HAL
open science

Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors

Sanju Ahuja, Johanna Gunawan, Nataliia Bielova, Cristiana Teixeira Santos

To cite this version:

Sanju Ahuja, Johanna Gunawan, Nataliia Bielova, Cristiana Teixeira Santos. Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors. 2025. ⟨hal-05301214v1⟩

HAL Id: hal-05301214

https://hal.science/hal-05301214v1

Preprint submitted on 7 Oct 2025 (v1), last revised 3 Mar 2026 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors

SANJU AHUJA, Inria Centre at Université Côte d'Azur, France
JOHANNA GUNAWAN, Maastricht University, Netherlands
NATALIIA BIELOVA, Inria Centre at Université Côte d'Azur, France
CRISTIANA TEIXEIRA SANTOS, Utrecht University, Netherlands

Dark patterns are design practices that undermine users' ability to make autonomous and informed choices in digital experiences. The EU Digital Services Act (DSA) seeks to protect users from such designs and their effects, with Article 25 prohibiting three autonomy violation types: *deception, manipulation* and *distortion/impairment*. Demonstrating such regulatory violations, however, requires design-oriented reasoning necessary to articulate why an observed design practice constitutes a specific autonomy violation type. This paper maps 59 known dark patterns onto the three autonomy violation types from the DSA and identifies eight new design factors which can help determine when a dark pattern violates autonomy. Our mapping of dark patterns to autonomy violations grounds ongoing regulatory debates in design while opening pathways for translational research that reimagines how HCI engages with the governance of design practices.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Human computer interaction (HCI)**.

Additional Key Words and Phrases: dark patterns, deceptive design, autonomy violations, Digital Services Act

ACM Reference Format:
Sanju Ahuja, Johanna Gunawan, Nataliia Bielova, and Cristiana Teixeira Santos. 2025. Dark Patterns and the EU Digital Services Act: Mapping Autonomy Violations and Design Factors. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 41 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Commonly referred to as "dark patterns", [1] deceptive, manipulative, or coercive design practices are used in digital systems to increase revenue, maximize user engagement, and collect personal data. These practices impact user autonomy and influence users into making decisions they did not intend, or decisions against their best interests [4, 5, 44, 63]. Dark patterns are highly pervasive: 95% of mobile apps [25] and over 10% of global e-commerce websites [62] feature

[1]Inspired by the recent workshop on dark patterns at ACM CHI 2024 [43], we adapt the workshop's statement on the usage of the term "dark patterns". We use this term to connect our efforts to prior scholarship across domains and legal codified concepts, recognizing that other terms, notably "deceptive design" or "manipulative user interface design," are also used but do not yet encapsulate the broad remit of practices or concepts from academic or regulatory perspectives. Though we retain this term in the absence of a fully-encompassing replacement, we direct readers to the knowledge that ACM Diversity and Inclusion Council now includes the term "dark patterns" on a list of problematic terms (https://www.acm.org/diversity-inclusion/words-matter).

Authors' Contact Information: Sanju Ahuja, sanju.ahuja@inria.fr, Inria Centre at Université Côte d'Azur, Sophia Antipolis Cedex, France; Johanna Gunawan, johanna.gunawan@maastrichtuniversity.nl, Maastricht University, Maastricht, Netherlands; Nataliia Bielova, nataliia.bielova@inria.fr, Inria Centre at Université Côte d'Azur, Sophia Antipolis Cedex, France; Cristiana Teixeira Santos, c.teixeirasantos@uu.nl, Utrecht University, Utrecht, Netherlands.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

at least one dark pattern. Similarly, an EU Commission report [59] found evidence of dark patterns in 97% of the 200 most popular apps and websites across the EU. Such patterns appear in many different contexts: in the web and mobile apps [25, 49], e-commerce [62, 67], social media [64, 65, 76], privacy interfaces [11, 46, 50], games [36, 84], voice interfaces [70] and video streaming platforms [17]. More recent research has analysed and measured the immediate effect of dark patterns on end users' behavior within digital systems, such as within consent banners [9, 12, 35, 48, 51, 68], video-on-demand platforms [77] or even specific interfaces for children [78]. While the body of literature on dark patterns is continuously growing [42], various definitions of dark patterns types have been proposed in various contexts. To standardise these definitions into a unified body of knowledge, Gray et al. [47] harmonized multiple taxonomies of dark patterns spanning those from regulators [18, 31, 37, 59, 69], academia [11, 44, 58, 62], and industry practice [15, 16] into a unified hierarchical ontology of dark patterns [47]. This ontology will be considered as a fundamental body of knowledge in our paper, even though it is possible to extend the existing ontology with new types of dark pattern practices.

As regulators around the world increasingly draft policies for digital technologies, the demand grows for technologists' subject matter expertise and for translational frameworks that can define concepts across disciplinary boundaries. Within this space, HCI has a critical role to play in articulating how technical design practices intersect with regulatory concerns and in shaping tools that make such reasoning actionable. Though prior laws may have jurisdiction over some aspects of modern technologies, more recent laws often target specific practices discovered in scholarship. One such example is the EU Digital Services Act (DSA) [27][2], which prohibits the use of dark patterns in online platforms. Article 25 of the DSA states that *"[p]roviders of online platforms shall not design, organise or operate their online interfaces in a way that deceives or manipulates the recipients of their service or in a way that otherwise materially distorts or impairs the ability of the recipients of their service to make free and informed decisions."* This provision is further concretised by Recital 67, which explicitly mentions the term dark patterns, and contains the motivations, definitions and examples relevant to Article 25 [74]. Most importantly, this Recital mentions more directly the ability of recipients of a given service to make autonomous and informed choices or decisions. Both the Article and the Recital scope *autonomy* as a protected value in the context of dark patterns. Autonomy loss refers to the user's lack of capacity to have meaningful control over their choices and make free, informed, and meaningful choices or decisions [55]. Further, interdisciplinary interpretations of these provisions reveal that the DSA prohibits dark patterns by explicitly protecting user autonomy [39], articulating three different types of autonomy violations (*deception, manipulation,* and *distortion/impairment*) [74], and providing examples of potentially prohibited practices. Given that the European Commission can impose fines of up to 6% of the online platform's global revenue for significant breaches of the DSA[3] [27], clear reasoning is imperative for asserting that a given practice constitutes an autonomy violation.

With a wide variety of dark patterns present in many different contexts, it remains a challenge to map observed design practices to the autonomy violations outlined in such legal provisions [13, 54]. Therefore, there is a challenge in mapping observed design practices in different contexts to specific violations outlined in the DSA Article 25. Thus, we argue for methodologies and frameworks to help both regulators and industry reason about dark patterns in terms of the three prohibited autonomy violation types. Moreover, there is a need not only in determining if a given design practice violates user autonomy, but also in identifying the type of autonomy violation(s), bridging this reasoning closer to the legal provisions of the DSA. In this paper, we investigate the following research questions:

[2]Entered into force February 17, 2024.
[3]As per Article 52, paragraph 3.

RQ1: How do the dark patterns from Gray et al. [47] unified ontology map to autonomy violation types from Article 25 of the DSA?

RQ2: What design factors contribute to the reasoning about autonomy violations within dark pattern practices?

RQ3: How can the identified factors be applied to determine autonomy violations for any given design practice?

To answer these questions, we first qualitatively analyse the definitions of dark patterns from the Gray et al. [47] ontology in several iterative rounds, mapping them to three autonomy violation types from the DSA [74] in § 4.1. This mapping can enable regulators and platforms with a systematic tool to scrutinize known dark patterns and pinpoint precisely how and where these violate DSA provisions strengthening oversight and enforcement. Second, we further analyse the corpus of dark patterns to identify factors contributing to each identified autonomy violation. As a result, we identified *8 design factors* consisting of dark patterns' design traits that establish the reasoning for why that pattern may constitute a given autonomy violation (see § 4.2). Third, we articulate our reasoning to justify each dark pattern's autonomy violation types and contributing design factors in § 4.3. This reasoning forms one of the core findings of our work, wherein we describe the contributing mechanism of influence for each of the 59 dark patterns which leads to a particular autonomy violation.

We contribute our mapping of DSA autonomy violations to dark patterns and associated design factors as a framework enabling scholars, regulators, and industry practitioners to systematically analyze any design practice. We then demonstrate our framework's extensibility, as applied to new and emerging attention capturing deceptive practices recently identified by Monge Roffarello et al. [66] in § 5.1, as well as to an ongoing legal case in the EU against dark patterns in § 5.2.

By identifying the underlying design traits of dark patterns, we contribute to HCI scholarship, towards the development of shared concepts and a shared understanding of why certain types of designs may constitute dark patterns. Further, our work aligns dark patterns knowledge to a legal framework of autonomy violations, aiming to aid both regulators' enforcement and platform or designer compliance with the DSA in the European Union.

2 Background and Related Work

In this section, we review related work on dark patterns taxonomies and conceptualisations of "user autonomy" in the dark patterns context, as well as dark pattern regulation and enforcement. We then situate this work within broader literature.

2.1 Conceptualising Dark Patterns

2.1.1 Taxonomical Work. Since the term "dark patterns" was coined in 2010 [14], literature has identified and taxonomised dark patterns in various contexts. Brignull's original taxonomy was created as a practitioner-led work of awareness, proposing 12 dark patterns [14]. Zagal et al. [84] proposed a taxonomy of 7 dark patterns in games, categorised based on three broad intents, depending on whether they influence the user into expending time, money, or social capital. Bösch et al. [11] proposed a taxonomy of 7 privacy related dark patterns, while Mathur et al. [62] identified 15 dark patterns in e-commerce. Gray et al. [44] conducted one of the first harmonization efforts, creating a high level taxonomy of dark patterns inclusive of patterns from Brignull [14] and Bösch et al. [11]. Gray et al. [44] proposed 5 primary dark patterns, articulating strategic motivators behind such design practices. In addition to academic literature, regulatory sources have also proposed taxonomies of dark patterns, such as the European Commission [59], European

Data Protection Board (EDPB) [31], Organization for Economic Co-operation and Development (OECD) [69], United Kingdom's Competition and Markets Authority (CMA) [18], and United States' Federal Trade Commission (FTC) [37].

A recent comprehensive ontology of dark patterns by Gray et al. [47] unites ten taxonomies of dark patterns from academic, regulatory and practioner sources into three hierarchical levels. This ontology articulates systematic dark patterns definitions across 5 high-, 25 meso-, and 34 low-level patterns. The high-level consists of design *strategies* that characterize the broad nature of the influence afforded by a dark pattern. On the other hand, the low-level specifies the *means of execution*, potentially describing visual and/or temporal design elements that influence user decision making or choice. The high-level is the most abstract form of knowledge, while the low-level dark patterns are often context-specific. The meso-level bridges the high- and the low-levels by describing the *angle of attack*, i.e., the specific approach used to influence users and to undermine their decision making or choice.

2.1.2 Autonomy Risks. Dark patterns pose an acute cost to user autonomy and welfare, particularly for those with low digital literacy [58, 63]. Mathur et al. [63] analysed 19 definitions of dark patterns from literature, several of which included their tendency to *'confuse'*, *'deceive'*, *'exploit'*, *'manipulate'*, *'mislead'*, *'steer'*, and *'trick'* users, implying a violation of user autonomy. The literature review on harms caused by dark patterns [75] shows that loss of autonomy is a potential harm in itself, which is further correlated to additional harms and consequences, such as financial losses, loss of privacy, cognitive burden or social media addiction and conducive to a range of other harms. Autonomy concerns echo nearly all of dark patterns literature, with scholars working to articulate these. Ahuja and Kumar [4] analysed 151 dark patterns from literature and found that these designs can potentially undermine four different aspects of autonomy: *agency, freedom of choice, control* and *independence*. They defined each aspect of autonomy, mapping 151 patterns to seven specific autonomy concerns: inadequate information, biased evaluation, insufficient deliberation, restricted options, pressure to conform, lack of control, and compulsions. Mathur et al. [63] also recognized autonomy harms, and categorised dark patterns based on their design attributes: (a) those which modify users' decision space (asymmetric, restrictive, disparate treatment, covert); and (b) those which manipulate users' information flow (deceptive, information hiding). Despite all these efforts, defining autonomy remains challenging, as its meaning shifts across disciplinary and regulatory contexts.

2.2 Regulating and Enforcing Against Dark Patterns

A 2022 EU Commission report [59] found that "97% of the most popular websites and apps used by EU consumers deployed at least one dark pattern". In the EU, several regulations protect users from dark patterns and their effects. Existing regulations already apply in specific contexts: for instance, dark patterns in e-commerce can be covered by the Unfair Commercial Practices Directive (UCPD) [80], whereas data collection interfaces are governed by the General Data Protection Regulation (GDPR) [40]. Recently, several new EU regulations such as the DSA [27], DMA [26], Data Act [24] and AI Act [6] explicitly prohibit dark patterns, taking a formal stance against this design phenomenon. Nevertheless, all these legal provisions operate in different digital contexts (for example, DSA only covers "online platforms" while DMA targets "gatekeepers" and aim to protect users from different types of harms and within various visibility spectrums [57, Table 3][75]. Globally, other nations have also begun to issue formal prohibitions against the use of dark patterns. In the U.S., state privacy regulations explicitly ban dark patterns [1–3], while both state and federal rules may already have jurisdictional coverage (e.g. §5 of the FTC Act [22], which mandates the FTC to counter unfair and deceptive trade practices writ large).

2.2.1 Understanding DSA Article 25. Within the EU, the DSA is likely to encompass most online digital systems under its definition of "online platforms" [27, Art.3(i)][4]. Article 25(1) of the DSA prohibits dark patterns in online platforms, stating that: *"[p]roviders of online platforms shall not design, organise or operate their online interfaces in a way that deceives or manipulates the recipients of their service or in a way that otherwise materially distorts or impairs the ability of the recipients of their service to make free and informed decisions."* Following the interpretation of these concepts from a philosophy, law and HCI lens, Santos et al. [74] identified three different types of autonomy violations in the context of dark patterns prohibited by this law: *deception, manipulation* and *distortion/impairment*. We discuss these types further in § 3.1, as these three form our autonomy violation codebook. Santos et al. [74] describe the deception violation type as including design practices which create user perceptions that do not correspond to reality (i.e., false beliefs). The manipulation violation type includes design practices which have a steering effect on users' choices and decisions in a certain direction (potentially by bypassing due deliberation, or through trickery or pressure) – without being rationally persuasive, or deceptive; or coercive. Finally, the distortion/impairment violation type in Santos et al. [74] includes design practices which force or constrain users' choices and options.

These interpretations of the three autonomy violation types align with Benjamin Raue's legal commentary on DSA Article 25 [73]. This commentary describes deception as equivalent to giving an inaccurate impression, and manipulation as way to steer users, such as by steering their attentional or automatic decision making processes. The author argues that the transition from deception to manipulation is fluid and a clear distinction is often neither possible nor necessary. Similar to Santos et al. [74], Raue [73] also treats distortion and impairment as equivalent terms, which include designs that prevent or significantly impede users' freedom to make decisions in accordance with their preferences. This not only includes designs in which such freedom is completely suspended or eliminated, but also where obstacles are introduced to curtail this freedom.

2.2.2 Enforcing Against Dark Patterns. Recent enforcement actions show the growing regulatory focus on dark patterns across platforms and laws worldwide, illustrating both the momentum to curb manipulative practices and the early impact of the EU's DSA. Key cases from the EU and US illustrate how authorities are tackling these issues in practice. In the European context, from a consumer and competition law perspective, the European Consumer Organisation (BEUC) together with 25 members from 21 countries filed a complaint against SHEIN in June 2025 for multiple dark pattern practices, prompting the EU Commission to flag violations of consumer law [8, 30]. Similarly, the Polish Competition and Consumer Authority in 2024 fined Amazon €8 million in 2024 for misleading consumers about the conclusion of sales contracts on its online marketplace [81]. From a data protection view, Italy's data protection authority issued a €300,000 GDPR fine against a digital marketing company for using dark patterns to collect user consent [38]. Under the DSA, in May 2024, BEUC and 17 national consumer groups filed a complaint against Temu for presenting manipulative interface designs, misleading pricing/discounts potentially violating the DSA [7]. Although penalties are still pending, formal investigations are underway into X (formerly Twitter) for deceptive "verified accounts" [29], and Meta, for the use of dark patterns and addictive design features, especially those affecting minors and content recommendation systems [20].

In the U.S., the Federal Trade Commission (FTC) targeted major companies for using dark patterns. Epic Games misled minors into unintended in-game purchases without parental consent and faced a $520 million settlement [33]. Vonage made it hard for customers to cancel services, paying $100 million in refunds [32]. Amazon used dark patterns to enroll consumers in Prime without consent and making cancellation difficult through the multi-step 'Iliad Flow' [34].

[4]Indeed, several cases have recently showed an applicability of DSA to known online systems, such as an ongoing case against Meta since 2024 [20].

2.3 Comparison to Prior Work

We depart from prior interdisciplinary tech-law scholarship in a few key manners, including the direction of our analysis (from law-to-design instead of design-to-law), scope, and study design. Gunawan et al. [50] similarly conduct interdisciplinary tech-law scholarship, but perform literature review and case-law analysis rather than directly engaging with known dark patterns as we do in this study, and do not take a directional focus. We take a similar approach to Gray et al. [45], which starts with legal texts (specifically, FTC case documentation) and works in the direction of dark pattern analysis; they additionally take a narrow focus by inspecting a case against Amazon. Departing from Gray et al. [45], however, we work with legal scholarship and formal regulation as our starting texts, and instead of focusing on one platform, we inspect the broader Gray et al. [47] ontology of patterns.

In a prior work, though Mathur et al. [63] map dark patterns to design attributes, these attributes describe broader design strategies and are not further mapped to specific autonomy concerns; our work conducts the latter. Conversely, Ahuja and Kumar [4] map dark patterns to specific autonomy concerns grounded in philosophical literature; while our work is grounded in a legal framework of three autonomy violation types. Recent analyses of DSA Article 25 include Santos et al. [74] and Raue [73], both of which provide examples of dark patterns per the autonomy violation types. However, this recent literature does not explain *why* each specific dark pattern contains a given autonomy violation type. In this paper, we extend these works further, mapping a comprehensive set of 59 meso- and low-level dark patterns from Gray et al.'s ontology [47] to these autonomy violations. We also then identify a set of *eight design factors*, which can be used by regulators, scholars and companies to identify if and how a design practice violates user autonomy as outlined in Article 25 provisions. Our research methodology is explained below.

3 Methods

Our research team drew on collective expertise spanning human-computer interaction, computer science, law, and regulatory practice. Figure 1 summarizes various steps of our qualitative coding methods.

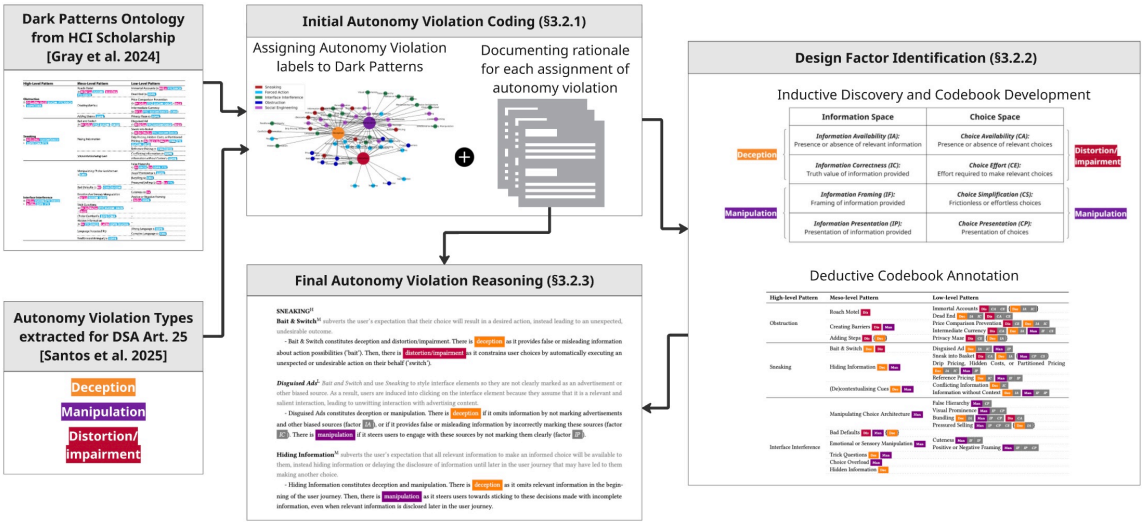

Fig. 1. Methodology used to identify design factors to determine dark pattern autonomy violations

3.1 Autonomy Violation Definition Selection

We use the three autonomy violation definitions as interpreted by Santos et al. [74], as these are the first dark patterns-specific autonomy taxonomy explicitly grounded in formal regulation (the EU DSA) and interpreted by legal scholars. This allows our work to engage directly with extant regulation for immediate impact.

Santos et al. [74] argue that dark patterns can potentially deceive users by making false statements or misleading statements that imply a falsehood; and also by omitting relevant information. They also note the potential to manipulate users through interface and/or linguistic design elements that steer users' perceptions, desires or emotions, without deceiving them or constraining their options [74]. Finally, Santos et al. [74] note that dark patterns might distort or impair users' autonomy by altering their choice set or by altering the conditions associated with available choices (such as threats, incentives, or unnecessary effort). For the purposes of our analysis and annotation, we take the following structured definitions of each autonomy violation, in accordance to and inferred from Santos et al. [74]:

- **Deception** violations *deceive, mislead,* or otherwise *interfere with user perceptions of truth.*
- **Manipulation** violations *steer* user behavior.
- **Distortion/impairment** violations *constrain* user behavior.

Naturally, persuasive technologies and designs all exert some form of influence upon user autonomy. However, as this paper is specific to the scope of dark patterns, we only interpret these three autonomy violations under the premise that the examined patterns result in unwanted consequences or harm to the user.

3.2 Dark Patterns Corpus Selection

We strictly utilize the Gray et al. [47] ontology described in § 2.1.1 to center our work, both to align with the broader dark patterns community and to work with a centralized known resource. We exclude the five high-level patterns (Obstruction, Sneaking, Interface Interference, Forced Action and Social Engineering) from the analysis, as they are broad design strategies and do not specify the manner in which a design practice may influence users' choices or decision making. Instead, we map the remaining meso- (N=25) and low-level (N=34) patterns to specific autonomy violations, by analysing the definition of each dark pattern from Gray et al. [47], while using the interpretations of autonomy from Santos et al. [74] as described in § 3.1. These 59 total patterns thus constitute our corpus of exemplars for analysis.

3.3 Analysis Procedure

To code these dark patterns systematically and map them to the autonomy violations, we conducted annotations and related discussions in the following iterative stages:

3.3.1 Initial Autonomy Violation Coding. We conducted this coding round to establish an initial mapping of dark patterns to the three autonomy violations. Two authors with interdisciplinary expertise in dark patterns independently coded all 59 patterns using AirTable, assigning one or more autonomy violations to each and documenting their rationale in brief memos. These two authors then met to discuss all labels, assessing discrepancies in three rounds. For every round, we recorded instances of perfect agreement and deliberated on discrepancies to reach consensus, producing new memos to document our joint rationale.

In the first round, we inspected dark patterns that were assigned a single autonomy violation by both coders (N=22). We were in agreement for 17 patterns, and we discussed the five differing patterns towards consensus, either choosing

one best-fit violation type or choosing both violation types. For example, *Complex Language* is a low-level dark pattern that makes information difficult to understand by using obscure word choices and/or sentence structure [47]. One coder labeled it as deception, and the other as manipulation. At the discussion stage, we concurred on using both labels, as the autonomy violation would depend on whether the complex language is potentially misleading (deception), or it simply discourages the user from engaging with the information provided (manipulation).

In the second round, we inspected dark patterns which were assigned a single label by one coder, but multiple labels by the other coder (N=22). For these, we noted that the single label assigned by one coder always appeared as part of the multiple labels assigned by the other coder. We again discussed these patterns towards consensus, either choosing one or multiple violation types which fit best. For example, *Countdown Timer* is a low-level dark pattern which indicates that a deal or discount will expire by displaying a countdown clock or timer [47]. One coder labeled it as deception, and the other as both deception and manipulation. During discussions, we concurred on using both labels, as the autonomy violation would depend on whether the timer is fake (deception), or it is genuine but it creates an undue sense of urgency that steers users towards a purchase (manipulation).

Third and last, we inspected dark patterns which were assigned multiple labels by both coders (N=15). We agreed upon 13 patterns, and we discussed two differing patterns towards consensus.

3.3.2 Design Factor Identification. Following our discussions in § 3.3.1, we noted common factors arising in our rationale memos. To explore these further, we conducted an inductive exploration round to identify and enumerate these factors, and a second deductive round to deepen our coding efforts. For this stage, we excluded the meso-level patterns, as the meso-level definitions describe only an angle of attack, i.e., the approach used to influence users, but do not specify the exact means of execution – making it difficult to infer specific design aspects leading to an autonomy violation (even if the violation type is clear). As such, we identify design factors only for the 34 low-level ontology patterns [47].

Inductive Discovery and Codebook Development. The same two authors coding the violations in § 3.3.1 revisited the low-level pattern definitions, the autonomy violation labels assigned to them, and rationale memos of both authors. This time inspecting the reasoning memos, the authors inductively identified and iterated upon the design factors which contributed to our decision to assign particular autonomy violation label(s) to each dark pattern. We refined these factors in discussions with all authors, finally identifying a set of eight factors, each of which contributed to a particular autonomy violation type (Deception – 2 factors; Manipulation – 4 factors; and Distortion/impairment – 2 factors). We categorised these eight factors further into two broad design *spaces*: Information Space and Choice Space. We explain these design factors and spaces in § 4.2.

Deductive Codebook Annotation. With this set of eight factors as a codebook, we deductively re-coded each of the 34 low-level patterns. Like § 3.3.1, initial coding was done independently across all patterns in the subset by two authors via AirTable, with coders assigning one or more design factor and providing a rationale memo. At this stage of coding, we only labeled design factors for the autonomy violation(s) already assigned to that pattern in § 3.3.1. However, if authors (inductively) noted the potential for other design factors, and thus the potential for other autonomy violations, this was documented in the memos and flagged for future discussion.

In total, for 34 low-level patterns, the two authors independently coded 162 binary data points for the presence or lack thereof of a given factor (Deception: 26 labels x 2 factors; Manipulation: 20 labels x 4 factors; Distortion/impairment: 15 labels x 2 factors). For independently-annotated codes, authors agreed on 136 out of 162 items resulting in Cohen's Kappa of $k = 0.679$ representing substantial agreement [19]. Remaining discrepancies were discussed towards consensus.

3.3.3 Final Autonomy Violation Reasoning. In the final step, we created a textual reasoning structure to justify each dark pattern's autonomy violation labels and contributing design factors. For this, we consulted all prior labels and memos, refining our reasoning and flagging potential changes to the autonomy violation labels for the meso-level patterns as well (The low-level patterns were flagged for discussion in the previous step described in § 3.3.2). The same two authors from § 3.3.1 wrote the initial reasoning for each of the 59 dark patterns. We then discussed the reasoning for all flagged dark patterns with a third author, resulting in changes to autonomy violation labels for 11 out of 59 dark patterns (we added a new autonomy violation to 9 dark patterns and removed an autonomy violation from 2 dark patterns). Out of these 9 dark patterns where a new autonomy violation was added, 7 were low-level patterns, and together we identified their contributing design factors. All these changes supplemented our reasoning. Lastly, a fourth author read through and validated our reasoning for all 59 dark patterns, resulting in no changes to the autonomy violation labels and only one addition of a contributing design factor in one of the dark patterns. Our final reasoning structure is explained in §4.3 and the detailed reasoning for all 59 dark patterns is provided in § 7.1 of the Appendix.

4 Results

Our analysis described in § 3.3 resulted in the final mapping of the meso- and low-level dark patterns from Gray et al. [47] to one or more autonomy violations from Santos et al. [74]; and then to the eight design factors underlying these autonomy violations. This section describes our results, providing an overview on the different types of mappings of dark patterns to the autonomy violations (§ 4.1), articulating design factors that determine such violations (§ 4.2) and presenting the final mapping with the accompanying reasoning for each dark pattern (§ 4.3).

4.1 Mapping Dark Patterns to Autonomy Violations

Overall, out of the 59 dark patterns, 17 mapped to a single autonomy violation and 42 mapped to multiple autonomy violations, and we discuss several instances of both types below[5].

4.1.1 Single Autonomy Violation Type. In our mapping, 17 dark patterns mapped to a single type of autonomy violation, either deceiving *or* manipulating *or* distorting/impairing user autonomy, as illustrated in the following examples:

- **Deception** violations are associated with 2 meso- and 2 low-level patterns. For example, the *Pay-to-Play*^{L} dark pattern claims that aspects of a service or a product are available via purchase or download, but later charging users further to actually obtain that functionality [47]. We labeled it as deception – either because of an explicit false or misleading claim that users can access certain functionality, or because of information being omitted that users need to pay further to obtain this functionality. Another example is the *Conflicting Information*^{L} dark pattern, which includes two or more sources of information that conflict with each other [47]. We labeled it as deception – because it implies that at least one source or piece of information is false or misleading.
- **Manipulation** is associated with 4 meso- and 5 low-level patterns. For example, the *Visual Prominence*^{L} dark pattern places an element relevant to user goals in visual competition with a more distracting and prominent element [47]. We labeled it as manipulation – as it steers user perception through distracting and prominent interface elements, hence steering their choices. Another example is the *Confirmshaming*^{L} dark pattern, which frames a choice to opt-in or opt-out of a decision through emotional language or imagery that relies upon shame or guilt [47]. We also labeled it as manipulation – as it steers users' choices by associating emotions of shame or guilt with certain choices.

[5]In the following examples, we label each dark pattern type with "M" or "L" superscript indicating the type of the pattern: meso- or low-level.

- **Distortion/impairment** is associated with 3 meso- and 1 low-level pattern. For example, the *Grinding*^{L} dark pattern requires repeated, often cumbersome and labor-intensive actions over time in order to obtain certain game functionality [47]. We labeled it as distortion/impairment – because the user is either forced to grind or forced to pay in order to obtain this functionality.

4.1.2 Multiple Autonomy Violation Types. The remaining 42 dark patterns were mapped to multiple autonomy violations, out of which 31 were mapped to two violations and 11 to all three violation types. These multiple violations can occur *separately*, i.e. one *or* the other ; or *together*, i.e. one *and* the other.

Autonomy violations that occur separately: In some dark patterns, the multiple autonomy violations occur *separately*, i.e., one *or* the other. For example, the *Activity Messages*^{L} dark pattern presents users with data about other users' purchases, views, visits, or contributions [47]. We labeled it as deception – if the information presented is false or misleading; or manipulation – if the information presented is not false or misleading but framed in a way to create a sense of urgency. In this example, the dark pattern can *either* deceive *or* manipulate.

Autonomy violations that occur together: Other dark patterns may implicate two or all autonomy violations occurring *together*. We stratify these into the following cases:

- **Autonomy violations that follow a temporal progression:** The multiple autonomy violations may follow a temporal progression in some dark patterns. For example, in the *Drip Pricing, Hidden Costs, or Partitioned Pricing*^{L} dark pattern, an e-commerce website hides information about full costs of a product or a service [47]. We identified it as deception in the beginning – because of hidden cost information. However, as these costs are revealed later in the user journey, a user may still be steered towards making a purchase, not wanting to abandon their progress – because of which the dark pattern is also labeled as manipulation. In this case, the initial deception transforms into manipulation based on the temporal progression of the user journey.
- **Autonomy violations that reinforce each other:** Other cases of dark patterns may violate autonomy in multiple ways that reinforce each other. For example, the *Sneak into Basket*^{L} dark pattern adds unwanted items to a user's shopping cart without their consent [47]. We identify this as distortion/impairment – because items are added to cart without any user action, whereas removing them requires direct action. At the same time, there is also deception – because the user is not explicitly informed that such items have been added. Both these autonomy violations occur together.
- **"Additional" autonomy violations:** For 15 out of the 42 dark patterns with more than one autonomy violation labels, we identified that some violations were present only in a specific context or condition (depending on the task flow surrounding the 'core' dark pattern or the exact implementation of the pattern). For example, extending the case of *Sneak into Basket*^{L} discussed in the previous point, this dark pattern *always* forces the user to take extra action to remove the unwanted item from cart (distortion/impairment) and it *always* leaves the user uninformed about these unwanted items (deception). However, *conditionally*, *if* the items sneaked into users' carts are actually tempting, it may steer users towards keeping them at the last minute (manipulation). Hence, manipulation is the additional autonomy violation depending on the exact item added to cart.

4.1.3 Autonomy Violation Types Across High-level Categories. The distribution of autonomy violations also varies across the five broad high-level categories in the Gray et al. [47] ontology: *Obstruction, Sneaking, Interface Interface, Forced Action,* and *Social Engineering*. While the resulting labeling of each dark pattern may contain one, two, or all three

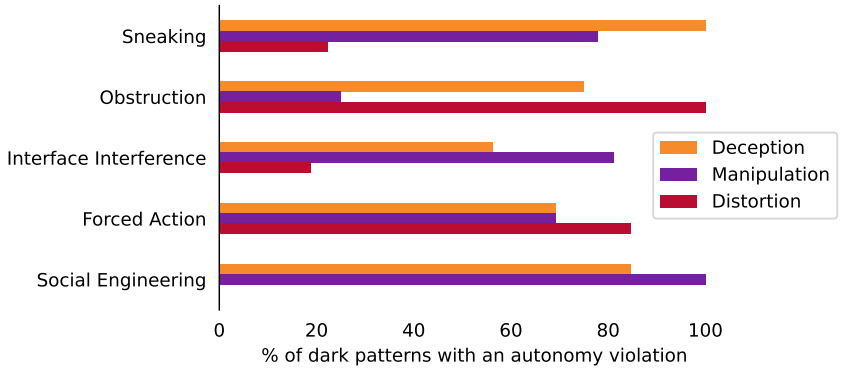

Fig. 2. Bar chart depicting percentages of patterns including an autonomy violation, grouped by the high-level pattern type

autonomy violations, including "additional" violations, in this section we have included all such autonomy violations when providing visual representations of our mapping.

Figure 2 presents the normalized percentages of each of the three autonomy violation labels (*deception, manipulation,* and *distortion/impairment*) across these five high-level categories of dark patterns. This stresses *Sneaking, Obstruction* and *Social Engineering* dark patterns to be the most consistent internally; that is, all patterns (100%) within the high-level category share one of the autonomy violation types: *Sneaking* patterns all constitute deception, all *Obstruction* patterns constitute distortion/impairment, and all *Social Engineering* patterns constitute manipulation. This indicates the contextual nature of dark patterns, which is consistent with the findings from Gray et al. [47] regarding the necessity of a nuanced approach to dark pattern identification. That is, though some types of dark patterns always trigger a specific autonomy violation, diversity in design choices may result in different autonomy-violative mechanisms.

We note more comparative "diversity" in autonomy violations of the *Interface Interference* high-level strategy. For *Interface Interference* the most prominent autonomy violation is manipulation, in which the interface might steer users towards specific choices without necessarily hiding or eliminating them (such as *Manipulating Choice Architecture*^{M}). Deception features secondarily, if the interface is also used to disguise relevant information or choices (such as in *Hidden Information*^{M}) or to create a mismatch between information and action possibilities (such as in *Feedforward Ambiguity*^{M}). We find that the *Forced Action* category is comparably the most diverse in terms of autonomy violations across all high-level strategies. The most prominent autonomy violation in this category is distortion/impairment, for example, whenever a user is mandated or pressured to take certain actions. Deception also features secondarily here, as dark patterns can create a false perceptions, making an action appear mandatory (hence, forced) even when it is not actually mandatory (such as in *Forced Registration*^{M}). Steering effects for manipulation may also be present, but this largely depends on the manner in which the pattern is deployed.

Distribution across combination subsets. Figure 3 contains the network graph visually presenting the relationships between autonomy violations and all 59 dark patterns, with a color coding of each dark pattern based on its high-level categorisation in the Gray et al. [47] ontology. The graph provides more detail as to the distribution of the autonomy violations beyond Figure 2 and stresses two main trends: first, that some high-level strategies cluster around certain combinations of autonomy violations, and second, differences in pairwise frequency.

Regarding the distribution of strategies within each combination subset, we often note a particular high-level strategy constituting the majority of the subset. For manipulation-only patterns, the prominent high-level pattern is *Interface Interference*^{H} with 7/9 patterns; for distortion-only patterns this is *Forced Action*^{H} with 3/4 patterns. For two-violation groups, *Obstruction*^{H} patterns overwhelm the deception-distortion pair (4/6 patterns). We believe this affirms the overall internal consistency of most Gray et al. [47] high-to-low mappings and highlights dark patterns that are comparatively less complicated to reason through, while still acknowledging areas for further tightening dark pattern conceptualisation.

Out of all potential combinations of the three violations, the manipulation-distortion pair had the fewest (N=2) dark patterns, with the next-lowest combinations being single-violation groups for deception and distortion with 4 patterns each. These two patterns are *Creating Barriers*^{M} and *Auto-Play*^{L}, which correspond to separate high-level strategies. Conversely, the deception-manipulation subset was the largest subset across all possible combinations with 21 patterns total. We hypothesise that these two extremes stem from conceptual similarity, as in our many iterative discussions it was far easier to define distortion cases as distinct, whereas deception and manipulation at times seemed to present as two sides of a continuum of influence (separated only by a severity threshold). Our reasoning in § 7.1 explains how we delineated deception and manipulation, but we expect this to be an interesting distinction to explore in future scholarship.

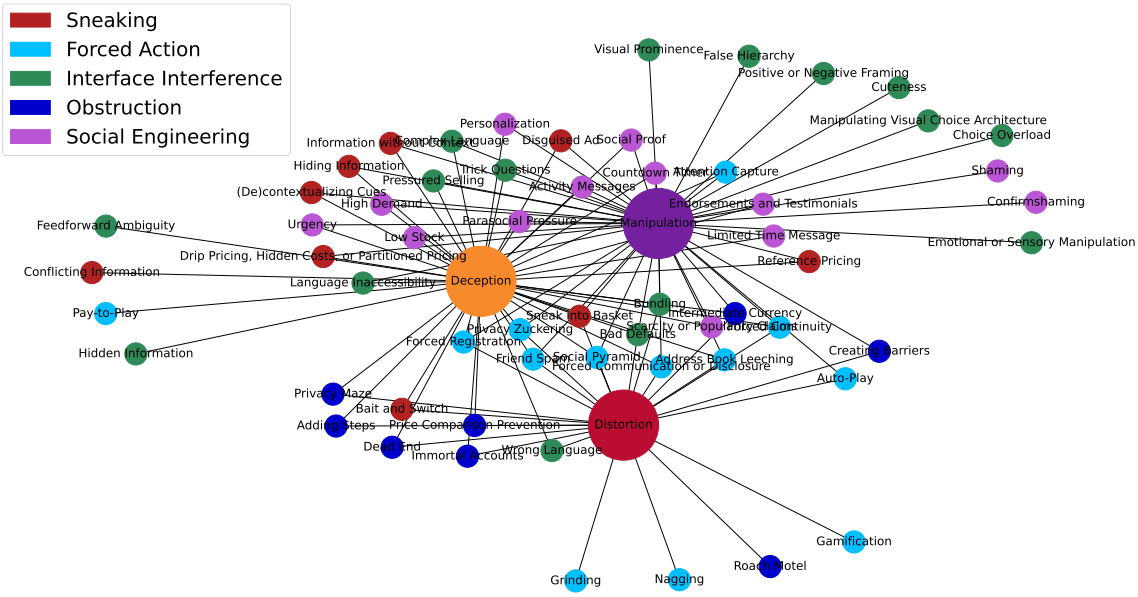

Fig. 3. Network graph of relationships between autonomy violations and dark patterns, with nodes colored according to their Gray et al. [47] high-level pattern type

4.2 Articulating Design Factors to Determine Dark Pattern Autonomy Violations

As described in § 3.3, after Rounds 1 and 2 of analysis, we identified *eight design factors* to help determine which of the three autonomy violation type(s) apply to any observed dark pattern. Each of these eight design factors maps to a particular autonomy violation (Deception – 2 factors; Manipulation – 4 factors; and Distortion/impairment – 2 factors) (see §3.3.2). We further categorised these factors into two broad design *spaces*: the *Information Space* – influencing users by altering information – and the *Choice Space* – influencing users by altering their choices – presented in Figure 4.

	Information Space	Choice Space	
Deception	***Information Availability (IA):*** Presence or absence of relevant information	***Choice Availability (CA):*** Presence or absence of relevant choices	Distortion/ impairment
	Information Correctness (IC): Truth value of information provided	***Choice Effort (CE):*** Effort required to make relevant choices	
Manipulation	***Information Framing (IF):*** Framing of information provided	***Choice Simplification (CS):*** Frictionless or effortless choices	Manipulation
	Information Presentation (IP): Presentation of information provided	***Choice Presentation (CP):*** Presentation of choices	

Fig. 4. Design factors which help determine the dark pattern autonomy violation type(s)

The ***Information Space*** consists of all the information made available to the user by an online platform to support user choices and decisions. This includes information that pertains to a choice, but also the information about the existence of a choice itself. It consists not only of textual information, but also iconography, graphics, colors, or any other form of information and the manner in which it is conveyed to the user. These different aspects of the information space are categorised into *four design factors* within our framework. The ***Choice Space*** consists of the set of choices and options made available to the user by the online platform. It consists of the relationship between different choices, i.e., if some choices are easier or more difficult than others, if some choices are incentivised, and the way the different choices are presented to the user. We categorise these different aspects of the choice space into *four design factors*. Together, the eight factors can help reason about why a particular dark pattern constitutes a particular autonomy violation, as shown in Figure 4. We describe these design factors and explain how they contribute to the three autonomy violation types.

4.2.1 Deception. We identified two design factors, both in the information space, which can contribute to deception. These are: *Information Availability (IA)* and *Information Correctness (IC)*.

(1) ***Information Availability (IA)*** is concerned with the *presence or absence of relevant information*. This includes information which is relevant to make a choice, but also information about the existence of a choice itself. This factor is implicated in deception when a dark pattern omits or excludes relevant information entirely; or it does not provide relevant information at a relevant time or in the relevant context, thus limiting its discoverability.
(2) ***Information Correctness (IC)*** is concerned with the *truth value of information provided*. This includes not just textual information, but also iconography, graphics, colors, or any other form of information. This factor is implicated in deception when a dark pattern provides user with information that is false or misleading.

4.2.2 Manipulation. We identified four design factors – two in the information space and two in choice space – which can contribute to manipulation. These are: *Information Framing (IF), Information Presentation (IP), Choice Simplification (CS)* and *Choice Presentation (CP).*

(1) ***Information Framing (IF)*** is concerned with the *framing of information provided.* This means that the way that information is conveyed shapes users' understanding and perception of choices in one way or another – in addition to relevant facts. This can include emphasising positive or negative aspects of a choice, creating trust, or exploting emotions such as guilt, fear, confusion or doubt. This factor is implicated in manipulation when a dark pattern steers users towards or away from certain choices by conveying information in a certain way.
(2) ***Information Presentation (IP)*** is concerned with the *presentation of information provided* in the interface. This includes elements of layout, color, order, structure, hierarchy or timing of the information provided. This factor can also be implicated in manipulation, as dark patterns can use these interface elements to perceptually steer users towards or away from particular information, without actually omitting any relevant information.
(3) ***Choice Simplification (CS)*** is concerned with *frictionless or effortless choices* that require little or no effort on part of the user. The absence of friction may steer users towards these choices, even though there are no constraints or restrictions preventing alternative actions. This factor is implicated in manipulation when the absence of friction steers users towards certain choices due to reduced reflection or deliberation.
(4) ***Choice Presentation (CP)*** is concerned with the *presentation of choices* in the interface. This includes elements of layout, color, order, structure, hierarchy or timing of choices presented to the user. This factor can also be implicated in manipulation, as dark patterns can use these interface elements to perceptually steer users towards or away from certain choices, without actually omitting any options.

4.2.3 Distortion/impairment. We identified two design factors, both in the choice space, which contribute to distortion/impairment. These are: *Choice Availability (CA)* and *Choice Effort (CE).*

(1) ***Choice Availability (CA)*** is concerned with the *presence or absence of relevant choices.* This includes restrictions on the user's choice set by omitting relevant choices, or by mandating or forcing a user to perform certain actions. It also includes pre-selecting or executing a choice on behalf of the user without any user action. This factor is implicated in distortion/impairment, as it causes a user to act unwillingly or involuntarily, or prevents them from choosing their desired course of action.
(2) ***Choice Effort (CE)*** is concerned with the *effort required to make relevant choices.* This includes making certain choices unreasonably difficult or effortful, in terms of time, number of steps, or any other form of effort. This factor is implicated in distortion/impairment, when it causes a user to unwillingly choose the easy course of action, even if they are aware of alternate choices.

4.3 Mapping dark patterns to autonomy violations and design factors

Following our deductive codebook annotation procedure described in § 3.3.2, we have finalized both labels for each meso- and low-level dark pattern: autonomy violations and contributing design factors. Table 1 presents the resulting mapping of dark patterns to autonomy violations, along with the contributing design factors identified for the low-level patterns. Autonomy violations are organized according to their relationship with a given dark pattern. When a pattern consistently results in a violation type, the violation is listed directly. When the violation arises only conditionally or contextually, it is marked as "additional" (see § 4.1.2) and shown in curly braces.

Table 1. Mapping of meso- and low-level dark patterns [47] to autonomy violation types [74] and their associated design factors. Dec = deception, Man = manipulation, Dis = distortion/impairment. Design factors are shown in *gray* (e.g., *IA* = "Information Availability"). Additional autonomy violations with associated design factors are shown inside curly brackets { }.

High-level Pattern	Meso-level Pattern	Low-level Pattern
Obstruction	Roach Motel Dis	Immortal Accounts Dis *CA* *CE* {Dec *IA* *IC*}
		Dead End Dec *IA* *IC* Dis *CA* *CE*
	Creating Barriers Dis Man	Price Comparison Prevention Dis *CE* Dec *IA* *IC*
		Intermediate Currency Dis *CA* Dec *IA* Man *IF* *CS*
	Adding Steps Dis {Dec}	Privacy Maze Dis *CE* {Dec *IA*}
Sneaking	Bait & Switch Dec Dis	Disguised Ad Dec *IA* *IC* Man *IP*
	Hiding Information Dec Man	Sneak into Basket Dis *CA* Dec *IA* {Man *CP* *CS*}
		Drip Pricing, Hidden Costs, or Partitioned Pricing Dec *IA* *IC* Man *IP*
		Reference Pricing Dec *IC* Man *IF* *IP*
	(De)contextualising Cues Dec Man	Conflicting Information Dec *IC*
		Information without Context Dec *IA* Man *IF* *IP*
Interface Interference	Manipulating Choice Architecture Man	False Hierarchy Man *CP*
		Visual Prominence Man *IP* *CP*
		Bundling Dec *IA* Man *IP* *CP* Dis *CA*
		Pressured Selling Man *IP* *CP* *CS* {Dec *IA*}
	Bad Defaults Dis Man {Dec}	
	Emotional or Sensory Manipulation Man	Cuteness Man *IF* *IP*
		Positive or Negative Framing Man *IF* *IP* *CP*
	Trick Questions Dec Man	
	Choice Overload Man	
	Hidden Information Dec	
	Language Inaccessibility Man Dec	Wrong Language Dis *CA* {Dec *IA*}
		Complex Language Man *IF* Dec *IC*
	Feedforward Ambiguity Dec	
Forced Action	Nagging Dis	
	Forced Continuity Dis {Dec Man}	
	Forced Registration Dis Dec Man	
	Forced Communication or Disclosure Dis Dec Man	Privacy Zuckering Dis *CA* Dec *IA* *IC* {Man *IF* *IP*}
		Friend Spam Dis *CA* {Dec *IA* *IC* Man *IF* *IP*}
		Address Book Leeching Dis *CA* Dec *IA* *IC* {Man *IF* *IP*}
		Social Pyramid Dis *CA* {Dec *IA* *IC* Man *IF* *IP*}
	Gamification Dis	Pay-to-Play Dec *IA* *IC*
		Grinding Dis *CA* *CE*
	Attention Capture Man {Dec}	Auto-Play Man *CS* {Dis *CA* *CE*}
Social Engineering	Scarcity and Popularity Claims Dec Man	High Demand Dec *IC* Man *IF* *IP*
	Social Proof Man {Dec}	Low Stock Dec *IC* Man *IF* *IP*
		Endorsements and Testimonials Man *IF* *IP* Dec *IC*
		Parasocial Pressure Man *IF* *IP* Dec *IC*
	Urgency Dec Man	Activity Messages Dec *IC* Man *IF* *IP*
		Countdown Timer Dec *IC* Man *IF* *IP*
		Limited Time Message Dec *IA* *IC* Man *IF* *IP*
	Shaming Man	Confirmshaming Man *IF*
	Personalization Man Dec	

Intriguingly, the autonomy violation labels of low-level dark patterns generally but do not always match with those for corresponding meso-level patterns. In some cases, the low-level patterns may have more autonomy violation labels than the meso-level, because the low-level definition often describes additional and specific means of execution, making it possible to infer more autonomy violations. For example, *Roach Motel*^{M} is assigned only the distortion/impairment label, whereas two of its low-level patterns (*Immortal Accounts*^{L} and *Dead End*^{L}) have also been assigned the deception label. In fewer cases, there are meso-level patterns (such as *(De)contextualising Cues*^{M}), whose definition is a summation of its low-level patterns (*Conflicting Information*^{L} and *Information without Context*^{L}). Therefore, the low-level patterns might have fewer autonomy violation labels, as each of them cover only a partial aspect of the meso-level pattern. Overall, though we noted these quirks after reviewing patterns altogether, we first prioritise adherence to the ontological text and analyse each definition as a standalone dark pattern. Thus the meso- to low-level mismatches in Table 1 are left intentionally as such.

The exact reasoning supporting the assignment of the autonomy violations and the design factors for the 59 meso- and low-level dark patterns is reported in the Appendix (§7.1). Though we did not aim to create a unified reasoning structure for each dark pattern that explains our mappings, we have still tried to make such reasoning consistent across patterns. Through our process, we have considered important not only to describe which autonomy violations each dark pattern constitutes, but also, whenever possible, the expectation of the user and the mismatch between this expectation and what has happened, similarly to the logic behind the original definitions of dark patterns by Gray et al. [47, §5]. Nevertheless, though original definitions of dark patterns from Gray et al. [47] implicitly contained references to deception or manipulation or distortion/impairment, these autonomy violation types were not clearly labeled and unsupported by the reasoning of design factors – a gap we have started closing with this work.

Based on our iterative generation of the reasoning with four contributing co-authors described in § 3.3.3, we have identified the common structure for the reasoning described below. This structure differed for patterns with only a single autonomy violation type from those with multiple types (see § 4.1). For each autonomy violation type, we identified the *"Contributing Mechanism of Influence"*, which is by nature different, depending on the autonomy violation type, and is inspired by the interpretations of Santos et al. [74], presented in § 3.1. As such, for *deception*, the mechanism would explain if a dark pattern *omits information* or *provides false or misleading information*, thus creating a false perception for the user. For *manipulation*, the mechanism explains how a dark pattern *steers user choices*. For *distortion/impairment*, the mechanism explains how a dark pattern *constrains user choices*. Additionally, for low-level patterns, the *"Contributing Mechanism of Influence"* also utilises the eight contributing design factors to describe the reasoning underlying the autonomy violations (see § 4.2), such as "Information Correctness" labeled accordingly with *IC*.

For each dark pattern, we take inspiration from the its original definition from Gray et al. [47] and describe the mechanism causing the appearance of the claimed autonomy violation. However, we do not aim to claim that mapping these dark patterns to autonomy violations always means that they must be prohibited under DSA Article 25. For example, some dark patterns may be covered under the GDPR or the UCPD, as discussed in § 2.2, whereas others may not surpass the threshold of scale or severity required for regulatory prohibitions. Therefore, our mappings demonstrate the autonomy violation potential for each dark pattern, without making any arguments for legal prohibitions, which would depend on the exact context and the exact implementation of a design.

4.3.1 Single Autonomy Violation Type. For dark patterns with a single autonomy violation type (see § 4.1.1), we identified the following structure:

{Dark Pattern} constitutes **{Autonomy Violation}** as **{Contributing Mechanism of Influence}**.

For example, the *Choice Overload*^{M} original definition from Gray et al. [47] is: "[it]subverts the user's expectation that the choices they make should be understandable and comparable, instead providing too many options to compare or encouraging users to overlook relevant information due to the volume of choices provided.". Our reasoning is the following:

> Choice Overload constitutes manipulation as it steers users towards particular choices or options by increasing the volume of available choices, leading users to overlook some options or relevant information about these options.

Another example is the *Pay-to-Play*^{L} pattern, which "initially claim[s] that aspects of a service or product are available via purchase or download, but then later charging users to actually obtain that functionality. As a result, the user incorrectly assumes that a service or product will allow them certain functionality, leading to them downloading or purchasing the product or service under false pretenses." [47] Our reasoning, detailed below, illustrates how multiple design factors may contribute to a single autonomy violation:

> Pay-to-Play constitutes deception as it omits information (factor *IA*) or provides false or misleading information (factor *IC*) to create a false perception that a certain functionality of a product or a service is available via purchase or download, whilst later charging users additionally for that functionality.

4.3.2 Multiple Autonomy Violation Types. For dark patterns with more than one autonomy violation type (see §4.1.2), we identified the following structure:

> **{Dark Pattern}** constitutes **{First Autonomy Violation}** and[/or] **{Second Autonomy Violation}**.
> There is **{First Autonomy Violation}** as/if **{Contributing Mechanism of Influence}**.
> There is **{Second Autonomy Violation}** as/if **{Contributing Mechanism of Influence}**.

For example, *Creating Barriers*^{M} "subverts the user's expectation that relevant user tasks will be supported by the interface, instead preventing, abstracting, or otherwise complicating a user task to disincentive user action." [47]. Our reasoning is explained below, with the two autonomy violation types separated by an *"and/or"*, indicating that the violations could occur *separately* or *together* (as explained in §4.1.2):

> Creating Barriers constitutes distortion/impairment and/or manipulation. There is distortion/impairment if it constrains user choices by preventing certain tasks or by making them unreasonably difficult or complicated. There is manipulation if it steers user choices by preferring or prioritizing certain tasks in the interface, without any constraints or restrictions.

Another example is *Reference Pricing*^{L}, which "include[s] a misleading or inaccurate price for a product or service that makes a discounted price appear more attractive. As a result, the user is misled into believing that the price they pay is discounted, leading them to make a decision to purchase a product or service on false pretenses." [47]. Our reasoning is explained below, with the two autonomy violation types separated by *"and"*, indicating that both violations occur *together* (as explained in §4.1.2):

> Reference Pricing constitutes deception and manipulation. There is deception as it provides false or misleading information about the 'original' price of a product or service (factor *IC*). There is manipulation as it steers users towards a purchase, because the false original price frames the 'actual' price as discounted in comparison (factor *IF*), often accompanied by aesthetic or visual cues highlighting this discount (factor *IP*).

4.3.3 Additional Autonomy Violations. For both meso- and low-level patterns, we identified cases where an autonomy violation occurs only under specific contexts or conditions (see § 4.1.2). In such cases, we add an additional statement in the reasoning, labeled *Additional autonomy violation*, which is appended after the main reasoning text and follows the structure described below:

> ***Additional autonomy violation:*** There is also a possibility of **{Autonomy Violation}** if **{Additional Contributing Mechanism of Influence}**.

For example, the *Sneak into Basket*^{L} "add[s] unwanted items to a user's shopping cart without their consent. As a result, a user assumes that only the items they explicitly added to their cart will be purchased, leading to unintentional purchase of additional items." [47] Our complete reasoning, including additional autonomy violation, is as follows:

> Sneak into Basket constitutes distortion and deception. There is distortion/impairment as it constrains user choices by adding items to users' cart without any user action, whereas removal of these items necessitates direct action (factor *CA*). There is deception as it omits explicit information that such items have been added (factor *IA*).
> ***Additional autonomy violation:*** There is also a possibility of manipulation if it steers users towards a purchase through last minute presentation of potentially tempting items in the cart (factor *CP*) and the reduced friction in adding them to cart (factor *CS*).

5 Utilising and Extending the Framework

Herein we discuss how our framework can be applied to dark patterns that are currently not represented in the ontology of Gray et al. [47]. Indeed, Gray et al. [47] posited that the dark patterns ontology is inherently extendable, supporting both the addition of new patterns and strengthening contextual or domain specific examples of existing patterns. In this section, we demonstrate that our framework is intended as a complement to extant ontologies and taxonomies, and may similarly be applied to new patterns as they are identified. First, we analyse in §5.1 the case of attention capturing deceptive practices, which extend the ontology of dark patterns, as well as are of interest to both EU regulators and upcoming EU law. Then, in §5.2, we apply our framework to an ongoing EU Commission investigation under DSA Article 25, demonstrating the usefulness of our approach in reasoning about specific dark pattern cases.

5.1 Beyond the Ontology: Attention Capturing Design Practices

Monge Roffarello et al. [66][6] identified that *attentional harms* have not been covered extensively in the existing taxonomies and ontologies of dark patterns. Monge Roffarello et al. [66] conduct a systematic literature review conceptualising *Attention Capture Deceptive Patterns (ACDPs)*: patterns that lay "at the intersection of damaging design patterns and digital wellbeing" and are defined as *"recurring pattern[s] in digital interfaces that a designer uses to exploit psychological vulnerabilities and capture attention, often leading the user to lose track of their goals, lose their sense of time and control, and later feel regret"* [66, § 3.1]. This work identifies 11 ACDPs observed primarily in social media platforms but also present in games and video streaming services. The 2023 EU Parliament resolution [71] also emphasizes that many digital services are deliberately designed to prolong user engagement by exploiting psychological vulnerabilities

[6]Through this work occurred prior to the Gray et al. [47] ontology, this work was not considered as a source for the ontology and we have noticed many patterns that can extend the ontology.

– what the EU Parliament refers to as "addictive design". The EU Parliament report [72] leading to this resolution explicitly mentioned eight of these 11 design practices.[7]

In this section, we demonstrate the practical utility of our framework to these timely and understudied attention capture damaging patterns (ACDPs). To illustrate the use of our reasoning for such dark patterns that extend the Gray et al. [47] ontology, we first analyse the definitions of the 11 ACDPs from Monge Roffarello et al. [66] to identify which of them map to existing dark patterns within the ontology, and which may warrant extending the ontology. We find that three out of 11 practices – Neverending Autoplay, Disguised Ads and Recommendations, and Grinding – either completely or largely match with existing dark patterns in the ontology. For these, we draw on the mappings presented in Table 1 to identify the associated autonomy violations and design factors. We analyse the remaining eight design practices based on the original interpretations of autonomy violations from Santos et al. [74] (see §2.2.1) and design factors we identified in Section §4.2 respectively. This analysis was done independently by one of the authors and verified by two other authors, and we held discrepancy discussions until consensus.

Table 2 presents the final mapping of these 11 ACDPs to the Gray et al. [47] ontology, as well as to the autonomy violations and design factors. As we analyse the definitions of the 11 ACDPs from Monge Roffarello et al. [66], we find slight mismatches in the autonomy violations and design factors mappings between Table 2 and our previous results in Table 1. For example, according to Monge Roffarello et al. [66], the *Grinding* dark pattern is defined as *"forc[ing users] to repeat the same process several times to unlock an achievement..."* Therefore, in Table 2, it maps to distortion/impairment, constraining user choices by making the process of unlocking an achievement difficult or effortful (*Choice Effort* factor). However, the definition of *Grinding* in the Gray et al. [47] ontology contains the additional aspect in which a user can *"mak[e] unwanted additional in-app purchases to unlock the same functionality without 'grinding'... "* Therefore, in Table 1, *Grinding* also maps to the *Choice Availability* factor, wherein a user is forced to make in-app purchases to avoid grinding for extended periods of time. Similarly, wherever the Monge Roffarello et al. [66] ACDPs extend the ontology as new low-level patterns, they may not always match with the autonomy violations of existing meso-level patterns, again depending on the definitions of each dark pattern. Hence, in practical contexts, the mapping of any given design practice to autonomy violations and design factors will depend on the exact implementation of the design.

Table 2. Attention capture damaging patterns and their definitions from Monge Roffarello et al. [66] mapped to the Gray et al. [47] ontology of dark patterns and to the autonomy violations and design factors. The superscript in the "Ontology Mapping" column represents the level of the pattern ("M" for meso- and "L" for low-level patterns).

Design Practice	Definition	Ontology Mapping	Autonomy Violations and Design Factors
Infinite Scroll	As the user scrolls down a page, more content automatically and continuously loads at the bottom.	New low-level of *Attention Capture*^{M}	manipulation *Choice Simplification* *CS* : Steers users to watch more content by creating a frictionless scrolling experience which does not allow reflection over time spent

[7]"[W]hereas addictive practices have been empirically studied and include design features such as *'infinite scroll'*, *'pull-to-refresh'* page reload, *'never ending auto-play'* video features, *personalised recommendations, 'recapture notifications'*, meaning notifications to regain users' attention after leaving a service or app, *'playing by appointment'* at certain moments during the day, design leading to *'time fog'* causing a diluted perception of time or *'fake social notifications'* creating the illusion of updates within the user's online social circle..." [72, par. L].

Design Practice	Definition	Ontology Mapping	Autonomy Violations and Design Factors
Casino Pull-to-refresh	When the user swipes down on their smartphone, there is an animated reload of the page that may or may not reveal new appealing content.	New low-level of *Attention Capture*^{M}	manipulation *Information Framing* IF: Steers users to spend more time by framing the upcoming content as 'novel' using animated reload *Choice Simplicification* CS: Steers users to spend more time by creating a frictionless refreshing experience which does not allow reflection over time spent
Neverending Autoplay	A new video is automatically played when the current one finishes. There is never a point for the user to stop and refect, and the option to turn of autoplay is hidden or non-existent.	Mathes with existing low-level *Auto-Play*^{L}	manipulation *Choice Simplicification* CS: Steers users to watch more videos by creating a frictionless viewing experience which does not allow reflection over time spent ***Additional autonomy violation:*** distortion/impairment *Choice Availability* CA, *Choice Effort* CE: If it constrains user choices by making it impossible or difficult to turn off auto-play deception *Information Availability* IA: If it omits or hides the option to turn off auto-play
Guilty Pleasure Recommendations	Personalized suggestions that prey on individual consumer frailty to target user's guilty pleasures and increase use time.	New low-level of *Personalization*^{M}	manipulation *Choice Presentation* CP: Steers user choices by prioritizing presentation of personalized suggestions targeting users' guilty pleasures while potentially demoting other content
Disguised Ads and Recommendations	Advertisements and recommendations, e.g., posts and sponsored pages, that are disguised as normal content into social networks' newsfeeds.	Matches with existing low-level *Disguised Ad*^{L}	deception or manipulation *Information Availability* IA, *Information Correctness* IC: If it omits the marking of advertisements and recommendations or provides false or misleading markings *Information Presentation* IP: If it steers users to engage with advertisements and recommendations by not marking them clearly
Recapture Notifications	Notifications that are deliberately sent to recapture users' attention and have them start a new usage session, e.g., notifications with recommended content or notifications about content the user has never interacted with.	New low-level of *Attention Capture*^{M}	manipulation *Choice Simplicification* CS: Steers users to start a usage session after receiving notifications without due reflection or deliberation ***Additional autonomy violation:*** distortion/impairment *Choice Availability* CA, *Choice Effort* CE: If it constrains user choices by making it impossible or difficult to turn off notifications

Design Practice	Definition	Ontology Mapping	Autonomy Violations and Design Factors
Playing by Appointment	Users are forced to use a digital service at specific times as defined by the service, otherwise the user may lose points and achievements.	New low-level of *Gamification*[M]	distortion/impairment *Choice Availability* CA: Constrains user choices by forcing them to use a service at specific times to keep points or achievements
Grinding	Users are forced to repeat the same process several times to unlock an achievement, e.g., a new level in a video game or a badge on a social network.	Matches with existing low-level *Grinding*[L]	distortion/impairment *Choice Effort* CE: Constrains user choices by making the process of unlocking an achievement difficult or effortful
Attentional Roach Motel	Registering to and accessing attention-capture digital services is easy, while operations like logout or canceling an account are painfully difficult.	New low-level of *Roach Motel*[M] partly matches with *Immortal Accounts*[L]	distortion/impairment *Choice Effort* CE: Constrains user choices by making it difficult to logout of a service or to cancel an account
Time Fog	A pattern through which designers reduce users' awareness of time spent, e.g., by hiding the smartphone's clock.	New low-level of *Attention Capture*[M]	manipulation *Choice Simplification* CS: Steers users to spend more time through lack of time cues which create a frictionless user experience, preventing reflection over time spent
Fake Social Notifications	The platform sends messages pretending to be another user or social notifications about some content the user has never interacted with.	New low-level of *Attention Capture*[M]	deception *Information Correctness* IC: Provides false or misleading messages or notifications pretending to be from another user ***Additional autonomy violation:*** distortion/impairment *Choice Availability* CA, *Choice Effort* CE: If it constrains user choices by making it impossible or difficult to turn off notifications

5.2 *In-Situ* Case Study: X's Paid Blue Checkmarks

To demonstrate how our framework directly supports legal analysis, particularly in ongoing or former enforcement proceedings, we present a case study of the paid blue checkmarks on X (formerly known as "Twitter") . X presently faces enforcement proceedings[8] under Article 25 of the DSA. This case study illustrates how our framework may be immediately used to identify autonomy concerns and their contributing factors in current cases.

5.2.1 Case Description. Formal proceedings of the European Commission were initiated in 2023 against the known online platform X due to the suspected manipulative design of the interface, particularly with regard to X's use of their "blue check" icons in user profiles [28]. Typically, blue checkmarks signify 'verified accounts', i.e., accounts confirmed by the platform to be authentic. However, in April 2023, X allowed any user to subscribe to an account that obtain a 'verified' status by paying a subscription fee [83]. In July 2024, the European Commission stated that: "... X designs and operates its interface for the "verified accounts" with the "Blue checkmark" in a way that does not correspond to

[8]Enforcement proceedings indicates that a regulatory authority has initiated formal measures to ensure compliance with the law.

industry practice and *deceives* users. Since anyone can subscribe to obtain such a "verified" status, it *negatively affects users' ability to make free and informed decisions about the authenticity of the accounts and the content they interact with.* There is evidence of motivated malicious actors abusing the "verified account" to *deceive* users [*emphasis ours*]"[29].

5.2.2 Applying our design factor framework. This legal case analyses an observed design practice, the paid "verified account" indicated by a "blue checkmark". This practice clearly maps to the dark pattern *Parasocial Pressure* in the Gray et al. [47] ontology, as the blue checkmark indicates that the account has been verified by a trusted entity (X in this case). According to our mapping (see Table 1 and Appendix §7.1), this design practice can constitute both manipulation and deception. It can constitute manipulation, as users' choices and decisions can be steered by verified accounts enabled by the blue checkmark, which frames such accounts as 'trustworthy' (factor *Information Framing* *IF*), as well as placement and aesthetics of such blue checkmarks (factor *Information Presentation* *IP*). However, when verified accounts are not genuinely authentic, as in the case of a paid account, this practice constitutes deception as such endorsement is false or misleading (factor *Information Correctness* *IC*). In the case of X, we argue that the design practice of verified accounts is problematic precisely because of this deception, which is the subject of the European Commission's investigation.

6 Discussion

We consider our framework to be a tool both for design and HCI stakeholders, as well as legal and regulatory stakeholders. We now discuss both the design and legally-oriented implications of our resulting framework, then discuss limitations and opportunities for future work.

6.1 Design Implications

As a design tool, our framework can help both design practitioners and their stakeholders understand what design elements contribute to dark pattern deployment, as well as contribute to regulation-informed design or even empower practitioners and scholars seeking legal engagement or improved compliance.

Articulating Design Differences. In this work, we mapped known dark patterns from Gray et al.'s [47] ontology to the three autonomy violation type(s) from DSA Article 25 [74], identifying 8 design factors and categorising them into two design spaces – *Information Space* and *Choice Space* – that can help determine these autonomy violations. In this work, we consider our identified deception factors (*Information Availability* and *Information Correctness*) as exclusive to the information space, as deception alters user perceptions of truth, and is indeed facilitated by alterations of information provided (or not provided) the user. On the other hand, we consider both distortion/impairment factors (*Choice Availability* and *Choice Effort*) as exclusive to the choice space, as distortion/impairment is indeed about constraints on the set of available choices for the user. The four manipulation factors (*Information Framing, Information Presentation, Choice Simplification and Choice Presentation*) are distributed in both spaces. This is because manipulation is often interpreted and understood as a *steering effect in absence of* outright deception or distortion/impairment (see § 2.2.1), which are relatively stronger violations. In both spaces, any influences which do not surpass the threshold of these two violation types can potentially be manipulative. Therefore, we posit that deception and manipulation can be visualised on a spectrum in the information space, with the stronger violations being deceptive, and the weaker violations being manipulative. There may be real world design practices which may not be clear cut between the two, such as when information is presented in a way that it is nearly imperceptible to the user. In such cases, the decision of the violation type(s) would depend on the exact implementation of the design, as well as how it is likely to be

experienced by an average or reasonable user. Similarly, distortion/impairment and manipulation can be visualised on a spectrum in the choice space, with the stronger violations distorting or impairing users' choice set and the weaker violations being less restrictive, i.e., manipulative in nature.

Regulation-Informed Design. By operationalising the DSA text and legal scholarship's [74] distinction of three autonomy violation types – *deception, manipulation* and *distortion/impairment*, our framework can feed forward into HCI literature on dark patterns, unifying the understanding of what types of designs constitute dark patterns. As the three autonomy violations are broadly formulated and open to interpretation (even from different disciplines such as HCI, philosophy and law), the design factors help articulate which design choices or elements contribute to different types of autonomy problems. In this sense, these factors make an *evaluative* contribution to dark patterns literature, acting as an indicator and helping to reason about autonomy violations. However, designers can also take a *generative* lens using these factors to identify alternate design practices or countermeasures to dark patterns. For example, a deception violation that occurs due to *Information Availability* factor can be mitigated by making relevant information available and more explicit. Similarly, a manipulation violation that occurs due to *Choice Simplicification* factor can be mitigated by introducing friction in choices where due deliberation would benefit users. As such, this framework provides a toolkit for regulation-informed design against the deployment of dark patterns. Teams seeking compliance-by-design may take the generative approach earlier in design stages, or evaluate extant interfaces using this framework without needing to force designers to interpret legal texts on their own. This framework thus provides guidance that translates concepts for immediate, practical application.

Practitioners could also use this framework to leverage ongoing or decided enforcement cases against competitors. For example, in the US Federal Trade Commission's case against Fortnite game developer Epic Games, employees were cited as having raised concerns over dark patterns, but executives and managers "feared" a reduction in impulse purchases [23, para. 35]). Such cases intend for a deterrent effect: competitors are meant to take notice and adjust their own behavior accordingly. In this regard, we turn to Richmond Wong's work on user experience (UX) professionals' tactics of "soft resistance" in values work, which notes that design practitioners' values may be "disputed by other organizational stakeholders" [82] . Wong [82] calls for, among other things, tools and structures that "place responsibilities for values in collectives beyond these individual UX professionals" and also consider that such practitioners may experience different levels of internal empowerment. We present our framework as a tool from a broader interdisciplinary collective, as a regulation-informed resource that UX professionals might direct their organizational stakeholders to when challenging anti-autonomy practices.

Supporting Design and HCI Experts in Formal Proceedings. Our framework supports increasing interdisciplinary engagement, with particular benefits for expert testimony. Some dark patterns scholars in design and HCI have already provided expert testimony in important dark patterns cases [41, 56], as have legal scholars [52]. In dark patterns cases, opposing legal teams may refer to common stances in industry [60], which argue that dark patterns are simply common designs or are beneficial to users (thus, such patterns should not be called "sludges" and instead be considered the more benign or even desirable "nudges" [79]). Our systematic mapping of designs to autonomy violations from codified law may help experts push back on such objections. That is, structured frameworks like the Gray et al. [47] ontology that are explicitly grounded in legal source material assert the robustness that non-academics (especially those not practiced in qualitative or design work) sometimes fail to prioritise or recognise. This strengthening, both as a tool to use or a resource to reference, can facilitate experts' arguments against opposition, as well as help designers push back internally on unethical managerial requests.

Compliance-minded or organizationally empowered practitioners could similarly use the framework directly and apply our mappings to the designs implicated in current or future cases against companies within their market (as demonstrated in § 5.2.2), then communicating identified autonomy violations to relevant organisational stakeholders.

6.2 Legal Implications

Informed by legal scholarship and the DSA, this study contributes to dark pattern regulation, enforcement and policymaking in the following three areas.

Compound Effects and Cumulative Impact. Though all autonomy violations result from a single provision in the EU DSA (Article 25), our findings note that 42 out of 59 meso- and low-level dark patterns map to *multiple* potential autonomy violations (§ 4.1.2). Triggering multiple violations can have several implications. In particular, patterns deployed by platforms with multiple violation types inherently carry *higher risk* upon infringing user autonomy. That is, more avenues for a problem to occur leads to higher risk. This finding has immediate relevance to DSA Article 34, concerning platform's risk assessment obligation: multiple potential autonomy violations could impact the systemic risks[9] that the DSA seeks to prevent, and should be further considered in platforms' mandatory risk assessments.

Further, while our study does not measure the harm resulting from a dark pattern autonomy violation, we do note that multiple-violation mappings may still have a potentially amplifying effect on resultant harms [75] and overall severity [61]. Future research could investigate and, where relevant, measure whether dark patterns that operate in multiple ways are more effective, as users may be more susceptible or vulnerable, although this vulnerability depends on the specific practice and context. The occurrence of multiple violations may also serve as an indicator of intention or purpose, helping to demonstrate whether the use of a dark pattern is deliberate (see [74] on the distinction between purpose and effect). Therefore, we recommend that enforcement of DSA Article 25 takes into account the cumulative impact of multiple autonomy violation types within a single design, or when multiple dark patterns are used in a single interface and exacerbate autonomy violative potential.

Operationalizing Laws for Enforcement. Recently enacted EU laws like the DSA, DMA, Data and AI Act emphasize the purpose of protecting user autonomy in their provisions targeting dark patterns and related design practices. Although extant regulation, and DSA Article 25 in particular, already establish a legal basis for prohibiting dark patterns, abstract or broad formulations can leave open questions about which concrete practices impact users autonomy and thus fall within its jurisdictional scope. This paper translates Article 25's abstract autonomy violations into eight design-relevant factors (within a shared vocabulary) that connects design features to legal language. These design factors (§ 4.2) provide regulators with objective evidentiary indicators that can shape how investigators and auditors collect evidence (e.g., documenting missing information or measuring choice effort), and for platform compliance checks. Therefore, the framework offers enforcers structured guidance for detecting and addressing dark patterns, while clarifying compliance obligations for the platforms that implement them. This approach supports ongoing legal cases under the DSA (§ 2.2.2) and can also inform future enforcement efforts.

Informing Provisions in Future Regulation. Among other objectives, the EU's forthcoming (in-draft) Digital Fairness Act (DFA) aims to address "unsolved unethical techniques and commercial practices related to dark patterns, and the addictive design of digital products" [21]. Addictive designs are widely recognized as problematic across multiple

[9]Under Article 34 of the DSA, systemic risks consist of risks stemming from the design or functioning of online platforms, particularly those that may negatively affect the exercise of fundamental rights, including, dignity, protection of personal data, private life, freedom of expression and information, and high-level of consumer protection.

domains, including in the HCI research community [66], by independent NGOs [10], and in the legal field (e.g. in an EU Parliament report [72] and a proposed US Senate bill [53], among others worldwide). We consider this to be a timely opportunity for this work: in this drafting period, our framework could inform dark pattern provisions or recitals in the DFA, help policymakers ensure the DFA's compatibility with prior rules' dark patterns provisions, and serve as a tool for reasoning through dark patterns in the DFA's other interest areas (e.g. harm in video games). Specifically, the EU Commission could refer to mappings directly (from Table 1) or follow the example of attention-capturing design practices (provided in Section § 5.1) to articulate which dark patterns or practices should be presented in the final text.

6.3 Limitations and Future Work

Scope. Regarding jurisdictional scope, the applicability of the proposed framework is anchored in the legal interpretation of the DSA [74] and may not generalise to other non-EU-based autonomy-oriented regulations or international jurisdictions. Research, particularly across jurisdictions and cultures, is needed to better understand how different users or jurisdictions approach or perceive autonomy in relation to dark patterns. Moreover, future work could both expand upon the types of autonomy violations induced by dark patterns and the design factors that contribute to them within both HCI research and policy-oriented applications.

We exclusively use the consolidated Gray et al. ontology [47] and do not conduct autonomy violation or design factor mappings for the five high-level pattern categories (due to their level of abstraction), inspecting only the N=59 meso- and low-level patterns, while more dark patterns may be identified in other literature or disciplines. However, we believe this work, grounded in the DSA, offers a springboard for future extensions of our framework, as described in § 5.1. Across jurisdictional and dark pattern scopes, our work presents an early, but novel effort to bridge the technology-legal gap with regards to autonomy problems from dark patterns. Future avenues for research include validating this framework with users, regulators, and/or industry practitioners to ensure that our mappings are practical, relevant, and effective in addressing present challenges in dark patterns compliance.

Design interpretations versus user perception. Design interpretations vis-à-vis user perception may not perfectly align with our inferred autonomy violations, as users may still experience a sense of deception from manipulation or distortion patterns, and vice versa. This may be for variety of reasons, including having an overbroad definition of each autonomy violation without the granularity favored by legal and policy-oriented experts, personal distinctions or interpretations, and more. Our work does not explore the resultant consumer perception of each dark pattern's autonomy violation; future scholarship might conduct user studies or surveys to see whether our mappings align with firsthand user perceptions.

Additional context beyond definitional constraints. For consistency in our methodology, we interpreted the Gray et al. [47] ontology's dark pattern definitions strictly and did not presume other situations beyond what the definitional text provided. As such, our work remains closely aligned to the ontological definition. However, the nature of design and dark patterns–and the flexibility of the ontology– do not capture all possible dark pattern types, cases, or situations. In some respects, this contributed to autonomy violation assignments between a meso-level pattern and its corresponding low-level patterns that may not match 1:1, as explained in §4.3. Future work might take more expansive interpretations to capture a broader range of autonomy violations, or possibly to sub-type dark patterns to greater detail.

Acknowledgments

This work has been supported by the ANR 22-PECY-0002 IPoP (Interdisciplinary Project on Privacy) project of the Cybersecurity PEPR, Inria DATA4US Exploratory Action project, and the Inria International Chair funding.

References

[1] California Consumer Privacy Act. 2020. California Consumer Privacy Act (Final Text of Proposed Regulations). https://www.oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf

[2] Colorado Privacy Act. 2021. Colorado Privacy Act of 2021. https://leg.colorado.gov/sites/default/files/2021a_190_signed.pdf

[3] California Privacy Rights Act. 2020. California Privacy Rights Act. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5

[4] Sanju Ahuja and Jyoti Kumar. 2022. Conceptualizations of user autonomy within the normative evaluation of dark patterns. *Ethics and Information Technology* 24, Article 52 (2022). doi:10.1007/s10676-022-09672-9

[5] Sanju Ahuja and Jyoti Kumar. 2024. Layered Analysis of Persuasive Designs: A Framework for Identification and Autonomy Evaluation of Dark Patterns. In *Proceedings of the Mobilizing Research and Regulatory Action on Dark Patterns and Deceptive Design Practices (DDPCHI 2024) Workshop at CHI conference on Human Factors in Computing Systems* (Honolulu, HI, USA). Article 1, 14 pages. https://ceur-ws.org/Vol-3720/paper1.pdf

[6] AI Act 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (AI Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj OJ L 2024/1689, 12.7.2024, p. 1–151.

[7] BEUC – The European Consumer Organisation. 2024. Taming Temu: Why the fast-growing online marketplace fails to comply with the EU Digital Services Act. https://www.beuc.eu/enforcement/taming-temu Complaint filed under the Digital Services Act; includes press release, report, annex, and more.

[8] BEUC – The European Consumer Organisation. 2025. Consumer groups file complaint against SHEIN for dark patterns fuelling over-consumption. https://www.beuc.eu/press-release/consumer-groups-file-complaint-against-shein-dark-patterns-fuelling-over-consumption Press release filed with the European Commission and consumer protection authorities; reference BEUC-PR-2025-023..

[9] Nataliia Bielova, Laura Litvine, Anysia Nguyen, Mariam Chammat, Vincent Toubiana, and Estelle Hary. 2024. The effect of design patterns on (present and future) cookie consent decisions. In *Proceedings of the 33rd USENIX Conference on Security Symposium* (Philadelphia, PA, USA) *(SEC '24)*. USENIX Association, USA, Article 158, 18 pages.

[10] Bits of Freedom. 2025. *Investigation: Platforms still use manipulative design despite DSA rules.* https://dsa-observatory.eu/2025/08/07/investigation-platforms-still-use-manipulative-design-despite-dsa-rules/

[11] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies* 2016, 4 (2016), 237–254. doi:10.1515/popets-2016-0038

[12] Elijah Bouma-Sims, Megan Li, Yanzi Lin, Adia Sakura-Lemessy, Alexandra Nisenoff, Ellie Young, Eleanor Birrell, Lorrie Faith Cranor, and Hana Habib. 2023. A US-UK Usability Evaluation of Consent Management Platform Cookie Consent Interface Design on Desktop and Mobile. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI'23)*. Association for Computing Machinery.

[13] Martin Brencke. 2024. A Theory of Exploitation for Consumer Law: Online Choice Architectures, Dark Patterns, and Autonomy Violations. *Journal of Consumer Policy* 47 (2024), 156.

[14] Harry Brignull. 2010. *What are deceptive patterns?* Retrieved February 13, 2025 from http://darkpatterns.org/

[15] Harry Brignull. 2018. Deceptive Patterns: User Interfaces Designed to Trick People. http://darkpatterns.org/

[16] Harry Brignull. 2023. Deceptive Patterns. https://www.deceptive.design

[17] Akash Chaudhary, Jaivrat Saroha, Kyzyl Monteiro, Angus G. Forbes, and Aman Parnami. 2022. "Are You Still Watching?": Exploring Unintended User Behaviors and Dark Patterns on Video Streaming Platforms. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 776–791. doi:10.1145/3532106.3533562

[18] CMA 2022. *Evidence review of Online Choice Architecture and consumer and competition harm.* Technical Report. https://www.gov.uk/government/publications/online-choice-architecture-how-digital-design-can-harm-competition-and-consumers/evidence-review-of-online-choice-architecture-and-consumer-and-competition-harm

[19] Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20 (1960), 37–46. Issue 1. doi:10.1177/001316446002000104

[20] European Commission. 2024. Formal Proceedings Against Meta under the Digital Services Act. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2664 Press release, 30 April 2024. Ongoing investigation; no final decision issued..

[21] European Commission. 2025. Questions and Answers on the Digital Fairness Act. https://ec.europa.eu/commission/presscorner/detail/en/qanda_24_4909 Press corner Q&A on the Digital Fairness Act, accessed 2025-09-09.

[22] Federal Trade Commission. 1914. Section 5 of the US FTC Act (15 U.S.C. §45).

[23] Federal Trade Commission. 2022. Epic Games, In the Matter of. File No. 1923203. [Complaint]. https://www.ftc.gov/system/files/ftc_gov/pdf/1923203EpicGamesComplaint.pdf

[24] Data Act 2023. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (Data Act). https://eur-lex.europa.eu/eli/reg/2023/2854/oj OJ L 2023/2854, 22.12.2023, p. 1–64.

[25] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376600

[26] DMA 2022. Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act). https://eur-lex.europa.eu/eli/reg/2022/1925/oj OJ L 265, 12.10.2022, p. 1–66.

[27] DSA 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). http://data.europa.eu/eli/reg/2022/2065/oj

[28] European Commission 2023. Commission opens formal proceedings against X under the Digital Services Act. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709

[29] European Commission 2024. Commission sends preliminary findings to X for breach of the Digital Services Act. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761

[30] European Commission. 2025. Commission and national authorities urge SHEIN to respect EU consumer protection laws. https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1331 Press Release IP_25_1331.

[31] European Data Protection Board. 2023. *Guidelines 3/2022 on Dark patterns in social media platform interfaces: How to recognise and avoid them.* Technical Report Version 2.0. https://edpb.europa.eu/system/files/2023-02/edpb_03-2022_guidelines_on_deceptive_design_patterns_in_social_media_platform_interfaces_v2_en_0.pdf

[32] Federal Trade Commission. 2022. FTC Action Against Vonage Results in $100 Million to Customers Trapped by Illegal Dark Patterns and Junk Fees When Trying to Cancel Service. https://www.ftc.gov/news-events/news/press-releases/2022/11/ftc-action-against-vonage-results-100-million-customers-trapped-illegal-dark-patterns-junk-fees-when-trying-cancel-service Accessed: 2025-09-10.

[33] Federal Trade Commission. 2023. FTC Finalizes Order Requiring Fortnite Maker Epic Games to Pay $245 Million for Tricking Users into Making Unwanted Charges. https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-finalizes-order-requiring-fortnite-maker-epic-games-pay-245-million-tricking-users-making Accessed: 2025-09-10.

[34] Federal Trade Commission. 2023. FTC Takes Action Against Amazon for Enrolling Consumers in Amazon Prime Without Consent and Sabotaging Their Attempts to Cancel. https://www.ftc.gov/news-events/news/press-releases/2023/06/ftc-takes-action-against-amazon-enrolling-consumers-amazon-prime-without-consent-sabotaging-their Accessed: 2025-09-10.

[35] Carlos Bermejo Fernandez, Dimitris Chatzopoulos, Dimitrios Papadopoulos, and Pan Hui. 2021. This Website Uses Nudging: MTurk Workers' Behaviour on Cookie Consent Notices. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2, Article 346 (October 2021), 22 pages. doi:10.1145/3476087

[36] Dan Fitton and Janet C. Read. 2019. Creating a Framework to Support the Critical Consideration of Dark Design Aspects in Free-to-Play Apps. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19)*. Association for Computing Machinery, New York, NY, USA, 407–418. doi:10.1145/3311927.3323136

[37] FTC 2022. *Bringing Dark Patterns to Light Staff Report.* Technical Report. Federal Trade Commission. https://www.ftc.gov/system/files/ftc_gov/pdf/P214800%20Dark%20Patterns%20Report%209.14.2022%20-%20FINAL.pdf

[38] Garante per la protezione dei dati personali. 2023. Provvedimento prescrittivo e sanzionatorio nei confronti di Ediscom S.p.A. https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870014 Doc. web n. 9870014.

[39] Maximilian Gartner. 2022. Regulatory Acknowledgment of Individual Autonomy in European Digital Legislation: From Meta-Principle to Explicit Protection in the Data Act. *European Data Protection Law Review* 8, 4 (2022). doi:10.21552/edpl/2022/4/6

[40] GDPR 2016. 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 (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj OJ L 119, 4.5.2016, p. 1–88.

[41] Colin M Gray. 2022. Expert Report of Colin M. Gray, Ph.D. Public Redacted Version (State of Arizona v. Google LLC). https://www.azag.gov/sites/default/files/2022-09/Expert%20Report%20of%20Colin%20M.%20Gray%2C%20Ph.D..pdf

[42] Colin M. Gray, Lorena Sanchez Chamorro, Ike Obi, and Ja-Nae Duane. 2023. Mapping the Landscape of Dark Patterns Scholarship: A Systematic Literature Review. In *Companion Publication of the 2023 ACM Designing Interactive Systems Conference (DIS '23 Companion)*. Association for Computing Machinery, 188–193. doi:10.1145/3563703.3596635

[43] Colin M. Gray, Johanna T. Gunawan, René Schäfer, Nataliia Bielova, Lorena Sanchez Chamorro, Katie Seaborn, Thomas Mildner, and Hauke Sandhaus. 2024. Mobilizing Research and Regulatory Action on Dark Patterns and Deceptive Design Practices. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 482, 6 pages. doi:10.1145/3613905.3636310

[44] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3173574.3174108

[45] Colin M. Gray, Thomas Mildner, and Ritika Gairola. 2025. Getting Trapped in Amazon's "Iliad Flow": A Foundation for the Temporal Analysis of Dark Patterns. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 225, 10 pages. doi:10.1145/3706598.3713828

[46] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 172, 18 pages. doi:10.1145/3411764.3445779

[47] Colin M. Gray, Cristiana Teixeira Santos, Nataliia Bielova, and Thomas Mildner. 2024. An Ontology of Dark Patterns Knowledge: Foundations, Definitions, and a Pathway for Shared Knowledge-Building. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 289, 22 pages. doi:10.1145/3613904.3642436

[48] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. 2021. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research* 3, 1 (Feb. 2021), 1–38. doi:10.33621/jdsr.v3i1.54

[49] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. 2021. A Comparative Study of Dark Patterns Across Web and Mobile Modalities. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 377 (Oct. 2021), 29 pages. doi:10.1145/3479521

[50] Johanna Gunawan, Cristiana Santos, and Irene Kamara. 2022. Redress for Dark Patterns Privacy Harms? A Case Study on Consent Interactions. In *Proceedings of the 2022 Symposium on Computer Science and Law* (Washington DC, USA) *(CSLAW '22)*. Association for Computing Machinery, New York, NY, USA, 181–194. doi:10.1145/3511265.3550448

[51] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, Whatever": An Evaluation of Cookie Consent Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 621, 27 pages. doi:10.1145/3491102.3501985

[52] Woodrow Hartzog. 2023. Prepared Testimony and Statement for the Record. https://www.judiciary.senate.gov/imo/media/doc/2023-09-12_pm_-_testimony_-_hartzog.pdf

[53] Josh Hawley. 2020. 116th Congress, 1st Session – Social Media Addiction Reduction Technology Act. https://www.hawley.senate.gov/wp-content/uploads/files/2019-07/Social-Media-Addiction-Reduction-Technology-Act.pdf

[54] Natali Helberger, Betül Kas, Hans W. Micklitz, Monika Namysłowska, Laurens Naudts, Peter Rott, Marijn Sax, and Michael Veale. 2024. *Digital Fairness for Consumers*. Technical Report. BEUC. https://www.beuc.eu/sites/default/files/publications/BEUC-X-2024-032_Digital_fairness_for_consumers_Report.pdf

[55] Information Commissioner's Office and Competition and Markets Authority. 2023. Harmful Design in Digital Markets: How Online Choice Architecture Practices Can Undermine Consumer Choice and Control over Personal Information. https://www.drcf.org.uk/siteassets/drcf/pdf-files/harmful-design-in-digital-markets-ico-cma-joint-position-paper.pdf?v=380506 Accessed: 2025-09-10.

[56] Jennifer King and Adriana Stephan. 2021. Regulating Privacy Dark Patterns in Practice - Drawing Inspiration from the California Privacy Rights Act. *Georgetown Law Technology Review* 5 (2021), 26 pages. Issue 250.

[57] Mark Leiser and Cristiana Santos. 2024. Dark Patterns, Enforcement, and the Emerging Digital Design Acquis: Manipulation beneath the Interface. *European Journal of Law and Technology* 15, 1 (2024). https://ejlt.org/index.php/ejlt/article/view/990/1084 BILETA Special Issue.

[58] Jamie Luguri and Lior Jacob Strahilevitz. 2021. Shining a Light on Dark Patterns. *Journal of Legal Analysis* 13, 1 (March 2021), 43–109. doi:10.1093/jla/laaa006

[59] Francisco Lupiáñez-Villanueva, Alba Boluda, Francesco Bogliacino, Giovanni Liva, Lucie Lechardoy, and Teresa Rodríguez de las Heras Ballell. 2022. *Behavioural study on unfair commercial practices in the digital environment : dark patterns and manipulative personalisation : final report*. Publications Office of the European Union, Brussels, Belgium. doi:10.2838/859030

[60] Daniel Luque. 2023. Examining the FTC's Hostility to Common Design Practices. *Disruptive Competition Project* (2023). https://project-disco.org/competition/examining-the-ftcs-hostility-to-common-design-practices/

[61] Gianclaudio Malgieri and Cristiana Santos. 2025. Assessing the (severity of) impacts on fundamental rights. *Computer Law & Security Review* 56 (2025), 106113. doi:10.1016/j.clsr.2025.106113

[62] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 81 (Nov. 2019), 32 pages. doi:10.1145/3359183

[63] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 360, 18 pages. doi:10.1145/3411764.3445610

[64] Thomas Mildner, Merle Freye, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2362–2374. doi:10.1145/3563657.3595964

[65] Thomas Mildner, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 192, 15 pages. doi:10.1145/3544548.3580695

[66] Alberto Monge Roffarello, Kai Lukoff, and Luigi De Russis. 2023. Defining and Identifying Attention Capture Deceptive Designs in Digital Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing

Machinery, New York, NY, USA, Article 194, 19 pages. doi:10.1145/3544548.3580729

[67] Carol Moser, Sarita Y. Schoenebeck, and Paul Resnick. 2019. Impulse Buying: Design Practices and Consumer Needs. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3290605.3300472

[68] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *CHI*.

[69] OECD. 2022. *Dark commercial patterns*. Technical Report. doi:10.1787/44f5e846-en

[70] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. 2022. Exploring Deceptive Design Patterns in Voice Interfaces. In *Proceedings of the 2022 European Symposium on Usable Security* (Karlsruhe, Germany) *(EuroUSEC '22)*. Association for Computing Machinery, New York, NY, USA, 64–78. doi:10.1145/3549015.3554213

[71] European Parliament. 2023. European Parliament Resolution of 12 December 2023 on Addictive Design of Online Services and Consumer Protection in the EU Single Market (2023/2043(INI)). https://www.europarl.europa.eu/doceo/document/TA-9-2023-0459_EN.html Resolution adopted 12 December 2023, based on report A9-0340/2023.

[72] European Parliament. 2023. Report on Addictive Design of Online Services and Consumer Protection in the EU Single Market (2023/2043(INI)). https://www.europarl.europa.eu/doceo/document/A-9-2023-0340_EN.html Committee on the Internal Market and Consumer Protection, Rapporteur: Kim van Sparrentak.

[73] Benjamin Raue. 2025. Article 25. In *Digital Services Act Article-by-Article Commentary*, Benjamin Raue and Franz Hofmann (Eds.). Bloomsbury Publishing.

[74] Cristiana Santos, Sanju Ahuja, Nataliia Bielova, and Christine Utz. 2025. Understanding the scope of Article 25 of the DSA in regulating dark patterns. In *Dark patterns and deceptive design patterns: Conceptualising and systematising a key contemporary phenomenon*. Edward Elgar. Forthcoming, Available at SSRN: https://ssrn.com/abstract=4899559.

[75] Cristiana Santos, Viktorija Morozovaite, and Silvia De Conca. 2025. No harm no foul: how harms caused by dark patterns are conceptualised and tackled under EU data protection, consumer and competition laws. *Information & Communications Technology Law* (2025), 1–47. doi:10.1080/13600834.2025.2461958

[76] Brennan Schaffner, Neha A. Lingareddy, and Marshini Chetty. 2022. Understanding Account Deletion and Relevant Dark Patterns on Social Media. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 417 (Nov. 2022), 43 pages. doi:10.1145/3555142

[77] Brennan Schaffner, Yaretzi Ulloa, Riya Sahni, Jiatong Li, Ava Kim Cohen, Natasha Messier, Lan Gao, and Marshini Chetty. 2025. An Experimental Study of Netflix Use and the Effects of Autoplay on Watching Behaviors. *Proceedings of the ACM on Human-Computer Interaction* 9, 2, Article CSCW030 (May 2025), 22 pages. doi:10.1145/3710928

[78] René Schäfer, Sarah Sahabi, Lucia Karl, Sophie Hahn, and Jan Borchers. 2025. "If They Have No Choice, They'll Accept!": How Children and Adolescents Assess Deceptive Designs. In *Proceedings of the 24th Interaction Design and Children (IDC '25)*. Association for Computing Machinery, 863–871. doi:10.1145/3713043.3731497

[79] Richard H. Thaler. 2018. Nudge, not sludge. *Science* 361, 6401 (2018), 431–431. doi:10.1126/science.aau9241 arXiv:https://www.science.org/doi/pdf/10.1126/science.aau9241

[80] UCPD 2005. 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 (Unfair Commercial Practices Directive). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32005L0029 OJ L 149, 11.6.2005, p. 22–39.

[81] Urząd Ochrony Konkurencji i Konsumentów (UOKiK). 2024. 31 mln zł kary dla Amazon. https://uokik.gov.pl/31-mln-zl-kary-dla-amazon Decision imposing a fine of 31 850 141 zł on Amazon EU SARL for misleading consumer practices, including dark patterns regarding product availability, delivery times, and Guaranteed Delivery service..

[82] Richmond Y. Wong. 2021. Tactics of Soft Resistance in User Experience Professionals' Values Work. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 355 (Oct. 2021), 28 pages. doi:10.1145/3479499

[83] X 2023. How to get the blue checkmark on X. https://help.x.com/en/managing-your-account/about-x-verified-accounts

[84] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark patterns in the design of games. In *Foundations of Digital Games 2013*. https://www.diva-portal.org/smash/get/diva2:1043332/FULLTEXT01.pdf

7 Appendix

7.1 Dark patterns autonomy violations and contributing design factors

Below we present the final reasoning behind our autonomy violation labels, including the design factors influencing low-level dark patterns. We take the original definition of each meso- and low-level dark pattern from Gray et al. [47] presented in gray, followed by our reasoning, with each autonomy violation is highlighted in color – such as deception – and design factors are also highlited – such as *CA* (using factor abbreviations from Figure 4). The level of dark pattern (high-, meso- or low-) is indicated by the corresponding

superscript (H, M or L).

OBSTRUCTION^{H}

Roach Motel^{M} subverts the user's expectation that an action will be as easy to reverse as it is to make, instead creating a situation that is easy to get into, but difficult to get out of.

- Roach Motel constitutes distortion/impairment as it constrains user choices by increasing the effort or difficulty of getting out of a situation.

Immortal Accounts^{L} create a *Roach Motel* and use *Obstruction* to make it difficult or impossible to delete a user account once it has been created. As a result, the user may create an account or share data with the false assumption that they can later delete this information, even though that account and/or data are then unable to be removed by the user.

- Immortal Accounts constitutes distortion/impairment as it constrains user choices by omitting the option to delete a user account (thus making this choice unavailable, factor *CA*) or by making it unreasonably difficult and therefore increasing users' choice-making effort (factor *CE*).

- ***Additional autonomy violation:*** There is also a possibility of deception if it omits information (factor *IA*) or provides false or misleading information (factor *IC*) about the possibility of account deletion before account creation.

Dead Ends ^{L} create a *Roach Motel* and use *Obstruction* to prevent users from finding information through inactive links or redirections that limit or completely prevent the display of relevant information. As a result, the user may seek to find relevant information or action possibilities but instead be left unable to achieve their goal.

- Dead Ends constitutes deception and distortion/impairment. There is deception as it provides false or misleading information in the form of links or redirection chains that are inactive in reality (factor *IC*), and thus, it omits information or controls by preventing users from finding them (factor *IA*). There is distortion/impairment as it constrains user choices by increasing the effort required to find relevant information or controls (factor *CE*) or by omitting relevant controls entirely (factor *CA*).

Creating Barriers^{M} subverts the user's expectation that relevant user tasks will be supported by the interface, instead preventing, abstracting, or otherwise complicating a user task to disincentive user action.

- Creating Barriers constitutes distortion/impairment and/or manipulation. There is distortion/impairment if it constrains user choices by preventing certain tasks or by making them unreasonably difficult or complicated. There is manipulation if it steers user choices by preferring or prioritizing certain tasks in the interface, without any constraints or restrictions.

Price Comparison Prevention^{L} *Creates Barriers* and uses *Obstruction* by excluding relevant information, limiting the ability of a user to copy/paste, or otherwise inhibiting a user from comparing prices across two or more vendors. As a result, the user cannot make an informed decision about where to buy a product or service.

- Price Comparison Prevention constitutes distortion/impairment and/or deception. There is distortion/impairment if it constrains user choices by increasing the effort required to compare prices (factor *CE*). There is deception if it omits information required to compare prices (factor *IA*) or if it provides false or misleading information such as incorrect prices or prices which change on the fly (factor *IC*).

Intermediate Currencies^{L} *Create Barriers* and use *Obstruction* to hide the true cost of a product or service by requiring the user to spend real money to purchase a virtual currency that is then used to purchase a product or service. As a result, the user is unable to easily ascertain the true monetary cost of a product or service, leading them to make an uninformed purchase decision based on an obscured cost.

- Intermediate Currencies constitutes distortion/impairment, deception and manipulation. There is distortion/impairment as it constrains user choices because the user cannot avoid purchasing a virtual currency for the purchase of a product or service (factor *CA*). There is deception as it omits information about the true monetary cost of the product at the moment of purchase (factor *IA*). There is manipulation as it steers user choices by framing the cost of a product as a virtualised cost, subverting deliberation or reflection over its true monetary costs (factor *IF*). In addition, it also steers users towards a purchase as virtualised purchases can be executed without the friction of formalised payment methods (factor *CS*).

Adding Steps[M] subverts the user's expectation that a task will take as few steps as technologically needed, instead creating additional points of unnecessary but required user interaction to perform a task.

- Adding Steps constitutes distortion/impairment as it constrains user choices by adding extra and unnecessary steps to complete certain tasks.

- ***Additional autonomy violation:*** There is also a possibility of deception if it provides false or misleading information about how many steps a task will take.

Privacy Mazes[L] *Add Steps* and use *Obstruction* to require a user to navigate through many pages to obtain relevant information or control without a comprehensive and exhaustive overview. As a result, the user is prevented from easily discovering relevant information or action possibilities, leaving them unable to make informed decisions regarding their privacy.

- Privacy Mazes constitutes distortion/impairment as it constrains user choices by increasing the effort to obtain relevant information or controls, requiring users to navigate through many pages (factor *CE*).

- ***Additional autonomy violation:*** There is also a possibility of deception if it omits information or controls by limiting their discoverability within many pages and without a comprehensive overview (factor *IA*).

SNEAKING[H]

Bait & Switch[M] subverts the user's expectation that their choice will result in a desired action, instead leading to an unexpected, undesirable outcome.

- Bait & Switch constitutes deception and distortion/impairment. There is deception as it provides false or misleading information about action possibilities ('bait'). Then, there is distortion/impairment as it constrains user choices by automatically executing an unexpected or undesirable action on their behalf ('switch').

Disguised Ads[L] *Bait and Switch* and use *Sneaking* to style interface elements so they are not clearly marked as an advertisement or other biased source. As a result, users are induced into clicking on the interface element because they assume that it is a relevant and salient interaction, leading to unwitting interaction with advertising content.

- Disguised Ads constitutes deception or manipulation. There is deception if it omits information by not marking advertisements and other biased sources (factor *IA*), or if it provides false or misleading information by incorrectly marking these sources (factor *IC*). There is manipulation if it steers users to engage with these sources by not marking them clearly (factor *IP*).

Hiding Information[M] subverts the user's expectation that all relevant information to make an informed choice will be available to them, instead hiding information or delaying the disclosure of information until later in the user journey that may have led to them making another choice.

- Hiding Information constitutes deception and manipulation. There is deception as it omits relevant information in the beginning of the user journey. Then, there is manipulation as it steers users towards sticking to these decisions made with incomplete information, even when relevant information is disclosed later in the user journey.

Sneak into Basket[L] *Hides Information* and uses *Sneaking* to add unwanted items to a user's shopping cart without their consent. As a result, a user assumes that only the items they explicitly added to their cart will be purchased, leading to unintentional purchase of additional items.

- Sneak into Basket constitutes distortion and deception. There is distortion/impairment as it constrains user choices by adding items to users' cart without any user action, whereas removal of these items necessitates direct action (factor *CA*). There is deception as it omits explicit information that such items have been added (factor *IA*).

- ***Additional autonomy violation:*** There is also a possibility of manipulation if it steers users towards a purchase through last minute presentation of potentially tempting items in the cart (factor *CP*) and the reduced friction in adding them to cart (factor *CS*).

Drip Pricing, Hidden Costs, or Partitioned Pricing[L] *Hides Information* and uses *Sneaking* to reveal new charges or costs, present only partial price components, or otherwise delay revealing the full price of a product or service through late or incomplete disclosure. As a result, the user is misled about the total or complete price of the product or service, leading to them to make a purchase decision after they have expended effort on false pretenses.

- Drip Pricing, Hidden Costs, or Partitioned Pricing constitutes deception and manipulation. There is deception as it provides false or misleading information in the form of a partial or incomplete price of a product at first (factor *IC*), and it omits information about additional cost components (factor *IA*). Then, there is manipulation as it steers users towards sticking to their decisions and making a purchase even after the full price is revealed later in the user journey (factor *IP*).

Reference Pricing[L] *Hides Information* and uses *Sneaking* to include a misleading or inaccurate price for a product or service that makes a discounted price appear more attractive. As a result, the user is misled into believing that the price they pay is discounted, leading them to make a decision to purchase a product or service on false pretenses.

- Reference Pricing constitutes deception and manipulation. There is deception as it provides false or misleading information about the 'original' price of a product or service (factor *IC*). There is manipulation as it steers users towards a purchase, because the false original price frames the 'actual' price as discounted in comparison (factor *IF*), often accompanied by aesthetic or visual cues highlighting this discount (factor *IP*).

(De)contextualising Cues[M] subverts the user's expectation that provided information will guide the user to making an informed choice, instead confusing the user and/or preventing them from locating relevant information due to the context where information is presented.

- (De)contextualising Cues constitutes deception and/or manipulation. There is deception if it omits relevant information by placing it out of context, thus preventing it from being located. There is manipulation if it steers user choices in a particular direction by confusing the user with out of context information.

Conflicting Information[L] uses *(De)contextualizing Cues* and *Sneaking* to include two or more sources of information that conflict with each other. As a result, the user is unsure what the consequences of their actions will be and will be more likely to accept default settings that may not be in their best interest.

- Conflicting Information constitutes deception as it provides false or misleading information in the form of conflicting sources or pieces of information (factor *IC*).

Information without Context[L] uses *(De)contextualizing Cues* and *Sneaking* to alter the relevant information or user controls to limit discoverability. As a result, the user is unlikely to find the information or action possibility they are interested in.

- Information without Context constitutes deception and/or manipulation. There is deception if it omits relevant information or controls by placing them out of context (therefore, information is not available, factor *IA*). There is manipulation if it steers user

choices through the information framing (factor *IF*) or presentation of out of context information (factor *IP*).

INTERFACE INTERFERENCE

Manipulating Choice Architecture[M] subverts the user's expectation that the options presented will support their desired goal, instead including an order or structure of options that makes another outcome more likely

- Manipulating Choice Architecture constitutes manipulation as it steers user choices through interface elements such as order or structure of options, without hiding or restricting user choices.

False Hierarchy[L] *Manipulates the Choice Architecture*, using *Interface Interference* to give one or more options visual or interactive prominence over others, particularly where items should be in parallel rather than hierarchical. As a result, the user may misunderstand or be unable to accurately compare their options, making a selection based on a false or incomplete choice architecture.

- False Hierarchy constitutes manipulation as it steers user choices by presenting some options as more prominent than others (factor *CP*), without hiding or restricting any choices.

Visual Prominence[L] *Manipulates the Choice Architecture*, using *Interface Interference* to place an element relevant to user goals in visual competition with a more distracting and prominent element. As a result, the user may forget about or be distracted from their original goal, even if that goal was their primary intent.

- Visual Prominence constitutes manipulation as it steers user choices through the presentation of distracting and prominent information elements (factor *IP*) or choice elements (factor *CP*), which can potentially lead users away from their original goals.

Bundling[L] *Manipulates the Choice Architecture*, using *Interface Interference* to group two or more products or services in a single package at a special price. As a result, the user may incorrectly assume that these items must be purchased as a bundle or be unaware of the unbundled price for the component elements, possibly leading to an uninformed purchasing decision.

- Bundling constitutes deception and/or manipulation and/or distortion. There is deception if it omits information about the availability or prices of unbundled products and services (factor *IA*). There is manipulation if it steers users towards the bundle using visual or interactive information elements (factor *IP*) or choice elements (factor *CP*), even if the unbundled items are shown in the interface. There is distortion/impairment if it constrains user choices by making it impossible for them to purchase the unbundled items (factor *CA*).

Pressured Selling[L] *Manipulates the Choice Architecture*, using *Interface Interference* to preselect or use visual prominence to focus user attention on more expensive product options. As a result, the user may be unaware that a lower price is available or even desirable for their needs, steering the user into making a more expensive product selection than they otherwise would have.

- Pressured Selling constitutes manipulation as it steers users towards more expensive product options, either through preselection (e.g. an option is presented as "selected" in the interface) (factor *CP*), or through visual prominence (factors *IP*, *CP*, see *Visual Prominence* dark pattern). In the case of preselection, it also steers users through reduced friction in the purchase, subverting a reflective step within the user's decision to purchase the expensive product option (factor *CS*).

- ***Additional autonomy violation:*** There is also a possibility of deception if it omits information about less expensive product options (factor *IA*).

Bad Defaults[M] subverts the user's expectation that default settings will be in their best interest, instead requiring users to take active steps to change settings that may cause harm or unintentional disclosure of information.

- Bad Defaults constitutes distortion/impairment and manipulation. There is distortion/impairment as it constrains user choices by making the user expend unnecessary effort to change default settings. There is manipulation as it steers users towards keeping the default settings, due to a fear of going against the default.

- ***Additional autonomy violation:*** There is also a possibility of deception if it provides false or misleading information that the default settings are in the user's best interest.

Emotional or Sensory Manipulation[M] subverts the user's expectation that the design of the site will allow them to achieve their goal without manipulation, instead altering the language, style, color, or other design elements to evoke an emotion or manipulate the senses in order to persuade the user into a particular action.

- Emotional or Sensory Manipulation constitutes manipulation as it steers users' choices by affecting their emotions and/or perception through language, style, color or other design elements.

Cuteness[L] uses *Emotional or Sensory Manipulation* and *Interface Interference* to embed attractive cues in the design of a robot interface or form factor. As a result, a user may place undue trust in the robot, leading the user to inaccurately or incompletely assess the risks of interacting with the robot.

- Cuteness constitutes manipulation it steers users towards risky choices or options through undue feelings of trust created by the attractive presentation or 'cute' form of the robot interface (factor *IP*) and the framing of interactions through visuals, language or tone (factor *IF*).

Positive or Negative Framing[L] uses *Emotional or Sensory Manipulation* and *Interface Interference* to visually obscure, distract, or persuade a user from important information they need to achieve their goal. As a result, the user may assume that the system is providing equal access to relevant information, leading the user to be distracted by positive or negative aesthetic cues that distract them from important information or action possibilities or otherwise convince them to pursue a different goal.

- Positive or Negative Framing constitutes manipulation as it steers user choices through uses positive or negative linguistic framing (factor *IF*) or aesthetic cues in information presentation (factor *IP*) or choice presentatation (factor *CP*).

Trick Questions[M] subvert the user's expectation that prompts will be written in a straightforward and intelligible manner, instead using confusing wording, double negatives, or otherwise leading language or interface cues to manipulate a user's choice.

- Trick Questions constitutes deception and/or manipulation. There is deception if it provides false or misleading information by framing or wording prompts in a misleading way. There is manipulation if it steers user choices by using confusing or leading language or interface cues.

Choice Overload[M] subverts the user's expectation that the choices they make should be understandable and comparable, instead providing too many options to compare or encouraging users to overlook relevant information due to the volume of choices provided.

- Choice Overload constitutes manipulation as it steers users towards particular choices or options by increasing the volume of available choices, leading users to overlook some options or relevant information about these options.

Hidden Information[M] subverts the user's expectation that relevant information will be made accessible and visible, instead disguising relevant information or framing it as irrelevant.

- Hidden Information constitutes deception as it either omits relevant information or misleadingly frames it as irrelevant.

Language Inaccessibility[M] subverts the user's expectation that guidance will be provided in a way that is understandable and intelligible, instead using unnecessarily complex language or a language not spoken by the user to decrease the likelihood the user will make an informed choice.

- Language Inaccessibility constitutes manipulation and/or deception. There is manipulation if it steers user choices by preventing them from engaging with the information provided in complex or wrong language. There is deception if it provides false or misleading information or omits relevant information through the use of complex or wrong language.

Wrong Language[L] leverages *Language Accessibility*, using *Interface Interference* to provide important information in a different language than the official language of the country where users live. As a result, the user will not have access to relevant information about their interaction with the system and their ability to choose, leading to uninformed decisions.

- Wrong Language constitutes distortion/impairment as it constrains user choices by omitting the option to access relevant information in a language that users understand (factor *CA*).

- ***Additional autonomy violation:*** There is also a possibility of deception if it omits information by hiding it or limiting its discoverability because of the wrong language (factor *IA*).

Complex Language[L] leverages *Language Accessibility*, using *Interface Interference* to make information difficult to understand by using obscure word choices and/or sentence structure. As a result, the user will not be able to comprehend relevant information about their interaction with the system and their ability to choose, leading to uninformed decisions.

- Complex Language constitutes manipulation and/or deception. There is manipulation if it steers user choices by confusing the user or preventing them from engaging with the information provided due to obscure word choices and/or sentence structure (factor *IF*). There is deception if it the complex language is framed such that it provides false or misleading information (factor *IC*).

Feedforward Ambiguity[M] subverts the user's expectation that their choice will be likely to result in an action they can predict, instead providing a discrepancy between information and actions available to users that results in an outcome that is different from what the user expects.

- Feedforward Ambiguity constitutes deception as it provides false or misleading information which is not consistent with the resultant outcome of user actions.

FORCED ACTION

Nagging[M] subverts the user's expectation that they have rational control over the interaction they make with a system, instead distracting the user from a desired task the user is focusing on to induce an action or make a decision the user does not want to make by repeatedly interrupting the user during normal interaction.

- Nagging constitutes distortion/impairment as it constrains user choices by forcing or coercing users through repeated interruptions to take a particular action or make a decision that they do not want to make.

Forced Continuity[M] subverts the user's expectation that a subscription created in the past will not auto-renew or otherwise continue in the future, instead causing undesired charges, difficulty to cancel, or lack of awareness that a subscription is still active.

- Forced Continuity constitutes distortion/impairment as it constrains user choices by automatically renewing a subscription to a service without any user action, accompanied by an additional and often unreasonable effort to cancel such subscription.

- ***Additional autonomy violation:*** There is also a possibility of deception and/or manipulation, if it omits information in advance about the auto-renewal, or if it steers users to choose the auto-renewal option.

Forced Registration[M] subverts the user's expectation that they can complete an action without registering or creating an account, instead tricking them into thinking that registration is required, often resulting in the sharing of unneeded personal data.

- Forced Registration constitutes distortion/impairment and/or deception and/or manipulation. There is distortion/impairment if it constrains user choices by forcing or mandating them to register an account to complete an action. There is deception if it provides false or misleading information to lead a user into thinking that registration is required when it is not. There is manipulation if it

steers users into registering an account through linguistic framing or aesthetic interface elements.

Forced Communication or Disclosure[M] subverts the user's expectation that a system will only request information needed to complete their desired goals, instead tricking them into sharing more information about themselves or using their information for purposes that they do not desire.

- Forced Communication or Disclosure constitutes distortion/impairment and/or deception and/ or manipulation. There is distortion/impairment if it constrains user choices by forcing or mandating them to disclose information to complete their goals, or if the platform uses their information for extra purposes over which they have no control. There is deception if it omits information or provides false or misleading information to lead a user into thinking that sharing more information is required when it is not. There is manipulation if it steers users into sharing more information using linguistic framing or aesthetic interface elements.

Privacy Zuckering[L] uses *Forced Communication or Disclosure* as a type of *Forced Action* to trick users into sharing more information about themselves than they intend to or would agree to if fully informed. As a result, the user assumes that information they are requested to provide is vital for use of the service, even while this information is used or sold for other purposes.

- Privacy Zuckering constitutes distortion/impairment and/or deception. There is distortion/impairment if it constrains user choices by forcing or mandating them to share information for the use of a service, or if the information is used or sold for other purposes (factor *CA*). There is deception if it omits information (factor *IA*) or provides false or misleading information (factor *IC*) to lead a user into thinking that the information collected is vital or essential for the use of the service.

- ***Additional autonomy violation:*** There is also a possibility of manipulation if it steers users into sharing more information using linguistic framing (factor *IF*) or aesthetic interface elements (factor *IP*).

Friend Spam[L] uses *Forced Communication or Disclosure* as a type of *Forced Action* to collect information about other users through extractive means that results in unwanted contact from the service. As a result, the user assumes that information about their friends or social network is vital for use of the service, even while this information is used to spam other users.

- Friend Spam constitutes distortion/impairment as it constrains user choices by contacting their friends or social network without their permission (factor *CA*).

- ***Additional autonomy violation:*** There is also a possibility of deception and/or manipulation. There is deception if it omits information (factor *IA*) or provides false or misleading information (factor *IC*) to lead a user into thinking that information about other users is vital for the use of the service. There is manipulation if it steers users into sharing this information using linguistic framing (factor *IF*) or aesthetic interface elements (factor *IP*).

Address Book Leeching[L] uses *Forced Communication or Disclosure* as a type of *Forced Action* to collect information about other users through extractive means, which are often hidden to the user and/or conducted under false pretenses. As a result, the user assumes that only vital information will be collected when signing up for or using a service, even while this information is used to gain knowledge of other users or inform other purposes that have not been initially declared.

- Address Book Leeching constitutes distortion/impairment and/or deception. There is distortion/impairment if it constrains user choices by forcing or mandating them to share information about other users, or if this information is later used for undeclared purposes without user permission (factor *CA*) There is deception if it omits information (factor *IA*) or provides false or misleading information (factor *IC*) to lead a user into thinking that information about other users is vital for the use of the service.

- ***Additional autonomy violation:*** There is also a possibility of manipulation if it steers users into sharing more information using linguistic framing (factor *IF*) or aesthetic interface elements (factor *IP*).

Social Pyramid[L] uses *Forced Communication or Disclosure* as a type of *Forced Action* to manipulate existing users into recruiting new users to use a service, often by tying this recruitment to additional functionality or other benefits. As a result, the user assumes that social recruiting is necessary to continue to use aspects of the service, even while this information is primarily used to build the service's user base.

- Social Pyramid constitutes distortion/impairment as it constrains user choices by forcing them to recruit new users to receive additional functionality or other benefits (factor *CA*).

- ***Additional autonomy violation:*** There is also a possibility of deception and/or manipulation. There is deception if it omits information (factor *IA*) or provides false or misleading information (factor *IC*) to lead a user into thinking that social recruiting is necessary for the use of the service. There is manipulation if it steers users towards social recruiting, such as by framing (factor *IF*) or presenting (factor *IP*) the associated additional benefits in a tempting manner.

Gamification[M] subverts the user's expectation that system functionality is based on alignment with user goals and needs, instead coercing them into gaining access to aspects of a service through repeated (and perhaps undesired) use of aspects of the service.

- Gamification constitutes distortion/impairment as it constrains user choices by forcing or coercing them into using certain undesired aspects of a service to gain access to the desired aspects.

Pay-to-Play[L] uses *Gamification* as a type of *Forced Action* to initially claim that aspects of a service or product are available via purchase or download, but then later charging users to actually obtain that functionality. As a result, the user incorrectly assumes that a service or product will allow them certain functionality, leading to them downloading or purchasing the product or service under false pretenses.

- Pay-to-Play constitutes deception as it omits information (factor *IA*) or provides false or misleading information (factor *IC*) to create a false perception that a certain functionality of a product or a service is available via purchase or download, whilst later charging users additionally for that functionality.

Grinding[L] uses *Gamification* as a type of *Forced Action* to require repeated, often cumbersome and labor-intensive actions over time in order to obtain certain relevant functionality. As a result, the user may seek to avoid these repetitive actions, leading to them making unwanted additional in-app purchases to unlock the same functionality without "grinding" over an extended period of time.

- Grinding constitutes distortion/impairment as it constrains user choices by forcing or coercing them to perform repeated, cumbersome and labor-intensive actions (factor *CE*), or otherwise by forcing them to make in-app purchases (factor *CA*), so that they can gain access to relevant functionality.

Attention Capture[M] subverts the user's expectation that they have rational control over the time they spend using a system, instead tricking them into spending more time or other resources to continue use for longer than they otherwise would.

- Attention Capture constitutes manipulation as steers users to spend more time on a system by potentially bypassing rational and deliberative thinking.

- ***Additional autonomy violation:*** There is also a possibility of deception if it provides false or misleading information to capture users' attention.

Auto-Play[L] uses *Attention Capture* as a type of *Forced Action* to automatically play new video after an existing video has completed. As a result, the user may lose control over their viewing experience, leading them to watch more content than they intended or result in them watching content that is unexpected or harmful.

- Auto-Play constitutes manipulation as it steers users towards watching more videos by reducing friction in the video playing experience, subverting rational deliberation over time spent (factor *CS*).

- ***Additional autonomy violation:*** There is also a possibility of distortion/impairment if it constrains user choices by making it impossible (factor *CA*) or unreasonably effortful to turn off the auto-play (factor *CE*) .

SOCIAL ENGINEERING

Scarcity and Popularity Claims[M] subverts the user's expectation that information provided about a product's availability or desirability is accurate, instead pressuring the user to purchase a product without additional reflection or verification.

- Scarcity and Popularity Claims constitutes deception and/or manipulation. There is deception if it provides false or misleading information about a product's availability or desirability. There is manipulation if it steers users into a purchase by tempting them through a perception of scarcity or popularity, even if such information is not outrighly false.

High Demand[L] uses *Scarcity and Popularity Claims* as a type of *Social Engineering* to indicate that a product is in high-demand or likely to sell out soon, even though that claim is misleading or false. As a result, the user may assume that demand is high when it is not, leading to their uninformed purchase of a product or service.

- High Demand constitutes deception and/or manipulation. There is deception if it provides false or misleading information about a product's demand (factor *IC*). There is manipulation if it steers users towards a purchase by framing it as popular (factor *IF*), even if such claims are not outrightly false; and through the timing, placement and aesthetics of these claims (factor *IP*).

Social Proof[M] subverts the user's expectation that the indicated behavior of others in a specific situation is correct or desirable, instead accelerating user decision-making and encouraging the user to trust flawed implications through provided information.

- Social Proof constitutes manipulation as it steers users towards certain choices or decisions using their feelings of trust in other people's behaviors.
- ***Additional autonomy violation:*** There is also a possibility of deception if it provides false or misleading information about other people's behaviors.

Low Stock[L] uses *Social Proof* as a type of *Social Engineering* to indicate that a product is limited in quantity, even though that claim is misleading or false. As a result, the user may assume that a product is desirable due to demand, leading to undue or uninformed pressure to buy the product immediately.

- Low Stock constitutes deception and/or manipulation. There is deception if it provides false or misleading information about a product's limited quantity (factor *IC*). There is manipulation if it steers users towards a purchase by framing it as scarce (factor *IF*), even if such claims are not outrightly false; and through the timing, placement and aesthetics of these claims (factor *IP*).

Endorsements and Testimonials[L] use *Social Proof* as a type of *Social Engineering* to indicate that a product or service has been endorsed by another consumer, even though the source of that endorsement or testimonial is biased, misleading, incomplete, or false. As a result, the user may assume that the endorsement or testimonial is accurate and unbiased, leading to their uninformed purchase of a product or service.

- Endorsements and Testimonials constitutes manipulation and/or deception. There is manipulation if it steers users towards a purchase using their feelings of trust in endorsements by other consumers (factor *IF*) or through the timing, placement or aesthetics of these endorsements (factor *IP*). There is deception if it provides false or misleading information in such endorsements and testimonials (factor *IC*).

Parasocial Pressure[L] uses *Social Proof* as a type of *Social Engineering* to indicate that a product or service has been endorsed by a celebrity, influencer, or other entity that the user trusts, even though the source of that endorsement is biased, misleading, incomplete,

or false. As a result, the user may assume that the endorsement is accurate and unbiased, leading to their uninformed purchase of a product or service.

- Parasocial Pressure constitutes manipulation and/or deception. There is manipulation if it steers users towards a purchase using their feelings of trust in a celebrity, influencer or any other entity (factor *IF*); or through the timing, placement or aesthetics of these endorsements (factor *IP*). There is deception if it provides false or misleading information in these endorsements (factor *IC*).

Urgency[M] subverts the user's expectation that information provided about discounts or a limited-time deal for a product is accurate, instead accelerating the user's decision-making process by demanding immediate or timely action.

- Urgency constitutes deception and/or manipulation. There is deception if it provides false or misleading information about discounts or limited-time deals. There is manipulation if it steers users towards a purchase due to a sense of urgency and pressure, even if such information is not outrightly false.

Activity Messages[L] use *Urgency* as a type of *Social Engineering* to describe other user activity on the site or service, even though the data presented about other users' purchases, views, visits, or contributions are misleading or false. As a result, the user may falsely feel a sense of urgency, assuming that others users are purchasing or otherwise interested product or service, leading to their uninformed purchase of a product or service.

- Activity Messages constitutes deception and/or manipulation. There is deception if it provides false or misleading information about other users' activity (factor *IC*). There is manipulation if it steers users towards a purchase by framing it as urgent (factor *IF*), even if these claims are not outrightly false; or through the timing, placement and aesthetics of these claims (factor *IP*).

Countdown Timers[L] use *Urgency* as a type of *Social Engineering* to indicate that a deal or discount will expire by displaying a countdown clock or timer, even though the clock or timer is completely fake, disappears, or resets automatically. As a result, the user may feel undue urgency and purchasing pressure, leading to their uninformed purchase of a product or service.

- Countdown Timers constitutes deception and/or manipulation. There is deception if it provides false or misleading information in the form of fake timers, or if they disappear or reset automatically (factor *IC*). There is manipulation if it steers users towards a purchase by framing it as urgent (factor *IF*), even if the timers are not fake; or through the timing, placement and aesthetics of these timers (factor *IP*).

Limited Time Messages[L] use *Urgency* as a type of *Social Engineering* to indicate that a deal or discount will expire soon or be available only for a limited time, but without specifying a specific deadline. As a result, the user may feel undue urgency and purchasing pressure, leading to their uninformed purchase of a product or service.

- Limited Time Messages constitutes deception and/or manipulation. There is deception if it omits information about the deadline of the offer (factor *IA*) or if the deadline is false or misleading (factor *IC*). There is manipulation if it steers users towards a purchase by framing it as urgent (factor *IF*), even if the deadlines are not outrightly false; or through the timing, placement and aesthetics of these deadlines (factor *IP*).

Shaming[M] subverts the user's expectation that their decision is being guided by objective information and made in an informed way, instead using emotionally manipulative tactics to pressure the user to make certain choices over others in order for them to feel they have done the right thing.[10]

- Shaming constitutes manipulation as it steers users into making certain choices over others by using emotions to pressure them.

[10]The definition of *Shaming* is taken from the Gray et al. [47] Supplementary Materials.

Confirmshaming[L] uses ~~*Personalization*~~[11] *Shaming* as a type of *Social Engineering* to frame a choice to opt-in or opt-out of a decision through emotional language or imagery that relies upon shame or guilt. As a result, the user may be convinced to change their goal due to the emotionally manipulative tactics, resulting in being steered away from making a choice that matched their initial goal.

- Confirmshaming constitutes manipulation as it steers users towards or away from particular choices by framing information in a way that creates shame or guilt (factor *IF*).

Personalization[M] subverts the user's expectation that products or service features are offered to all users in similar ways, instead using personal data to shape elements of the user experience that manipulate the user's goals while hiding other alternatives.

- Personalization constitutes manipulation and/or deception. There is manipulation if it steers users towards particular choices over alternatives by using personal data to shape user experience. There is deception if it omits information by hiding certain choices or alternatives from users.

7.2 Supplementary figures

Figure 3 and Figure 2 provide visual context to Table 1, particularly for understanding overarching trends in how dark patterns are distributed across the autonomy violation types. Here we provide further granularity with factor-wise diagrams.

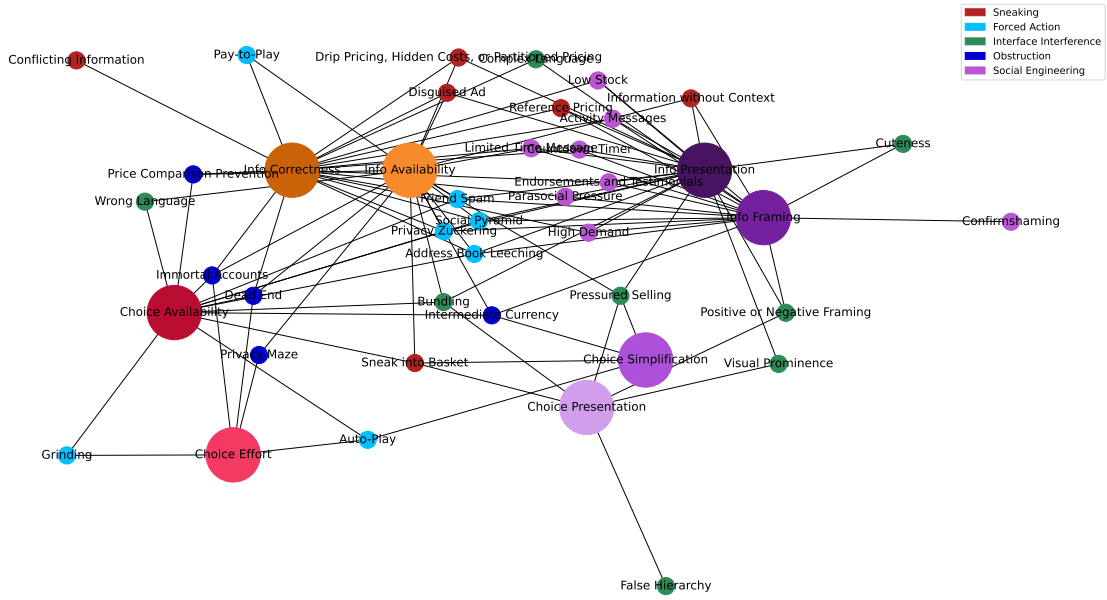

Fig. 5. Network graph of relationships between autonomy violations and dark patterns, with nodes colored according to their Gray et al. [47] high-level pattern type.

[11] *Confirmshaming* at the low-level maps to *Shaming* at the meso-level in the Gray et al. [47, Figure 5] ontology.

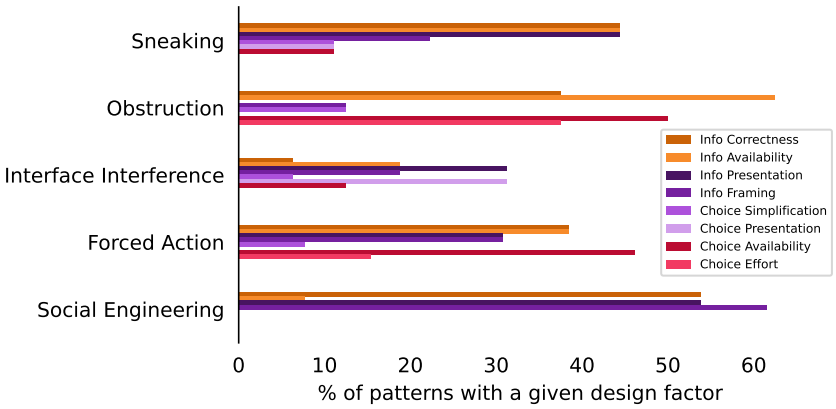

Fig. 6. Bar chart depicting percentages of patterns with a given design factor, as grouped by the high-level pattern type. Autonomy violations are inferred by the color of each bar, with factors presented as a gradient of their primary color: orange for deception, purple for manipulation, and red for deception.