Design: Exploiting the User' (2010) International World Wide Web Conference Committee 271, 1. Interfaces whose goal is 'to exploit cognitive vulnerabilities to guide users towards targeted choices' see Gregory Day & Abbey Stemler, 'Are Dark Patterns Anticompetitive?' *(2020)* 72 Ala. L.R 1, 3. 'Interface designs that try to guide end-users into desired behaviour through malicious interaction flows' – see Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, Lalana Kagal, 'Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence' (*2020)* CHI '20, 3. 'User interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might not make' - see Arunesh Mathur, Gunes Acar, Michael Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, Arvind Narayanan, 'Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites' (2019) Proceedings of the ACM Human-Computer Interaction 3 CSCW, Article 81, 2. 'Features of interface design crafted to trick users into doing things that they might not want to do, but which benefit the business in question' and 'deliberately misleading users through exploitative nudging' – see Forbrukerrådet, 'Deceived by Design: How Tech Companies Use Dark Patterns to Discourage us From Exercising our Rights to Privacy' (2018) at 7, at https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf. And lastly, the statutory definition in California: 'a user interface designed or manipulated with the substantial effect of subverting or impairing user autonomy, decision-making, or choice, as further defined by regulation'. See California Consumer Privacy Act (CCPA), Cal. Civ. Code § 1798.140(l).

[6] Richard Thaler & Cass Sunstein, *Nudge: Improving Decisions about Health, Wealth* and Happiness (Penguin Books 2009) 6.

Despite this author's negative stand on the deployment of nudges without openness and transparency on the agenda of the choice architect,[7] in this work, not all nudges and manipulative designs will be considered dark patterns.[8] To be considered a dark pattern, the design must be manipulative *and* malicious. A design that is manipulative but not malicious can be said to be morally problematic,[9] as the designer is trying to encourage the user to take a different course of action through the exploitation of cognitive biases. However, despite the wrongful means, the end goal might be beneficial, deriving from a background of ethical design,[10] as it happens with some nudges. For example, interface designs that nudge people to eat less calories or protect their privacy will not be considered dark patterns,[11] even if they eventually cause the affected person to be worse off according to his or her own preferences and wishes. Therefore, manipulation in itself and without the malicious element will not be a criterion to identify a dark pattern.

---

[7] For this author, nudges are problematic, as they rely on cognitive biases to steer people in a desired direction. A frequent example of nudges is the deployment of defaults that make people save money or donate organs after death. These examples might seem inoffensive and beneficial to the individual or to society. However, autonomy and human dignity are also important values and there might be situations in which both individual and society would benefit more from a well thought or personalised choice according to the concerned individual's values, beliefs, and rights. Moreover, the argument that the individual can still choose other alternatives when nudges are present is weak, as if nudges are effective, they leave little or no space for deeper reflection and consideration of values and ideas that are different than those advanced by the choice architect.

[8] Manipulative design will be defined as one that 'does not sufficiently engage or appeal to their capacity for reflection and deliberation'. Cass Sunstein, 'Fifty Shades of Manipulation' (2016) 1 J. Mark. Behav. 213, 218.

[9] On the deontological argument against manipulation, see Tal Z. Zarsky, 'Privacy and Manipulation in the Digital Age' (2019) 20 Theoretical Inq. L. 157, 175.

[10] 'Excellent designs would be those where the designer takes social responsibility and the ethical measures of a designed product into account. If we want to entrust a livable world for our children in the future, then the ethical design criterion should be adapted and implemented as one of the basic principles of design requirements'. See Ahmet Atak & Aydın Şık, 'Designer's Ethical Responsibility and Ethical Design' (2019) 7 Univers. J. Mech. Eng. 255, 262.

[11] Importantly, these interface designs can, nevertheless, cause harm, as they assume what is good or bad without consulting the affected individuals or considering their personal preferences. Mccrudden and King commented on the possible dangers of nudges: '(i)f we emphasize the tendency to 'counteract' biases, then we see that some forms of nudging seem to rely on psychological insights to try to ensure 'good' results. They 'attempt to harness cognitive irrationalities in aid of desired social policy outcomes'. See Christopher Mccrudden & Jeff King, The Dark Side of Nudging: The Ethics, Political Economy, and Law of Libertarian Paternalism (2015) Research Paper 485, U. Mic. Pub. L. 67, 105.

A design that is manipulative *and* malicious is a dark pattern. 'Malicious' will be defined as 'intended to harm or upset other people'.[12] In this work, based on the framework of ethical design and on the social responsibility of designers to care for the impact generated by their designs,[13] any design that makes the individual worse off, will be considered malicious, even if this was not the intention of the individual designer behind it.[14]

The present Article proceeds as follows: Section II proposes a definition for DP, taking into consideration the relevant elements for the present analysis as well as the distinction made above between dark patterns and nudges. Section III brings the general premises and the conceptions of privacy that influence this work. Section IV presents and discusses the cognitive biases involved in DP, showing how they affect data subjects, especially in the privacy context. Section V proposes a taxonomy for DP, offering various examples and a visual representation of each category. Section VI discusses DP's legal status, evaluating the possibilities within the EU legal framework that could legitimise or outlaw them, taking into consideration existing rules and principles. Section VII compares two decision-making paradigms that can be applied to data subjects: *Homo economicus* and *Homo manipulable*, arguing in favour of the latter, which reflects behaviours and traits observed in real life. Lastly, Section VIII concludes, pointing to the avenues to legally outlaw and prevent DP.

---

[12] Cambridge Dictionary, at https://dictionary.cambridge.org/dictionary/english/malicious. The word malicious assumes that there is an intention to harm; the word manipulative does not necessarily assume harm.

[13] See Atak & Şık, *supra* note 10 .

[14] The reason for this extension is that designers have access to a handful of tools and techniques to systematically affect the user's behaviour, including the exploitation of biases, which might work against users unconsciously. This is an enormous amount of power and influence, against which there is not much users can do, except for stopping using online services altogether, which can be deemed practical and, sometimes, unfeasible. Therefore, for the purpose of this work, designers will be considered objectively responsible by the work they create.

II. Definition of DP

Given the focus of the present Article, which aims at understanding the current legal standing of DP and to propose changes in the current framework that would improve the protection offered to data subjects, here I propose the following working definition for DP:

> *A dark pattern consists of user interface design choices that manipulate the data subject's decision-making process in a way detrimental to his or her privacy and beneficial to the service provider*

Below I explain each of its elements:

'user interface design choices': DP happen in the interface design of a product or service. The service can be provided through a browser or an app, and as long as the data subject can interact with the design of the service provider, DP can exist. The terminology 'design choices' is also important, as it highlights that behind the seemingly value-neutral appearance of a service, there are choices involved and professionals with expertise in designing interfaces that can be effective for business,[15] but also harmful for the data subjects' interests,[16] such as DP.

[15] 'If you work in human-computer interaction, you are probably a *choice architect*—even if you have been as unaware of that role' Anthony Jameson, Bettina Berendt, Silvia Gabrielli, Federica Cena, Cristina Gena, Fabiana Vernero & Katharina Reinecke, 'Choice Architecture for Human-Computer Interaction' (2014) 7 Foundations & Trends in Human–Computer Interaction 1, 3.

[16] For a deeper analysis on how design can be deployed to positively impact privacy, see Richmond Wong & Deirdre Mulligan, 'Bringing Design to the Privacy Table, Broadening 'Design' in 'Privacy by Design' Through the Lens of HCI' (2019) CHI 2019.

'that manipulate the data subjects' decision-making process: to manipulate someone can be defined as 'intentionally and covertly influencing their decision-making, by targeting and exploiting their decision-making vulnerabilities'.[17] DP are manipulative because they rely on cognitive biases to steer the data subject's decision-making process in a desired direction.

'in a way detrimental to his or her privacy': the decision must be detrimental to privacy in the sense that will be further developed on Section III below: there must be additional sharing of data or a negative influence on the subject's decision-making process regarding his or her privacy. A nudge or a manipulation that aims at protecting the data subject's privacy or that helps him or her to understand more about privacy, better navigate settings, share less data, or share data with less people will not be considered a dark pattern. DP differ in which cognitive biases they exploit, but all of them exacerbate the knowledge and power asymmetry between data subjects and data controllers. When a designer has an agenda to amplify data collection and surreptitiously twists interface elements to steer the data subject in a certain direction, the designer deprives the data subject of a fair decision-making process about their privacy.[18]

'and beneficial to the service provider': this is one of the outcomes of dark patterns. By keeping the data subject uninformed regarding privacy choices, turning data subjects away from scrutinising how their data is used, promoting unnecessary data sharing, fostering a culture of

[17] Daniel Susser, Beate Roessler & Helen Nissenbaum, 'Technology, Autonomy, and Manipulation' (2019) 8 Internet Pol'y. Rev. 1, 4 (2019).

[18] The harm to data subject can also be framed in terms of autonomy loss, see Daniel Susser, Beate Roessler & Helen Nissenbaum, 'Online Manipulation: Hidden Influences in a Digital World' (2019) 4 Geo. L. Tech. Rev. 1.

unrestrained and unconscious disclosure of personal data and so on, controllers and their partners are beneficiaries of DP.

In the next Section, I clarify the premises and the privacy theories that inform the present analysis.

III. Premises and Privacy Theories

Despite making us feel frustrated, surprised, trapped, ashamed, powerless, and so on, it is necessary to explain why and how DP are objectively harmful to privacy.

An important premise that needs to be clarified is that calling something a DP involves a moral judgement. The working definition, as set in the previous Section, requires a designer steering us to a certain direction that makes us worse off. But how can we objectively access what is a good or a bad outcome? Some cases are easy, as legal principles and rules make it clear what is the legally expected way to act. For example, a design technique that tricks someone into spending more money than she intended is bad, as it would be similar to swindling. On the other hand, a design that calls attention to safety issues of a certain product is good, as it is one of the principles of worldwide consumer protection laws.[19]

However, some cases are not so easy, as there are legal, cultural, and religious differences that result in different perceptions of what is good or bad, right or wrong. For example, what about a design that makes people work more? Or a design that makes people spend more time online interacting with other people? Or a design that promotes the desire to be thin? There are numerous grey areas, as it is a discussion on human values, not an exact science.

[19] See, *e.g.*, Directive 2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety. *OJ L 11, 15.1.2002, p. 4–17*.

Any honest analysis of dark patterns must be transparent regarding the values at stake. The present work explores dark patterns in the field of privacy; therefore, I need to be clear about what conception of privacy I am referring to and what will be considered a negative impact to privacy.

This work relies on two main definitions of privacy. The first is Alan Westin's, which states that privacy is 'the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others'.[20] Westin's definition reflects a model of control, in which the individual has agency over how his or her personal information is shared.[21] It implies that personal data can only be amplified or processed to the extent lawfully consented or legally authorised. This model correctly acknowledges the data subject's autonomy and freedom, expecting her to be the manager of the choices regarding her personal data. It is also aligned with Fair Information Practice Principles (FIPPs),[22] and data protection laws that rely on data subjects' consent as one of the possibilities for lawful processing.[23]

[20] Alan Westin, *Privacy and Freedom* (New York: Atheneum 1967) 5.

[21] According to Birnhack, 'privacy as control is best understood as a concretization of the overarching idea of dignity, applied to personal issues'. Michael Birnhack, 'A Process-Based Approach to Informational Privacy and the Case of Big Medical Data' (2019) 20 Theoretical Inq. L. 257, 263.

[22] According to Schwarz, 'fair information practices are the building blocks of modern information privacy law. They are centred around four key principles: (1) defined obligations that limit the use of personal data; (2) transparent processing systems; (3) limited procedural and substantive rights; and (4) external oversight'. Paul Schwartz, 'Privacy and Democracy in Cyberspace' (1999) 52 Vand. L. Rev. 1609, 1614.

[23] I.e., the Article 6(1)(a) of the GDPR establishes that 'processing shall be lawful only if and to the extent that at least one of the following applies: the data subject has given consent to the processing of his or her personal data for one or more specific purposes'.

Despite correctly emphasising user autonomy, there are numerous shortcomings of consent in privacy, which were studied and categorised by legal[24] and HCI[25] scholars. Given that we are specifically dealing with situations in which user agency is suppressed or inflected – and the data subject cannot properly choose – an additional definition is necessary, one that does not rely on the data subject's control and recognizes any negative interference in the decision-making process as a harm to privacy.

Following this line of thought, the other approach that will be relied on in this work is Ruth Gavison's, who defined the interest in privacy as 'related to our concern over our accessibility to others: the extent to which we are known to others, the extent to which others have physical access to us, and the extent to which we are the subject of others' attention'.[26] Despite being presented in the 1980s – the pre-internet era - Gavison's focus on access precisely captures the idea that we might want some people or entities not to have any access to us. It allows a conception of privacy as a state in which we can be shielded from external entities attempting - at all costs, sometimes even through malicious means, as it happens with DP - to collect more personal data.

I wish to emphasise that this work's approach is not context-dependent, meaning that it does not matter if we are dealing with medical, intimate, or social media data.[27] The threshold for

---

[24] See Luiza Jarovsky, 'Improving Consent in Information Privacy through Autonomy-Preserving Protective Measures (APPMs)' (2018) 4 Eur. Data Protection L. 447 (proposing a classification of the shortcomings of consent).

[25] See Daniel Solove, 'Privacy Self-Management and the Consent Dilemma' (2013) 126 Harv. L. Rev. 1880; Omri Ben-Shahar & Carl E. Schneider, 'The Failure of Mandated Disclosure' (2011) 159 U. Pa. L. Rev. 649; Lorrie Faith Cranor, 'Necessary but Not Sufficient, Standards Mechanisms for Privacy Notice and Choice' (2012) 10 J. Telecomm. & High Tech. L. 273; Lizzie Coles-Kemp & Elahe Kani-Zabihi, 'On-line Privacy and Consent: A Dialogue not a Monologue' (2010) Proceedings of the 2010 Workshop on New Security Paradigms - NSPW '10, 95.

[26] Ruth Gavison, 'Privacy and the Limits of Law' (1980) 89 Yale L.J. 421, 423.

[27] Here, I do not examine the nuances reflected in contextual informational norms. The appropriateness of information flows is important, and I intend to explore it in a subsequent analysis. In this work, I focus on limiting the access that controllers have to data subjects' personal data. For a more detailed approach on the value of contextual integrity for privacy law, see Hellen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford Law Books 2009).

trespassing privacy is the malicious attempt to abuse the data subject's decision-making capacity to access his or her personal data,[28] in whatever context it happens. Therefore, in this work, I assume that any manipulative technique that supports, by action or inaction, additional sharing of personal data or the disruption of the decision-making process regarding privacy, can be considered negative - a harm to privacy.

Another important premise to be stated at this point is that this Article deals only with malicious and manipulative practices that happen within design interfaces. There are, however, other types of malicious and manipulative techniques that can negatively impact privacy online. [29]

In the next Section, I turn to the topic of cognitive biases affecting DP, first explaining their general manifestation and then how they occur in the privacy context.

IV. Cognitive Biases affecting DP

An important part of the study of DP is understanding the cognitive biases they exploit. A cognitive bias is a 'systematic (that is, non-random and, thus, predictable) deviation from rationality in judgement or decision-making'.[30] Cognitive biases can also be understood in terms of the process

[28] Through the exploitation of cognitive biases and implementation of DP.

[29] There are, *i.e*., manipulative ads, which rely on personal data and might explore peer pressure or cognitive biases of the data subject to collect more data. For a broad review of manipulative ads, see Victor Danciu, 'Manipulative Marketing: Persuasion and Manipulation of the Consumer through Advertising' (2014) XXI Theoretical & Applied Economics 19 (2014). There are also discussions on manipulative artificial intelligence (AI), which can be used to deceit data subjects or to bypass privacy choices. For a fascinating forecast of possible negative impacts of AI, see Thomas King, Nikita Aggarwal, Mariarosaria Taddeo & Luciano Floridi, 'Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions' (2020) 26 Science & Engineering Ethics 89.

[30] Fernando Blanco, Cognitive Bias, in Encyclopedia of Animal Cognition and Behavior, (Jeniffer Vonk & Todd Shackelford eds., 2017) 1.

that leads to them and the effects generated by them.[31] The process is a heuristic process,[32] and the result is the bias.

Cognitive biases were originally discussed by cognitive psychologists such as Kahneman and Tversky,[33] who identified that 'people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors'.[34]

Cognitive biases are inherent human traits and can be empirically demonstrated.[35] If privacy law wants to protect real humans (*i.e.*, and not distorted or outdated theoretical models of human rationality), it must correctly acknowledge cognitive biases and their consequences to the decision-making capacity of data subjects.

Some of the biases exploited by DP are listed and explained below.[36] Designers can be very creative, so this list is not exhaustive: there may be new types of DP relying on other biases at any moment. The explanation of the main biases and their connection to the privacy context is

---

[31] Gaëlle Lortal, Philippe Capet & Alain Bertone, 'Ontology Building for Cognitive Bias Assessment in Intelligence' (2014) IEEE International Interdisciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support 237, 237-238.

[32] 'A strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods'. Gerd Gigerenzer & Wolfgang Gaissmaier, 'Heuristic Decision Making' 2011) 62 Annual Rev. Psy. 451, 454.

[33] A more recent discussion was conducted by Paul Slovic, Ellen Peters, Melissa Finucane & Donald MacGregor, 'Affect, Risk, and Decision Making' (2005) 24 Health Psy. 35.

[34] Amos Tversky & Daniel Kahneman, 'Judgement under Uncertainty: Heuristics and Biases' (1974), 185 Science 1124, 1124. There are also opposite views of the cognitive biases theory, which consider them to be advantages, such as Gerd Gigerenzer & Henry Brighton, 'Why Biased Minds Make Better Inferences' *(2009)* 1 Top. Cogn. Sci. 107.

[35] For a deeper view on cognitive biases and their empirical demonstration, see Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux 2011) and Dan Ariely, *Predictably Irrational* (Harper Perennial 2010).

[36] These biases are the ones predominantly present in the DP analysed on Section V *infra*. They are presented in alphabetical order.

relevant for the comprehension of the aspects of human nature that must be taken into consideration by data protection law.

Anchoring bias

'Systematic influence of initially presented numerical values on subsequent judgments of uncertain quantities, even when presented numbers are obviously arbitrary and therefore unambiguously irrelevant'.[37] This bias has been discussed extensively and demonstrated in behavioural psychology, including studies from Tversky and Kahneman[38] and popular culture books.[39] An offline example would be the case of a restaurant owner that adds dishes that are very expensive to the first pages of a restaurant menu; the client will then be anchored by higher values and will consider the remaining dishes' values to be lower and advantageous, even if they are expensive when compared to those of similar restaurants.

This bias has been exploited in the privacy context for example when presenting privacy options to the data subject. In a privacy menu, the pool of values from which the data subject will have to choose from is typically arbitrary, with the designer deciding what will be the broadest-sharing option and the least-sharing option. Relying on the anchoring bias, the designer can choose a first option that is privacy-negligent and additional options that are only mildly protective. The data subject will be 'anchored' by the first option and induced to perceive the additional options as being privacy protective.

---

[37] Predrag Teovanović, 'Individual Differences in Anchoring Effect: Evidence for the Role of Insufficient Adjustment' (2019) 15 Eur. J. Psychol. 8, 8.

[38] Tversky & Kahneman, *supra* note 34.

[39] See Ariely *supra* note 35.

Bandwagon effect

'The case where an individual will demand more (less) of a commodity at a given price because some or all other individuals in the market also demand more (less) of the commodity'.[40] This bias has similar roots to 'group think' and 'herd effect', where the behaviour of the individual is modified by the behaviour of the collective.[41] It can be observed in multiple contexts, such as when rating the attractiveness of female faces[42] or in Asch-type social conformity experiments.[43]

Recently, its occurrence in social networks and the impact on privacy has been particularly acknowledged. For example, it was observed that 'whether and how much one used Facebook was unequivocally coupled with its diffusion within the global, local, and communal contexts'.[44] We can observe the bandwagon effect anecdotally daily in social networks, where the quick popularity of memes and trending topics show the immense interconnectedness within these networks. In the privacy context, the consequence of the bandwagon effect is that if negligent privacy attitudes are

[40] Harvey Leibenstein, 'Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand' (1950) 64 Q.J. Econ. 183, 190.

[41] Lindsey Levitan & Brad Verhulst, 'Conformity in Groups: The Effects of Others' Views on Expressed Attitudes and Attitude Change' (2016) 38 Political Behav. 277.

[42] Vasily Klucharev Kaisa Hytönen Mark Rijpkema, Ale Smidts & Guillén Fernández, 'Reinforcement Learning Signal Predicts Social Conformity' (2009) 61 Neuron 140.

[43] Rod Bond, 'Group Size and Conformity' (2005) 8 Group Process Intergroup Relat. 331. 'The original Asch study on conformity is recognized as a classical experiment in social psychology. The experiment demonstrated the tendency of participants to conform when under the pressure of a unanimous majority'. Knud Larsen, 'The Asch Conformity Experiment: Replication and Transhistorical Comparisons' (1990) 5 J. Soc. Behav. Pers. 163, 163.

44 Wayne Fu, Jaelen Teo, Seraphina Seng, 'The Bandwagon Effect on Participation in and Use of a Social Networking Site' (2012) 17 First Monday.

trending (*i.e.*, people share photos, videos and personal data without restraint), they will be reproduced among other users of the network as well.[45]

Contrast effect

The contrast effect to which I am referring to here involves visual aspects: exploring the relationship between two objects (or two texts) to reduce readability or to generate a desired impression in the observer / reader. An example is that 'several studies indicate that increased contrast between the text and background results in increased readability'.[46] Therefore if the designer wishes that a certain text is not noticed or well read, the designer should apply low contrast in the colour scheme. A well-known example were the techniques used by search engines in the past to blur the difference between sponsored and organic results.[47] This effect is also commonly observed when the designer maliciously selects low contrast schemes for privacy protective options, to steer data subjects' attention off and reduce the probability that they will choose restrictive options.

Default Effect

[45] It is important to notice that there will always be culture and social norms. This bias, on the other hand, specifically highlights the constant correlation of the individual attribution of value to the collective attribution of value, without a dedicated thought process or contextual considerations.

[46] Robert Moore, Claire Stammerjohan & Robin Coulter, 'Banner Advertiser- Web site Context Congruity and Color Effects on Attention and Attitudes' (2005) 34 J. Advertising 71, 73.

[47] In 2013, the Federal Trade Commission (FTC) sent letters to search engine companies asking for more differentiation between paid and natural results, to avoid consumer deception and violation of Section 5 of the FTC Act. Similar letters had been sent in 2002, with the same argument. At https://www.ftc.gov/news-events/press-releases/2013/06/ftc-consumer-protection-staff-updates-agencys-guidance-search.

The default effect is the observation that the default option 'is chosen more often than expected if it were not labelled the default'.[48] There are various explanations for the 'stickiness' of defaults,[49] among them are the absence of effort needed and the appearance of an implied endorsement by the service provider. The default effect bias has been observed in multiple contexts, among the most famous are organ donation and retirement savings plans.[50]

Defaults are prevalent in the privacy field as well, and it has been shown that individuals will most commonly stick to the default privacy option instead of taking time to think and choose a more suitable alternative.[51] It is known that designers have been benefiting from this bias for a long time by defining privacy invasive defaults.[52] More recently, with the rise of Privacy by Design (PbD) and Data Protection by Design and by Default (DPbDD) and with the requirement for explicit consent,[53] defaults that are patently detrimental to the data subject are not so frequently seen anymore, but it is worth noticing that there is no express ban on defaults, therefore leaving room for companies to imply that certain data is necessary or essential for the performance of the service, with no external accountability.[54]

---

[48] Isaac Dinner, Eric Johnson, Daniel Goldstein, Kaiya Liu, 'Partitioning Default Effects: Why People Choose Not to Choose' (2011) 17 J. Exp. Psy.-Appl. 332, 335.

[49] For a deeper analysis of the properties of defaults, see Omri Ben-Shachar & John Pottow, 'On the Stickiness of Default Rules' (2006) 33 Fla. St. U.L. Rev 651.

[50] For a broader review, see Lauren Willis, 'When Nudges Fail: Slippery Defaults' (2013*)* 80 U. Chi. L. Rev. 1155.

[51] *Id.*

[52] See Matthew Keys, 'A Brief History of Facebook's Ever-changing Privacy Settings', Medium.com (2018), at https://medium.com/@matthewkeys/a-brief-history-of-facebooks-ever-changing-privacy-settings-8167dadd3bd0. On a more positive side, for an analysis on the optimization of access control defaults in online social networks, see Ron Hirschprung, Eran Toch, Hadas Schwartz-Chassidim, Tamir Mendel & Oded Maimon, 'Analyzing and Optimizing Access Control Choice Architectures in Online Social Networks' (2017) 8 ACM Trans. Intell. Syst. Technol., Article 57.

[53] GDPR, Article 4(11). The topics of PbD and DPbDD will be dealt in more detail on Section VII.2 *infra*.

[54] Unless it is a paid service or any specific business model that does not rely on advertising or data.

False Uniqueness Effect

'The tendency to underestimate the extent to which others possess the same beliefs and attributes as oneself or engage in the same behaviours, particularly when these characteristics or behaviours are positive or socially desirable'.[55] This is a type of egocentric bias well studied in psychology.[56] It the health context, it was observed that people who – health wise - behave in desirable ways, would underestimate the correct number of people who would behave like themselves.[57]

This bias' privacy implications can be observed in social networks. Designers can exploit this bias and create DP that lure data subjects into believing and feeling that they are unique, in the spotlight and part of the majority-opinion group.[58] In these digital environments, individuals are influenced by the false uniqueness effect and are urged by peer pressure, the bandwagon effect and by the design of the product to share more or more detailed personal data online.

Framing effect

'The 'framing effect' is observed when a decision maker's risk tolerance (as implied by their choices) is dependent upon how a set of options is described'.[59] More specifically, 'people appear to exhibit a general tendency to be risk seeking when confronted with negatively framed problems

---

[55] American Psychological Association, 'False-Uniqueness Effect', APA Dictionary of Psychology, at https://dictionary.apa.org/false-uniqueness-effect.

[56] See John Chambers, '*Explaining False Uniqueness: Why We are Both Better and Worse Than Others' (2008)* 2 Soc. Personal. Psychol. Compass 878. Linda Perloff & Philip Brickman, 'False Consensus and False Uniqueness: Biases in Perceptions of Similarity' (1982) 4 Academic Psychol. Bull. 475.

[57] See Jerry Suls, Choi Wan & Glenn Sanders, 'False Consensus and False Uniqueness in Estimating the Prevalence of Health-Protective Behaviors' (1988) 18 J. App. Soc. Psy. 66.

[58] Such as by keeping to show notifications of how many times a post was liked, commented, or shared, together with incentivizing messages such as 'well done, your post has already received x interactions'. It makes the individual feel unique and pushes him or her to share more.

[59] Cleotilde Gonzalez, Jason Dana, Hideya Koshino & Marcel Just, 'The Framing Effect and Risky Decisions: Examining Cognitive Functions with fMRI' (2005) 26 J. Econ. Psy. 2.

and risk averse when presented with positively framed problems'.[60] The framing effect has been observed in multiple contexts,[61] and studies have been on ways to reduce its impact.[62] An example of its manifestation is the performance comparison of advertising a yoghurt as 1% fat or 99% fat free. 'The percentage-fat-free format led to stronger endorsements of healthiness than the percentage-fat format,'[63] showing the impact of the framing effect in the perception of healthiness.

In the privacy context, the designer can frame choices and options in a way to elicit a less privacy-protecting option. For example, when asking if the data subject wants to start using a face identification service, the designer can highlight the novelty, the sophistication, and the surprises that the technology can bring, leaving problematic privacy issues as a side comment or something that does not deserve the same level of detail.

Functional fixedness

'The tendency to fixate on the typical use of an object or one of its parts'.[64] This bias has been associated with hindered problem-solving ability or creativity, as some complex challenges require overcoming design fixation (*i.e.,* 'thinking outside the box').[65]

---

[60] *Id.*

[61] See Sammy Almashat, Brian Ayotte, Barry Edelstein & Jeniffer Margrett, 'Framing Effect Debiasing in Medical Decision Making' (2008) 71 Patient Educ. Couns. 102 (2008). Antony Sanford, Nicolas Fay, Andrew Stewart & Linda Moxey, 'Perspective in Statements of Quantity, with Implications for Consumer Psychology' (2002) 13 Psy. Sci. 130.

[62] See Mathieu Cassotti, Marianne Habib, Nicolas Poirel, Ania Aïte, Olivier Houdé & Sylvain Moutier, 'Positive Emotional Context Eliminates the Framing Effect in Decision-Making' (2012) 12 Emotion 926.

[63] Antony Sanford et al., *supra* note 61, at 132.

[64] Tony McCaffrey, 'Innovation Relies on the Obscure: A Key to Overcoming the Classic Problem of Functional Fixedness' (2012*)* 23 Psy. Sci. 215, 217.

[65] *Id.* A simple and general example would be of a person that needs to open a plastic package and keeps looking for scissors, when the car key in her pocket would have been equally useful for this task.

In privacy, designers can maliciously exploit this bias by using colours, symbols, or online functions in an unusual way, therefore misleading the data subject to select the less privacy protective alternative. A simple example would be using a padlock symbol beside an option that does not protect privacy, thus inducing the data subject to error (as he or she would see the function of the padlock as if indicating privacy protection).

Hyperbolic discounting

'People's choices are often inter-temporally inconsistent, for example in the sense that people prefer a larger, later consumption bundle over a smaller, sooner one as long as both are sufficiently distant in time, but change their preference to the smaller, sooner bundle, as both draw near'.[66] In less technical terms, an individual would value $50 now more than $100 in a month.

In a privacy related context, it means that an individual often prefers to use a service immediately, even if it involves risks or possible long term privacy impacts, instead of not using the service now and preserving his or her privacy long term. Privacy scholars have acknowledged this bias in studies about privacy policies and how people prefer to just click 'I accept' and be submitted to aggressive terms so that they can immediately enjoy the service.[67]

Loss aversion

[66] Till Grüne-Yanoff, 'Models of Temporal Discounting 1937–2000: An Interdisciplinary Exchange between Economics and Psychology' (2015) 28 Science in Context 675, 677.

[67] See Alessandro Acquisti & Jens Grossklags, 'Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behavior' (2003) UC Berkeley 2nd Annual Workshop on Economics and Information Security; Alessandro Acquisti, Leslie John, George Loewenstein, 'What Is Privacy Worth?' (2013) 42 J. Legal Studies 249.

'The disutility of giving up an object is greater than the utility associated with acquiring it'.[68] It can be observed in the context of 'free trials,' in which you allow the person to have access to a product or service for a certain, limited period and then, to allow continuity, one or more payments are requested. The loss aversion bias will make the person more likely to agree to pay.

In the privacy context, it would manifest it the sense that 'people are willing to accept more money in exchange for disclosing personal information than they are willing to pay to regain control over the same information'.[69] Meaning that the aversion to losing personal data is bigger than the willingness to pay to acquire the same data.

Optimism bias

'The tendency for people to report that they are less likely than others to experience negative events and more likely than others to experience positive events'.[70] For example, people rate their chances of experiencing negative events as being less than the average person's.[71]

In the privacy context, it represents the tendency to think that one is less likely to suffer any privacy harm,[72] therefore possibly becoming a false reassurance that negligent, careless, and risky online behaviours are harmless.

---

[68] Daniel Kahneman, Jack Knetsch & Richard Thaler, 'Anomalies the Endowment Effect, Loss Aversion, and Status Quo Bias' (1991) 5 J. Econ. Perspect. 193, 194.

[69] Alessandro Acquisti, et al., 'Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online' (2017) 50 ACM Comput. Surv., Article 44, 8.

[70] Marie Helweg-Larsen & James Shepperd, 'Do Moderators of the Optimistic Bias Affect Personal or Target Risk Estimates? A Review of the Literature' (2001) 5 Personality & Social Psychol. Rev. 74, 74.

[71] Adam Harris & Ulrike Hahn, 'Unrealistic Optimism about Future Life Events: A Cautionary Note' (2011) 118 Psy. Rev. 135, 148.

[72] For an empirical study about the manifestation of optimism bias within judgements regarding personal privacy, see Hichang Cho, Jae-Shin Lee & Siyoung Chung, 'Optimistic Bias about Online Privacy Risks: Testing the Moderating Effects of Perceived Controllability and Prior Experience' (2010) 26 Comput. Hum. Behav. 987.

Restraint bias

'A tendency for people to overestimate their capacity for impulse control'.[73] This bias has been more frequently observed in the context of substance addiction, in which an individual inaccurately foresees his ability to restrain himself when exposed to a temptation (*i.e.,* a substance to which he is addicted), leading to overexposure and increase in impulsive behaviour.[74]

In the context of privacy, this bias can be observed in social networks, where three elements are coupled: a) the intrinsic peer pressure and desire to conform to the rules of the group stimulated by the design of online social networks (see above bandwagon effect); b) the repetitive push from service providers asking data subjects to provide more personal data; and c) the susceptibility to addiction related to certain social media uses.[75] Given these three items, the restraint bias will make individuals less likely to correctly estimate their impulse control capabilities, which leads to over-exposure and, in social networks, more personal data sharing.

*

As seen above, cognitive biases are diverse,[76] offering multiple opportunities for malicious designers wishing to manipulate the data subject's decision-making capacities. These biases have

[73] Loran Nordgren, Frenk van Harreveld & Joop van der Pligt, 'The Restraint Bias: How the Illusion of Self-Restraint Promotes Impulsive Behavior' (2009) 20 Psy. Sci. 1523, 1523.

[74] *Id.*

[75] For a deeper analysis of the correlation between social media use, impulsivity, and addiction, see Elisa Wegmann, Silke Müller, Ofir Turel & Matthias Brand, 'Interactions of Impulsivity, General Executive Functions, and Specific Inhibitory Control Explain Symptoms of Social-Networks-Use Disorder: An Experimental Study' (2020) 10 Sci. Rep.; also, Elisa Wegmann & Matthias Brand, 'A Narrative Overview About Psychosocial Characteristics as Risk Factors of a Problematic Social Networks Use' (2019) 6 Curr. Addict. Rep. 402.

[76] Kahneman & Ariely's books cited *supra* at note 35.

been extensively described in cognitive psychology and behavioural economics literature, but their integration with data protection law and the acknowledgement of their potential to support privacy harm is still missing.

The next Section proposes a taxonomy for DP, based on the different ways in which they can negatively impact the decision-making process of the data subject.

V. Taxonomy

Scholars and national organisations have proposed taxonomies for dark patterns in general, including those affecting money, emotions, attention, and data. Among the national organisations are the Norwegian Consumer Council (Forbrukerrådet),[77] and the French National Commission on Informatics and Liberty (CNIL).[78] Among the scholars are Hartzog,[79] Calo,[80] Frisch,[81] Brignull,[82]

[77] Forbrukerrådet, *supra* note 5.

[78] Laboratoire d'Innovation Numérique de la CNIL, 'IP Report: Shaping Choices in the Digital World, From dark patterns to data protection: the influence of UX/UI design on user empowerment' (2019), https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf.

[79] Hartzog, *Privacy's Blueprint, supra* note 2.

[80] Ryan Calo, 'Digital Market Manipulation' (2013) 27 Geo. Wash. L. Rev. 995.

[81] Lothar Frisch, 'Privacy Dark Patterns in Identity Management', *in* Open Identity Summit 2017: Proceedings 93 (Lothar Fritsch, Heiko Roßnagel, Detlef Hühnlein, eds., 2017).

[82] Brignull, *supra* note 3.

Conti & Sobiesk,[83] Nouwens et al.,[84] Bösch et al.,[85] Gray et al.,[86] Mathur et al.,[87] and Zagal et al.[88] Among all the above mentioned, only four (Forbrukerrådet, CNIL, Nouwens et al., and Zagal et al.) have proposed some form of classification system for dark patterns in the privacy field, but only the two national organisations (Forbrukerrådet and CNIL) offered a more overarching taxonomy for dark patterns in privacy.

Forbrukerrådet's categorization focuses on five forms of DP: 'default settings,' 'ease,' 'framing,' 'rewards and punishment,' and 'forced action and timing'. Despite the richness of the examples offered and the useful comparison between Facebook, Google and Microsoft's systems, this taxonomy does not fully portray the spectrum of DP as reflected in the categories that will be presented *infra* and do not show how cognitive biases can be used to form dark patterns or how the data subject is impacted.

CNIL developed a non-exhaustive typology including a broader spectrum of DP, offering a useful and deep overview of the topic. In CNIL's approach, the categories of dark patterns in privacy are enjoy / seduce / lure / complicate / ban. Each of them can either 'push the individual to accept sharing more than what is strictly necessary,' 'influence consent,' 'create friction on data protection actions' or 'divert the individual'. Despite offering no less than 18 examples that have certainly influenced this author to think more broadly about dark patterns, the logic behind their

---

[83] *Supra* note 5.

[84] *Supra* note 5.

[85] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp & Stefan Pfattheicher, 'Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns' (2016) Proceedings on Privacy Enhancing Technologies 237.

[86] Colin Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt & Austin Toombs, 'The Dark (Patterns) Side of UX Design' (2018) CHI 2018.

[87] *Supra* note 5.

[88] José Zagal, Staffan Björk, Chris Lewis, 'Dark Patterns in the Design of Games' (2013), Foundations of Digital Games Conference, at http://www.fdg2013.org/program/papers/paper06_zagal_etal.pdf.

classification does not match the analysis of the present work, which focuses on the ways in which decision making was affected.

Given the unsuitability of existing taxonomies to the present work's analysis, I propose a new classification, whose premises and methodology I will explain now.

The first step in elaborating the taxonomy was finding academic articles and other public sources with categories and examples of dark patterns, to understand which of them were specifically targeting personal data. The public repositories I found were the website https://darkpatterns.org, the Twitter account connected to the same website (@darkpatterns) - where the public can share their examples of dark patterns - and the Tumblr account https://confirmshaming.tumblr.com. The academic papers I used as sources of examples of dark patterns were those cited above, especially the reports by Forbrukerrådet and CNIL, which dealt with privacy related dark patterns.

After gathering a good number of examples of DP, my challenge was to develop categories and groups that matched them and helped me to advance the present analysis, understanding how exactly they impact the decision-making capacities of data subjects. For this purpose, I needed a deeper legal approach regarding decision making and the factors that would make a certain decision to be deemed invalid or unlawful.

In private law, the analysis of the integrity of the decision-making process is made through civil law and the study of consent and consent defects. Given that the jurisdictional focus of the present work is the EU, and the EU does not have a single Civil Code, I resorted to the Principles of European Contract Law (PECL).[89]

[89] Principles of European Contract Law, at https://www.trans-lex.org/400200.

The PECL provided me with a unified view of the EU approach to contractual principles and rules, including those pertaining to the topic of consent and consent defects. Among all the causes of invalidity of contracts,[90] those which are related to the decision-making process of one of the contracting parties are: a) mistake;[91] b) fraud;[92] c) threats;[93] and d) excessive benefit or unfair advantage.[94] In the presence of one of these situations, the damaged party may avoid the contract, according to the additional rules of the same chapter within the PECL.

The four categories above represent a central pillar in the study of the validity of consent within the realm of contract law.[95] Here, I am importing these ideas to the sphere of data protection law, as at this point, I want to expand the knowledge on how malicious designers can impact the data subject's decision-making process, including the validity of the obtained decision / consent.

When transposed to data protection law, the four contractual categories had to be adapted to match the typical DP situations. The idea of 'threat' was translated as 'pressure'.[96] In our typical commercial context, where a controller interacts with a data subject through a public interface design, this is the way a threat - an 'imminent and serious threat of an act… which it is wrongful to use as a means to obtain the conclusion of the contract'[97] would manifest (where the act would

[90] PECL, Chapter 4.

[91] PECL, Article 4:103.

[92] PECL, Article 4:107.

[93] PECL, Article 4:108.

[94] PECL, Article 4:109.

[95] For further study on the elements and validity of consent, see Nancy Kim, Relative Consent and Contract Law, 18 Nev. L.J. 165 (2017).

[96] According to Article 4:108 of the PECL, 'A party may avoid a contract when it has been led to conclude it by the other party's imminent and serious threat of an act: (a) which is wrongful in itself, or (b) which it is wrongful to use as a means to obtain the conclusion of the contract, unless in the circumstances the first party had a reasonable alternative'.

[97] *Id.*

be the impediment of the data subject to use the product or service). Despite having a softer meaning than the category in the civil law context, a pressure to share more or more in-depth data made by a data controller to a data subject in an environment of extreme asymmetry directly harms the decision-making process involved and negatively impacts privacy.

Second, the idea of 'excessive benefit or unfair advantage' was translated as 'hindrance'.[98] The essential idea is that data controllers are in a position of extreme advantage over data subjects, for having knowledge on data practices that is not usually available to data subjects and access to skilful designers that can manipulate the interface as wished. Controllers are aware of this imbalance and use it for their own benefit, in this case by hindering and making it difficult for data subjects to have access to and to express their privacy preferences in a usable and friendly way. Given data subjects' inexperience and general inability to navigate complex, misleading, and even tricky settings, they are hindered, and their decision making is negatively impacted.

Third, the idea of 'mistake' was translated as 'mislead,' [99] which is the essential action of the data controller regarding the data subject. However, the actions in the civil law and the data protection law context were very similar, as the aim of the controller, through design and psychology related means, is to confuse and make the data subject commit a mistake of fact, therefore acting in a way that is not the most beneficial for his or her privacy.

---

[98] According to the Article 4:109 of the PECL, '(1) A party may avoid a contract if, at the time of the conclusion of the contract: (a) it was dependent on or had a relationship of trust with the other party, was in economic distress or had urgent needs, was improvident, ignorant, inexperienced or lacking in bargaining skill, and (b) the other party knew or ought to have known of this and, given the circumstances and purpose of the contract, took advantage of the first party's situation in a way which was grossly unfair or took an excessive benefit'.

[99] According to Article 4:103 of the PECL, '(1) A party may avoid a contract for mistake of fact or law existing when the contract was concluded if: (a)(i) the mistake was caused by information given by the other party; or (ii) the other party knew or ought to have known of the mistake and it was contrary to good faith and fair dealing to leave the mistaken party in error; ..., and (b) the other party knew or ought to have known that the mistaken party, had it known the truth, would not have entered the contract or would have done so only on fundamentally different terms'.

Lastly, the concept of 'fraud' was translated as 'misrepresent'.[100] Here, in the data protection context, the controller benefits from a lack of accountability or stricter regulation to misrepresent facts to the data subject, such as the necessity of certain data to the performance of a task or the potential experience improvement after the data subject shares more or more in-depth data.

Below I present the four categories and their examples. Regarding the titles given to each example, they aim at summarising the characteristics of the respective DP and at facilitating comprehension. The overall goal of the proposed taxonomy is to help us to better understand and address the legal challenges behind DP, especially how designers negatively affect data subjects' decision-making process.

A) Pressure

Description: pressuring the data subject to share more (or more in-depth) than intended personal data to continue using a product or service.

Examples:

1. pressure to allow permissions: not allowing the use of a service unless multiple permissions are conceded (i.e., access to the camera, microphone, GPS, contact list and storage in a photo app).

2. *pressure to receive marketing:* requiring the data subject to check the box 'receive marketing offers per email' to conclude a purchase or sign up.

[100] According to Article 4:107 of the PECL, '(1) A party may avoid a contract when it has been led to conclude it by the other party's fraudulent representation, whether by words or conduct, or fraudulent non-disclosure of any information which in accordance with good faith and fair dealing it should have disclosed'.

3- *pressure to share:* requiring the data subject to reveal personal data to other users to use the service (*i.e.,* a running app that automatically shares geolocation with other users, not allowing the option to hide it).

4- *pressure to confirm:* using negative persuasive language to make the data subject feel bad about not accepting a certain modality of data collection (*i.e.,* a pop up window in an e-commerce site that shows two buttons, one saying 'yes I want to create a login to gain access to the best offers' and the other saying 'no, I do not like offers, I prefer paying a higher price').

B) Hinder

Description: delaying, hiding, or making it difficult for the data subject to adopt privacy protective actions.

Examples:

1- *difficult rejection:* consent pop ups that offer the options 'accept all' or 'more information'. If the data subject clicks on the latter, a page with dozens (sometimes hundreds) of green switches appears, containing every entity that currently collects data. Instead of having the option to 'reject all,' the data subject must go through each switch and turn it off or exit the site.

2- *difficult settings:* privacy settings that are built in a complex and multi-layered way, containing various links, explanations, sub-menus, third party URLs and other resources to make it difficult (and very unlikely) for the data subject to read and understand.

3- *difficult deletion:* making it difficult (*i.e.,* having to correctly navigate through multiple drop-down choices) or inconvenient (*i.e.,* requiring the data subject to speak with a representative through the phone) to delete the account, occasioning continuous data collection.

4- *privacy invasive defaults:* privacy invasive defaults (*i.e.,* a video social network app that shares videos publicly by default).

5- *hidden settings:* hiding privacy settings in a place that is not intuitive or that the data subject needs to correctly navigate through multiple drop-down choices to arrive.

C) Mislead

Description: using language, forms, and interface elements to mislead the data subject whilst taking privacy related actions.

Examples:

1- *ambiguity:* using confusing language such as in a pop-up stating 'do not share my data with third parties' with the options 'yes' and 'no'. It is unclear if 'yes' means 'share' or 'do not share'.

2- *double negative:* using double negatives as in 'I do not want to deny sharing my data with third parties'. It is difficult for the data subject to understand if he or she should check or uncheck this option in order not to share the data with third parties.

3- *twist:* using colours and symbols in a way that is different than what is customary to misguide the data subject (*i.e.,* using green in the 'deny' and red in 'accept' button or using the symbol of a padlock beside options that are privacy invasive).

4- *obfuscation:* adding elements to a privacy setting that are not connected to privacy, so that privacy information gets obfuscated.

5- *bad visibility:* using badly contrasted, light colours or small fonts to make privacy protecting options less visible.

6- *framing:* describing privacy invasive features in a positive way to distract the data subject from the downsides.

D) Misrepresent

Description: misrepresenting facts to induce data subjects to share more or (more in-depth) personal data than intended.

Examples:

1- False necessity: stating that collecting certain types of data is legally necessary or required for the performance of a task or for system functioning when they are not.

2- False experience improvement: stating that collecting certain types of data is necessary to bring a better experience or a better quality of service when there is no difference to the data subject

*

After discussing cognitive biases and the proposed taxonomy, in the next Section, I turn to data protection law and analyse the current legal status of DP, especially regarding the GDPR.

VI. DP and the GDPR

In this Section, my goal is to understand the legal status of DP regarding the GDPR. More specifically, I inquire: a) if DP fit any of the lawful grounds to collect and process data and b) if there is any rule or principle prohibiting DP.

The GDPR has various articles dealing with how, why, when, by whom, and for how long personal data may be processed. Moreover, the principles of data processing such as transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality and accountability are well defined and explained by the GDPR's recitals.[101] They aim at providing guidelines on how personal data should be handled by processors and controllers during the processing phase.

DP, however, do not happen in the processing phase. They happen during the collection of personal data, in the design interface, which we can call the pre-processing phase. The discussion in this Section is, therefore, whether the GDPR has mechanisms to curb and control manipulative behaviour from the controller that happens before data is collected – and which leads to it.

The rules that apply to the collecting phase are found in Article 6 of the GDPR. Additionally, Article 25 brings the concept of Data Protection by Design and by Default, which can potentially be helpful to set boundaries to design practices. Lastly, the fairness principle, although not defined in the GDPR, is mentioned multiple times by it, and per its philosophical and traditional legal meaning could serve as a key to restrain DP. I review each of these options in the following paragraphs.

VI.1 Lawfulness of Data Processing

[101] GDPR, Article 5(1)(a) et seq.

Article 6 of the GDPR establishes six situations in which data can be lawfully collected and processed. They can be succinctly stated as consent, performance of a contract, legal obligation, vital interest, public interest, legitimate interest.[102]

Among them, consent is the only option that could be affected by DP, as it comprises situations in which the decision-making capacity of the data subject serves as the legitimising factor to data collection. The other options are either related to public matters (legal obligation, vital interest, public interest task[103]) or refer to situations which involve a private interest from the data controller, but which the active participation of the data subject is not required (performance of a contract and legitimate interest).[104]

Regarding the rules for consent, Article 6(1)(a) establishes that 'processing shall be lawful only if and to the extent that at least one of the following applies: the data subject has given consent to the processing of his or her personal data for one or more specific purposes'. To be lawful, consent must be 'freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her'.[105]

To better understand what these requirements mean, I begin with the condition that consent must be 'freely given'. According to the GDPR: (a) the data subject should have genuine and free choice;[106] (b) there should not be a clear imbalance between the data subject and the data

[102] GDPR, Article 6(1).

[103] GDPR, Article 6(1)(c), (d) and (e).

[104] GDPR, Article 6(1)(b) and (f).

[105] GDPR, Article 4(11).

[106] Recital 42 of the GDPR states that '(…) consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment'.

controller;[107] (c) consent should not be a condition to the performance of a contract;[108] and (d) the data subject should be able to withdraw consent without detriment.[109]

The items that are most relevant to the present analysis are (a), (b) and (c) *supra*. As we saw, DP rely on cognitive biases to manipulate data subject to share personal data. If DP negatively interfere with the data subject's decision-making process, consent given under their influence cannot be said to be genuine and represent free choice, such as required by item (a) as the elements that have been manipulated by the controller are not under the awareness of the data subject.

Moreover, there is an important cognitive imbalance between data controllers and data subjects, contrary to what item (b) foresees. Controllers and their designers have the technical ability and the *know-how* to steer the decision-making process of the data subject in a desired direction. On the other hand, data subjects are usually not aware about the topic of cognitive biases and how they can be systematically manipulated by service providers.

Additionally, it is also important to comment that some of the DP infringe item (c), such as '*pressure to receive marketing:* requiring the data subject to check the box 'receive marketing offers per email' to conclude a purchase or sign up,' which is the second example I gave in the taxonomy regarding category 'Pressure'. In these cases, consent would be infringing specific GDPR tenets.

---

[107] Recital 43 of the GDPR puts that 'in order to ensure that consent is freely given, consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation.(…)'

[108] Article 7(4) of that GDPR states that 'when assessing whether consent is freely given, utmost account shall be taken of whether, *inter alia*, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract'.

[109] Recital 42 of the GDPR states that '(…) consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment'.

A second requirement for consent that is relevant to the present work is to be 'informed'. According to the GDPR, 'for consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended'.[110] DP seem not to be affected by this requirement. Controllers usually inform, at some point, data subjects about their identity and the general purposes of the processing for which the personal data are intended and, in another moment, they implement DP to collect the desired data. As data subjects tend not to read written declarations, especially long and complex ones, this information will not be properly noticed, and controllers will have a free pass to exploit cognitive biases through the manipulation of interface design.

Given this analysis, DP are incompatible with the various consent requirements established by the GDPR. However, the legal discussion of dark patterns in the privacy realm is still sparse and there are currently no consolidated definition, classification, and criteria to clearly determine what practices fall in the DP basket and which do not, making it difficult to single them out as practices that infringe the GDPR. Moreover, without a more thorough and constant debate between researchers, lawmakers, NGOs, privacy officers and privacy advocates on the characteristics of dark patterns in privacy and the way they infringe the GDPR, it will be difficult to enforce or implement a system of prevention of DP.

This work aims to answer this gap. It proposes a taxonomy and a legal definition for dark patterns in privacy and offers a thorough analysis of their characteristics and main manifestations. With increased awareness about practices that should be classified as DP - and therefore deemed illegal in case they are deployed to obtain consent – hopefully courts, lawmakers, privacy

[110] Recital 42 of the GDPR.

professionals and the public at large will move towards a more stringent view on the limits of lawful consent.

Although not at the legislative level yet, the European Data Protection Supervisor (EDPS) has started to raise the topic of issues involving legal design and dark patterns, denoting the Member States' interest in curbing these types of practices.[111] As we saw in previous Sections, CNIL in France and the Forbrukerrådet in Norway prepared detailed reports on the topic, joining the efforts to raise awareness, and perhaps paving the way for more interventions to come.

In the United States, the California Privacy Rights Act (CPRA) recently amended the California Consumer Privacy Act (CCPA),[112] adding a specific rule that made consent obtained through dark patterns invalid. It established that:

> (h) 'Consent' means any freely given, specific, informed, and unambiguous indication of the consumer's wishes by which the consumer, or the consumer's legal guardian, a person who has power of attorney, or a person acting as a conservator for the consumer, including by a statement or by a clear affirmative action, signifies agreement to the processing of personal information relating to the consumer for a narrowly defined particular purpose. Acceptance of a general or broad terms of use, or similar document, that contains descriptions of personal information processing along with other, unrelated information, does not constitute consent. Hovering over, muting, pausing, or closing a given piece of content does

[111] Giovanni Buttarelli, 'Legal Design Roundtable' (27 April 2019), at https://edps.europa.eu/sites/edp/files/publication/19-04-27_dark_patterns_en.pdf

[112] Cal. Civ. Code § 1798.140(h).

> not constitute consent. Likewise, agreement obtained through use of dark patterns does not constitute consent.[113]

It is the first time that a data protection legislation expressly states that an agreement obtained through dark patterns does not constitute consent. The CPRA represents an important advancement for privacy law, first because the terminology of dark patterns – which stems from interface design – is being imported into the privacy realm and helping to better shape and identify design practices that can negatively impact privacy rights. Second, California hosts an important tech hub and where some of the tech giants' headquarters are located, therefore legislation affecting them will likely have a wider national and global effect.

It must be pointed out, however, that this is the first step only. Further steps on the identification and understanding of practices that qualify as dark patterns must be carried on so that the Law is effective, and enforcement can work.

Also in the United States, Senators Mark Warner and Deb Fischer introduced a bipartisan federal legislation called Deceptive Experiences to Online Users Reduction (DETOUR) Act. According to the official description of this Bill, it:

> prohibits large online operators from manipulating their product to mislead consumers into providing personal information or giving consent. The bill further prohibits these operators from studying the behavioural patterns of subsets of users

[113] *Id.*

> without first obtaining informed consent, and it prohibits designing online products that lead to compulsive usage by children.[114]

This bill has a much broader scope, widening the practices that are considered dark patterns and being more detailed regarding enforcement practices. For example, it establishes among the prohibited conducts that:

> (A) to design, modify, or manipulate a user interface with the purpose or substantial effect of obscuring, subverting, or impairing user autonomy, decision-making, or choice to obtain consent or user data;
>
> (B) to subdivide or segment consumers of online services into groups for the purposes of behavioural or psychological experiments or studies, except with the informed consent of each user involved; or
>
> (C) to design, modify, or manipulate a user interface on a website or online service, or portion thereof, that is directed to an individual under the age of 13, with the purpose or substantial effect of cultivating compulsive usage, including video auto-play functions initiated without the consent of a user.[115]

Despite not being fully approved yet, it was introduced as a federal legislation, therefore having the potential for a larger impact. Moreover, it is interesting to see that, at least in the United

[114] S.1084 — 116th Congress (2019-2020).

[115] S.1084 — 116th Congress (2019-2020), Section 3(a)1.

States, the discussions about a legal framework for dark patterns advance fast and take into consideration their broad and interdisciplinary manifestations.

VI.2 Privacy by Design

Another concept that might help to deal with the legality of DP is Privacy by Design (PbD),[116] as advanced by Ann Cavoukian. It was absorbed by the GDPR as Data Protection by Design and by Default (DPbDD), and it is expressed in Article 25:

> Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.[117]

---

[116] The seven foundational principles of Privacy by Design, as idealised by Ann Cavoukian, the former Information and Privacy Commissioner for Ontario, Canada, are: '1. Proactive not Reactive, Preventative not Remedial; 2. Privacy as the Default Setting; 3. Privacy Embedded into Design; 4. Full Functionality – Positive-Sum, not Zero-Sum; 5. End-to-End Security – Full Lifecycle Protection; 6. Visibility and Transparency – Keep it Open; and 7. Respect for User Privacy – Keep it User-Centric'. Ann Cavoukian, 'Privacy by Design - The 7 Foundational Principles' - *Implementation and Mapping of Fair Information Practices*, at https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf.

[117] GDPR, Article 25.

Recital 78 of the GDPR specifies what it means with 'appropriate technical and organisational measures:'

> (…) such measures could consist, inter alia, of minimising the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features. When developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations.

Article 25 and Recital 78, when approaching DPbDD, focus mostly on the processing phase and measures that could be taken to reduce harm in this context, such as data minimization and pseudonymization. As we saw, DP happen in the moment of collection of personal data and within the design interface. Without a clearer connection between DPbDD tools and possible measures to be undertaken by controllers to avoid DP, it is unlikely that this framework might help curb dark patterns in privacy.

PbD / DPbDD are broad and overarching approaches that can help organisations to design products and services in a more privacy-protective way. [118] However, precisely because they are not technology-specific, for them to be effective in the context of DP there needs to be further developments regarding what measures, how and when should be applied by controllers. For that, again, hopefully this work can be useful by providing a legal definition, taxonomy, and a framework for DP, perhaps paving the way to further developments that help better operate PbD in the context of dark patterns in privacy.

VI.3 The Fairness Principle

The GDPR has numerous principles regarding the processing of personal data, such as lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, and accountability.[119] All these principles are clarified either by the GDPR Articles or by its Recitals, with one glaring exception: the fairness principle.[120]

Fairness is a complex concept. From H. L. A. Hart[121] to John Rawls,[122] different authors have attempted to frame and contextualise it. In the legal realm, it has been linked to justice and

[118] 'Privacy by design' consists of principles that can be applied from the onset of systems development to mitigate privacy concerns and achieve data protection compliance. However, these principles remain vague and leave many open questions about their application when engineering systems'.- Seda Gürses, Carmela Troncoso & Claudia Diaz, 'Engineering Privacy by Design' (2011) Proceedings of the 4th International Conference on Computers, Privacy & Data Protection, 1.

[119] GDPR, Article 5(1)(a) et seq.

[120] Fairness is mentioned in Articles 5, 6, 13, 14 and 40 and in Recitals 4, 39, 42, 45, 60, 71 and 129.

[121] H.L.A. Hart, 'Are There Any Natural Rights?' (1955) 64 Philos. Rev. 175.

[122] John Rawls, '*Justice as Fairness' (1958)* 67 Philos. Rev. 164.

equality,[123] absence of discrimination[124] and reasonableness,[125] and various fields have connected the idea of fairness to specific meanings and standards.[126] Within EU consumer protection law, for example, the Unfair Commercial Practices Directive (UCPD) has its own definition of what are unfair commercial practices,[127] as well as a list of commercial practices which are, in all circumstances, considered unfair.[128] The EU Unfair Contract Terms Directive (UCTD) follows the same model.[129]

Within data protection law, there are no consolidated views or rules on what unfair practices are. According to the United Kingdom's (UK) Information Commissioner's Office (ICO):

> [i]n general, fairness means that you should only handle personal data in ways that people would reasonably expect and not use it in ways that have unjustified adverse

[123] *Id.*

[124] Belinda Smith, 'Fair and Equal in the World of Work: Two Significant Federal Developments in Discrimination Law' (2010) 23 AJLL 199.

[125] Federico Ortino, 'From "Non-Discrimination" to "Reasonablenes": A Paradigm Shift in International Economic Law?' (2005) Jean Monnet Working Paper n. *01/05*.

[126] Within procedural law, the idea of procedural fairness is central to the validity and quality of the outcome. See Joel Brockner, 'Making Sense of Procedural Fairness: How High Procedural Fairness Can Reduce or Heighten the Influence of Outcome Favorability' (2002) 27 Acad. of Manage Rev. 58. Within competition law, economic fairness is central to recent debates in the field, see Alfonso Lamadrid de Pablo, 'Competition Law as Fairness' (2017) 8 JECL & Pract. 147.

[127] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'). *OJ L 149, 11.6.2005, p. 22–39*. Article 5.2 (a) & (b).

[128] *Id.* at Annex I.

[129] Council Directive 93/13/EEC of 5 April 1993 on Unfair Terms in Consumer Contracts. *OJ L 95, 21.4.1993, p. 29–34*. It has a definition of what is an unfair term on its Article 3 and a list of terms which are regarded as unfair in its Annex.

> effects on them (…). Assessing whether you are processing information fairly depends partly on how you obtain it. In particular, if anyone is deceived or misled when the personal data is obtained, then this is unlikely to be fair. In order to assess whether or not you are processing personal data fairly, you must consider more generally how it affects the interests of the people concerned – as a group and individually.[130]

This understanding is relevant to the present analysis because it allows us to integrate the idea of interface design manipulation with fairness in data protection. The ICO's definition is based on a philosophical interpretation and does not necessarily have a GDPR legal basis to rely on. For the ICO, the idea of fairness involves handling personal data in a way that: (a) respects the reasonable expectations of data subjects; (b) does not bring adverse effects to data subjects; (c) does not involve deception or misleading the data subject in the moment of collection of personal data; and (d) considers how it affects individual and collective interests of the concerned data subjects.

Based on this view, DP breach the principle of fairness, for cumulatively: (a) not respecting reasonable expectations of data subjects; (b) bringing negative adverse effects to data subjects' decision-making ability; (c) involving manipulation and exploitation of cognitive biases at the moment of data collection; and (d) negatively affecting data subjects' privacy rights.

Further discussions and doctrinal works on fairness in data protection are needed to clarify what is its meaning, correlated rights and whether it can be said to outlaw DP or not. In this author's

---

[130] Information Commissioner's Office, 'Principle (a): Lawfulness, Fairness and Transparency', at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency.

view, fairness is a key concept to curb dark patterns in privacy. It allows us to focus on the important elements raised by the ICO, which are the reasonable expectations of data subjects, the practical effects on data subjects, the existence of deception at the moment of collection of personal data, and the individual and collective rights affected by these practices. All these elements are not connected to a specific technology or to a specific rule on how data must be collected. They embrace the idea that it does not matter the shape or form of the technology or practice, the focus should be on the data subject: an individual with expectations, cognitive biases (therefore manipulable) and dignitary rights.

The discussion of fairness is central to the present work, and it is important to notice that the GDPR is still vague regarding the meaning of fairness, and despite constantly pairing it with transparency and lawfulness, it does not provide further guidance on how it should be applied or enforced in the real world. Advancements on the regulation of DP and other malicious online practices are dependent on a further development of the idea of fairness in data protection, which this work hopes to support.

*

In this Section, I analysed whether DP could be tackled by: a) the GDPR rules on lawfulness of data processing; b) PbD/DPbDD; and c) the fairness principle. Regarding 'a,' despite DP being incompatible with the various consent requirements established by the GDPR, the framework to define, characterise and repel DP still needs to be further developed so that they can be routinely identified and curbed. Legislative initiatives from the United States might serve as inspirations of practical steps to move forward. Regarding 'b,' these concepts do not seem to currently fit as mechanisms to stop the spread of DP, as more concrete standards and guidelines would need to be put in place – and perhaps this determinacy would go against the nature of PbD

and DPbDD themselves. Lastly, regarding 'c,' there is currently no clarification on what EU data protection law's approach to the fairness principle is. The GDPR does not define it and official documents from EU data protection supervisors do not clarify its meaning or role within data protection law. I have argued that further research and more practical steps on the unpacking and implementation of the fairness principle within data protection law are needed and have great potential to control DP and improve the level of protection afforded to data subjects.

In the next Section, I discuss a necessary change in the decision-making paradigm that supports data protection legislation. With the law reflecting a more accurate depiction of the data subject's traits and behaviour, hopefully legal principles and rules can be more effective in curbing novel malicious practices and protecting privacy.

VII. *Homo economicus* vs *Homo manipulable*

By applying cognitive psychology theories to the traditional economic model of rational decision making, as reflected by the *Homo economicus*,[131] the idea of bounded rationality gained strength.[132] It conveys that cognitive limitations such as biases affect the decision-making process and render the rational-agent model unrealistic.[133]

[131] The rational-agent model in the traditional economic theory is 'personified' as the *Homo economicus* – the economic man. He has determined preferences, is always self-interested, outcome oriented and 'has a rate of time preference that allows him to allocate consumption over time in a consistent manner, reflecting his welfare and his concern for the welfare of future generations'. - Herbert Gintis, 'Beyond Homo economicus: Evidence from Experimental Economics' (2000) 35 Ecol. Econ. 311, 312.

[132] 'Bounded rationality is a concept proposed by Herbert Simon that challenges the notion of human rationality as implied by the concept of *Homo economicus*. Rationality is bounded because there are limits to our thinking capacity, available information, and time (Simon, 1982)', Behavioral Economics, at https://www.behavioraleconomics.com/mini-encyclopedia-of-be/bounded-rationality.

[133] Daniel Kahneman, 'Maps of Bounded Rationality: Psychology for Behavioral Economics' (2003) 93 Am. Econ. Rev. 1449, 1449.

Bounded rationality is one of the basic assumptions of behavioural economics, which brings the understanding of cognitive psychology to the realm of economics.[134] Gintis proposed that the '*Homo economicus* is replaced by a more accurate model of individual choice and strategic interaction'[135] and Thaler forecasted that *Homo economicus* will evolve into *Homo sapiens*.[136]

The understanding of cognitive biases, as explained in Section IV *supra*, and the rejection of the *Homo economicus* model are of utmost importance to this work. As we saw, one of the main characteristics of DP is the designer's reliance on cognitive biases: the designer maliciously manipulates data subjects through the exploitation of cognitive biases, against which the data subject has little or no control. The vulnerability of the data subject and his or her incapacity to calculate, plan and easily avoid DP make the *Homo economicus* model strongly implausible – and even wrong – within the data protection realm.

However, currently, data protection law – and here I focus on the EU framework – does not expressly reject the *Homo economicus* model. Taking as an example the GDPR, despite its numerous principles and provisions to guarantee data subjects' rights, it allows consent as one of the possibilities for lawful processing of data, but it does not prevent manipulation and the exploitation of biases through the interface design, where the data subject first interacts with the data controller.

Despite not expressly endorsing the *Homo economicus* model, the GDPR and the EU data protection framework as a whole do not deny it or present a different approach that clarifies that

---

[134] '[b]ehavioral economics seeks to use psychology to inform economics, while maintaining the emphasis on mathematical structure and explanation of field data that distinguish economics from other social sciences'. Colin Camerer, 'Behavioral economics: Reunifying psychology and economics' (1999) 96 Proc. Natl. Acad. Sci. USA 10575, 10575.

[135] Gintis, *supra* note 131, at 320.

[136] Richard Thaler, 'From Homo Economicus to Homo Sapiens' (2000) 14 J. Econ. Perspectiv. 133, 140.

cognitive biases exist, and humans are vulnerable to manipulation online. There is indeed a protective framework which includes principles,[137] data subjects' rights,[138] the preference for opt in instead of opt[139] out and additional rules for consent.[140] However, despite allowing consent - a typical autonomy supporting mechanism – the GDPR concomitantly does not protect the data subject against possible downsides or traps associated with it,[141] therefore allowing DP to flourish.

To curb DP, many possibilities could be thought of, including ideas involving technology, the market, public awareness, and law. This work is specifically concerned with the legal path, given the incompatibility between what happens on the ground – the various DP – and the provisions afforded by the law, especially in the EU framework, which are insufficient to tackle the presented challenges.

My goal is to rethink the GDPR and EU Data Protection law in a way that correctly assigns liability and accountability to the deployers of unfair practices. For that, first a new paradigm must be put in place, one that recognizes that data subjects are manipulable and that absent rules and principles that curb malicious manipulation, data controllers are likely to maliciously exploit biases.[142]

The abuse of cognitive biases that takes place online is especially worrisome, as the negative impact on individuals is aggravated by the structure of the web and of the technology

[137] GDPR, Articles 5-11.

[138] GDPR, Articles 12-23.

[139] GDPR, Recital 32 and Article 4(11).

[140] GDPR, Articles 7 & 8.

[141] Such as malicious manipulation.

[142] Data controllers are companies in the pursuit of profit. Exploiting biases help data collection and more data can directly translate into profit through the advertising market. Therefore, absent constraints, data controllers will engage in the exploitation of biases.

market. Online, with only one click – one slip of attention – loads of personal data can be transferred to hundreds of data-thirsty third parties. All invisibly to the data subject's eye, but under the surveillance of dozens or hundreds of skilled engineers and marketers through their analytics dashboards.

Currently, the GDPR does not expressly acknowledge the existence of cognitive biases or the assumptions from cognitive psychology and behavioural economics. It assigns choice to the data subject and fails to foresee, prevent, and correct all the expected errors in judgement that will be associated with this choice.

It might be said that there are other safeguards to protect privacy and that choice is just part of the data protection framework offered by the GDPR. However, choice is an opening door to the data controller, giving it leeway to have access to the data. Additional rules modulate how, when and by whom these data can be used, but the threshold of protection will be lower if the controller can start with a greater amount of data that was maliciously extracted from the data subject.

To curb DP and to properly safeguard data subjects, data protection law must embrace the *Homo manipulable*, an individual who is affected by multiple cognitive biases and who is vulnerable to malicious agents, especially in the online context. Lawmakers, courts, privacy advocates and researchers have the task to produce content, awareness and legal change that updates data protection law and synchronises it with a more accurate depiction of us – the data subjects.

VIII. Conclusion

DP are defined in this work as *user interface design choices that manipulate the data subject's decision-making process in a way detrimental to his or her privacy and beneficial to the service provider.*

In this Article, I presented the premises, privacy theories, the cognitive biases involved, a proposed taxonomy, the legal status of DP within EU data protection law and the decision-making paradigm that inspires it. My purpose was to understand (a) how DP negatively impact the decision-making process of data subjects; and (b) why this is a source of unfairness and should be curbed by the legislation.

In a nutshell, the GDPR is silent about the exploitation of cognitive biases, manipulative interface designs and negative interferences in the decision-making process. It also misses the opportunity to unpack the fairness principle and to present occasions in which unfair practices could spread within the data protection realm, for example through design.

When dealing with consent by the data subject, the acknowledgment of cognitive biases and the preventive and corrective measures necessary to mitigate them is indispensable. Human choice will never be perfect, however, in the online environment, any asymmetry is aggravated by the immense processing and analytical powers owned by technology companies. Cognitive biases must be taken as granted, and any choice or interaction framework must overcome them.

As we saw, currently, the GDPR focuses strongly on the processing phase, granting broad protection to the data subject there. Regarding the pre-processing phase, however, it is practically unregulated, leaving the data subject vulnerable to malicious parties.

Interface design practices cannot be ignored. They must embody data protection values and, most important, fairness. The interface must be a place where the asymmetry between data subjects and controllers is reduced, not amplified.

To curb DP, fairness is a central concept, as it reflects the need to balance the asymmetries between controllers and data subjects. The GDPR refers to fairness multiple times, yet, has no definition thereof, either specificity or enforceability for the concept.[143] The way to advance data protection law is by unpacking the idea of fairness, so that it can encompass the right of fair decision making and fair interface design in privacy to data subjects.

[143] The GDPR refers to fairness or fair processing in five articles (5, 6, 13, 14, 40) and five recitals (38, 42, 45, 60, 71), however, there is no definition or specification of what it means.