What Makes XR Dark? Examining Emerging Dark Patterns in Augmented and Virtual Reality through Expert Co-Design

VERONIKA KRAUß, School of Information, University of Michigan, Ann Arbor, MI, USA and Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany
PEJMAN SAEGHE, University of Strathclyde, Glasgow, United Kingdom
ALEXANDER BODEN, Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany and Fraunhofer FIT, Sankt Augustin, Germany
MOHAMED KHAMIS and MARK MCGILL, University of Glasgow, Glasgow, United Kingdom
JAN GUGENHEIMER, Technical University of Darmstadt, Darmstadt, Germany and Telecom-Paris (IP-Paris), Paris, France
MICHAEL NEBELING, School of Information, University of Michigan, Ann Arbor, MI, USA

Dark patterns are deceptive designs that influence a user's interactions with an interface to benefit someone other than the user. Prior work has identified dark patterns in windows, icons, menus, and pointer (WIMP) interfaces and ubicomp environments, but how dark patterns can manifest in Augmented and Virtual Reality (collectively XR) requires more attention. We therefore conducted 10 co-design workshops with 20 experts in XR and deceptive design. Our participants co-designed 42 scenarios containing dark patterns, based on application archetypes presented in recent HCI/XR literature. In the co-designed scenarios, we identified 10 novel dark patterns in addition to 39 existing ones, as well as 10 examples in which specific characteristics associated with XR potentially amplified the effect dark patterns could have on users. Based on our findings and prior work, we present a classification of XR-specific properties that facilitate dark patterns: perception, spatiality, physical/virtual barriers, and XR device sensing. We also present the experts' assessments of the likelihood and severity of the co-designed scenarios and highlight key aspects they considered for this evaluation, for example, technological feasibility, ease of upscaling and distributing malicious implementations, and the application's context of use. Finally, we discuss means to mitigate XR dark patterns and support regulatory bodies to reduce potential harms.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; **Mixed/augmented reality**;

Additional Key Words and Phrases: Speculative design workshops, deceptive design, augmented reality, virtual reality, dark patterns

Authors' Contact Information: Veronika Krauß (Corresponding author), Technical University of Darmstadt, Darmstadt, Germany; e-mail: veronika.krauss@tu-darmstadt.de; Pejman Saeghe, University of Strathclyde, Glasgow, United Kingdom; e-mail: pejman.saeghe@strath.ac.uk; Alexander Boden, Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany and Fraunhofer FIT, Sankt Augustin, Germany; e-mail: alexander.boden@h-brs.de; Mohamed Khamis, University of Glasgow, Glasgow, United Kingdom; e-mail: Mohamed.Khamis@glasgow.ac.uk; Mark McGill, University of Glasgow, Glasgow, United Kingdom; e-mail: Mark.McGill@glasgow.ac.uk; Jan Gugenheimer, Technical University of Darmstadt, Darmstadt, Germany; e-mail: jan.gugenheimer@tu-darmstadt.de; Michael Nebeling, School of Information, University of Michigan, Ann Arbor, MI, USA; e-mail: nebeling@umich.edu.

ACM 1557-7325/2024/8-ART32
https://doi.org/10.1145/3660340

ACM Reference format:
Veronika Krauß, Pejman Saeghe, Alexander Boden, Mohamed Khamis, Mark McGill, Jan Gugenheimer, and Michael Nebeling. 2024. What Makes XR Dark? Examining Emerging Dark Patterns in Augmented and Virtual Reality through Expert Co-Design. *ACM Trans. Comput.-Hum. Interact.* 31, 3, Article 32 (August 2024), 39 pages.
https://doi.org/10.1145/3660340

1 Introduction

Our research explores potential **Deceptive Designs (DD)** in **Augmented Reality (AR)** and **Virtual Reality (VR)** (collectively referred to as **Extended Reality (XR)**) user experiences that have been frequently framed as *dark patterns*[1] in recent work [32]. Dark patterns describe "design practices that materially distort or impair [...] the ability of recipients of [a] service to make autonomous and informed choices or decisions" [27] without regard to the designer's motive [27].

Our efforts are motivated by a rise of dark patterns' prevalence and impact in research and practice [7, 31, 33, 49, 55] (also called "asshole design practices" [30]) paired with a lack of works focusing on dark patterns in XR as a technology on the brink of widespread adoption.

The dark pattern literature so far has predominantly studied e-commerce platforms, online services, and user consent [8, 10, 31, 43, 54, 62], typically enacted on **two dimensional (2D)** physical displays. The emergence and consumer uptake of immersive and spatial computing technologies however offers new platforms upon which dark patterns can be enacted. VR in particular has become known for its potential to let users engage in otherwise impossible situations, such as meeting dead relatives [36], switching your body [52], or experiencing situations that might trigger a user's fear of heights [6]. Furthermore, AR grows increasingly capable of mediating our perception of reality, meaning that DD that apply in virtual environments could soon find their way into our everyday physical environment. In this regard, our work is building on an increasing awareness in the **Human Computer Interaction (HCI)**/XR research communities regarding potential harms caused by certain XR designs (e.g., manipulative advertising [57], perception hacking [71, 73], memory manipulation [9], and privacy and security issues in XR [1, 13, 67]).

While manipulations enacted in XR could be used for good purposes, for example, treating anxiety disorders [6] or being able to experience small physical spaces as a spacious virtual area through redirected walking [69], considering the past dark pattern adoption on web and mobile, we anticipate an increased likelihood of key vendors and XR application developers exploiting the affordances of immersive and spatial computing to enact harmful and DD for personal gain. Further, novice designers might apply dark patterns without foreseeing their potential impact and consequently harm users and bystanders due to a lack of experience and training with XR.

As we still lack insights into how the application of such DD, especially dark patterns, impact users and society, we need to address this gap by investigating potential exploits of XR's unique affordances and capabilities.

Therefore, our work is guided by three research questions:

RQ1: What existing or new dark patterns could emerge in XR application scenarios?
RQ2: Which XR-specific properties would enable or potentially amplify dark patterns in XR applications?
RQ3: What aspects determine the probability and perceived severity of dark patterns in XR design?

[1] While *dark patterns* is a widely accepted term for coercive, deceptive, and manipulative design techniques in **User Interfaces (UI)**, we acknowledge potential issues raised by the ACM Diversity and Inclusion Council. Given the lack of widespread alternatives, we continue using *dark patterns* to connect with existing publications.

This article presents co-design workshops with experts in HCI and DD to explore possible scenarios and use cases of XR technologies and applications that were proposed in research as examples of up-and-coming XR interfaces. Specifically, we paired 10 XR experts (i.e., individuals who had recently published an article in a top HCI venue that proposed a novel XR technology or application) with 10 DD experts (i.e., individuals working in the field of DD in academia or industry). In 10 speculative design workshops, our participants constructed a total of 42 scenarios designed to explore dark patterns that could emerge from XR technologies and applications described in the literature. Analyzing the co-designed scenarios, we discovered four properties of XR that our participants commonly operationalized in an attempt to apply or create dark patterns: *perception*, *spatiality*, *physical/virtual barriers*, and *XR device sensing*.

To understand if the created DD could in fact be classified as dark patterns, we clustered them, expanding on Gray et al.'s dark pattern analysis strategies [31] and their preliminary dark patterns ontology [32]. We identified eight clusters, each leveraging at least one XR property, that formed around known dark patterns: persuading, imbalancing options, baiting, lying, directing attention, disguising (content, actions, or data collection), requiring a detour, and exploiting the drive to succeed. Four of the 42 co-designed scenarios created by our participants did not contain dark patterns; the remaining scenarios involved 39 known dark patterns per Mathur et al.'s catalog [55]. Further, we characterized 10 of the created DD as emergent dark patterns in XR because they did not fall under known definitions. Out of the identified dark patterns, we classified another 10 as existing dark patterns that were amplified through XR, for example, by presenting super-realistic virtual objects, personalizing a manipulation based on the XR device's sensor data, or making use of the user's body in the XR experience to inflict discomfort or physical harm. When our participants anticipated the risk of the co-designed scenarios and discussed their probability and severity given a 5- to 10-year time frame, there was a strong trend that dark patterns are likely to become, if not already, an imminent threat in emerging XR technology and applications.

This work presents the following four main contributions grounded in the analysis of our co-design workshop:

(1) Identifying four core properties of XR technology (perception, spatiality, physical/virtual barriers, and XR device sensors) that enable and amplify the impact of new and existing dark patterns in XR.
(2) An eight-tier classification of existing dark patterns based on previous work [31, 32, 55] and associated with the aforementioned XR properties.
(3) An experts' assessments of likelihood and severity of the co-designed scenarios as well as considered key aspects used to reason about their ratings.
(4) Forty-two co-designed scenarios detailing potential manipulative designs applicable in XR.

2 Related Work

Our work is grounded in recent literature on deceptive (or dark) design and XR research on manipulative design and privacy, safety, and security. With our work, we start to close the gap between the already described manipulative characteristics of XR and their potential exploitation through applying them as dark patterns in current and upcoming XR applications. Below, we provide an overview of deceptive and manipulative design in XR in addition to dark patterns as described in research and practice.

2.1 Deceptive and Manipulative Design in XR

XR has the potential to be a dominant platform for accessing digital spatial information in the coming decade. Consequently, mass adoption of XR [38], from consumer VR for gaming and collaboration, to everyday ubiquitous AR as envisioned by Apple, Meta, Microsoft, and so on, has the potential to offer a rich, always-available platform for real-time sensory deception [2, 51] and manipulation [73].

XR headsets act as mediators of our perception of reality [53]—with VR supplanting what we perceive in favor of virtuality, and AR supplementing or altering what we perceive. XR's capacity for ever-increasing perceptual realism [68], coupled with the ability to sense users' actions [56] and physiological signals [35], offers a sophisticated platform for understanding users' behavior with high degree of accuracy and provides the ability to manipulate users' behavior.

With respect to AR, as Wolf et al. noted, "should developers succeed in engineering the AR experience in such a way that augumented reality is indistinguishable (or nearly so) from physical reality, users might be deceived into believing in the physical existence of what they see, even though it is not physically present" [76]. In effect, seeing is (often-times) believing, and with XR we can manipulate what the user sees, in reality or virtuality, prodigiously, and we can do so on per-individual basis, i.e., two users side-by-side could perceive very different experiences unbeknownst to the other. This capacity for perceptual manipulation can notably be exploited to trick users into taking unintended actions. For example, a user's interactions with a spatial UI can be manipulated to lead them to select the wrong button [71], transposing a common dark pattern (*cursorjacking*) to a spatial XR context. Also, a user's movements could be more overtly manipulated. In defining virtual–physical perceptual manipulations, Tseng et al. [73] demonstrated how approaches such as redirected walking and redirected haptics could be maliciously co-opted, used to non-consensually direct user movements and actions toward unintended or unanticipated ends, including creating a basis for experiencing or inducing physical harm. While such overt safety-critical manipulations are possible (particularly in VR), the resultant harm is likely to be highly visible to the user, and consequently many of the approaches outlined by Tseng could be exploited once or twice only. In contrast, when considering DD, if the result of the deception is not immediately evident to the user being deceived, XR could facilitate deceptions that occur frequently or even continuously for significant periods of time. These deceptions could be enacted in **three dimensional (3D)** in virtuality or reality, and could be rendered in such a way as to be indistinguishable from reality, or seamlessly blended with intended elements of the virtual spatial UIs the user relies on.

While the ability of XR to enable perceptual manipulations is frequently researched within the field of interaction design to overcome technological limitations (e.g., locomotion [64, 70, 72] and haptics [4, 17, 65]), researchers additionally started to explore how these types of perceptual manipulations could be instrumented to induce physical harm [14, 57, 73] or economic benefits for the creator [57]. In 2014, Roesner et al. presented a set of potential challenges around security and privacy of AR technology [67] which was updated in 2021 [66]. They refer to the construct of perceptual manipulations as *output security*. Lebeck et al. explored how the *output challenge* can be structurally explored in AR to enable safe [46] and secure [47] augmentations. In an interdisciplinary approach, Baldassi et al. combined neuroscience and computer science in their exploration of the output challenge and identified one of the core challenges to be the ability of AR to manipulate the users' sensory, perceptual, and cognitive system [5].

In line with Baldassi et al.'s approach, our work is similarly arguing that the ability of XR technology to alter the perception of a user makes the potential for deception and manipulation more impactful. However, instead of bridging our work toward neuroscience, we are creating a link to design research and aim to understand how this ability of perceptual manipulations might impact existing and future DD patterns in XR.

2.2 Dark Patterns Described in Research and Practice

Since Conti and Sobiesk published their work studying malicious interface practices in 2010 [22] and Brignull et al. created the online pattern catalog *deceptive.design* [7][2] for "naming and shaming deceptive user interfaces" [11], dark patterns research has gained traction in research communities, media, and legislation [55].

In addition to an increasing number of legal regulations (e.g., [26, 74]), research expanded not only to investigate the user perspective on and potential impact of dark patterns [8, 48, 50] but also extended the scope of technologies and application domains to include privacy, security, and consent [10, 43, 62], gaming [78], or online retail [54]. Other work focuses on more or less mass market-ready technology beyond the screen, as they might also bear the potential of dark patterns [31], such as voice interfaces [63] and robots [45]. Further, Greenberg et al. [33] discuss eight dark patterns in proxemic interaction interfaces and speculate on potential solutions for identified root causes. Among others, they focus on the duality of meanings and the ambiguity of ownership in physical space as well as accidental proxemics and resulting privacy violations or data leakage [33], which also should be considered in XR scenarios as they utilize related interface properties.

Further, early work already considers XR as a technology that demands attention regarding emerging dark patterns [41, 57] and has focused on input and output security of XR [66, 67] (see Section 2.1). Nevertheless, there have been only few attempts to provide grounded understanding of purposeful manipulations and mechanisms that could be applied as dark patterns: In the context of ubiquitous AR, Eghtebas et al. [25] co-speculated on potential manipulative scenarios with diverse potential end-users. Based on their co-speculative design workbook study, they identify the three overarching themes (1) situatedness of information, (2) altering perception, and (3) ubiquitous sensing as source of unintended or intended harm. Moreover, they present seven scenarios describing a narrative of a potential manipulative or DD combined with a potential *antidote* or countermeasure. Finally, Wang et al. [77] implemented based on a scenario construction approach three potential dark patterns to test their effectiveness. In this line, we add to previous works through a speculative design workshop with XR and DD experts. Not only do we focus on potential scenarios of both AR and VR settings, but we are also interested in what kind of XR properties experts perceive as potential levers for future dark patterns and how they rate their probability and severity.

2.3 Adoption, Classification, and Mitigation of Dark Patterns

The question regarding the responsible actor for the usage of dark patterns and how to effectively countermeasure their activities is actively debated and evolves around user agency [8], designer's responsibility [31, 59, 60], and regulations enforced by law (e.g., [26, 74]). In this regard, Nelissen and Funk [60] investigated how designers approach creating interfaces with privacy in mind and solve potential conflicts of interest regarding ethical challenges and dark patterns. However, they found that designers are often unaware of how legal frameworks affect their design processes. They are even partly to blame for incorporating dark patterns in privacy-affecting design decisions [60]. Further, Nayaranan et al. [59] "urge the design community to set standards for itself, both to avoid onerous regulation and because it's the right thing to do" [59]. Also, Gray et al. address a designer's responsibility regarding ethical design practices and dark patterns and emphasize the need for "comprehensive ethics education in HCI and UX" [31] combined with a better integration of value-centered and pragmatic tools for practitioners and researchers [31].

[2]The website was originally called *darkpatterns.org* and was renamed in 2022 to be more inclusive [7].

However, the aforementioned broad variety of dark patterns research (e.g., [10, 18, 30, 31, 48]) also urges the need to align concepts, terminology, and taxonomies. There have been approaches focusing on more robust frameworks to describe dark patterns, for example, through the lens of choice architecture [20, 21, 55]. However, such an alignment has not yet been accomplished [55] which renders the regulation of dark patterns as well as implementation-agnostic debates and dark pattern education to be a difficult task. Gray et al. recently published a preliminary three-level ontology of dark patterns and respective knowledge that takes a first step toward harmonizing "regulatory and academic taxonomies of dark patterns" [32]. Our aim is to inform iterations of ontologies and classification frameworks such as Gray et al. [32], by providing an entry point for future work addressing dark patterns in the field of XR. In this sense, we aim to understand how dark patterns might be adapted to or amplified through XR-specific properties, or what potential emergent dark patterns could manifest with ongoing mass adoption. Therefore, we provide a new lens for establishing such robust definitions. Further, we offer early approaches to assess the risk such patterns could impose on users.

3 Study Design and Analysis

To address our research questions aiming to understand what types of DD patterns might occur in XR, how specific aspects of XR technology and applications might amplify them, and how probable and severe these aspects might be when applied in dark patterns, we ran a series of speculative design workshops between June and August 2022. In this section, we present our workshop design, including how it guided our recruitment strategy and our approach to reducing conflicts and bias as well as broadening topic coverage in our study.

3.1 Participant Recruitment and Pairing

Our study paired researchers in XR who recently proposed an XR technology or application (*XR experts*) with experts in ethical technology and DD from academia and industry (*DD experts*). We focused on XR researchers rather than developers from industry because the latter usually prioritizes using platforms and technologies that are accessible to a large share of the market. Researchers, however, aim to build forward looking interfaces and experiences and are less impacted by the availability and affordability of technologies as they do not aim to make profit from their prototypes.

Each XR expert was recruited based on one publication they contributed between 2018 and June 2022 to the field of XR. This publication was later used as the basis for co-designing potential DD patterns with the DD expert. We used these examples from the literature as a common ground and starting point for our workshop participants, ensuring there is relevant expertise in both XR and DD with each pairing. In total, 20 participants (7 female, 12 male, 1 non-conforming; age range 26–44 years; mean: 30 years) were recruited (see Table 2) according to our sampling strategy detailed below.

XR Expert Recruitment. Each XR expert was recruited based on one publication they contributed in the last 5 years to the field of XR as we wanted to ground our work in recent advances in XR application concepts. This publication was later used as the basis to create potential DD patterns. We ran a query on the SCOPUS database, targeting publications between January 2018 and June 2022 at ACM CHI, ACM UIST, ACM CSCW, IEEE VR, and IEEE ISMAR;[3] searching the abstract, title, and keywords using the following search query: `["augmented reality" OR "virtual reality" OR "mixed reality" OR "extended reality"] AND ["application" OR "system" OR "technique" OR "tool"]`. This resulted in

[3]We selected those conferences based on their impact score ranking on research.com, a known research portal ranking conferences and journals of various fields, and their frequency of having published articles addressing XR discourses.

Table 1. Topic Coverage of the Identified Seed Articles Adapted from Dey et al.'s Nine Application Areas [24]

WS	Type	Application area	Technique	Custom hardware	Venue
WS01	VR	Commercial and shopping	Haptics (force, physical properties)	Safe-to-touch quad-copter	CHI
WS02	AR	Navigation and driving	Visualization, x-ray vision	—	CHI
WS03	VR	Entertainment and gaming, social	Asymmetric co-located interaction	HMD with outward facing touch displays	CHI
WS04	AR	Interaction, medicine	3D Eye tracking	—	CHI
WS05	VR	Entertainment and gaming, perception	Haptics (facial force feed-back)	HMD-integrated pulley system	UIST
WS06	VR	Perception	Passive haptics	3D printed hair-like structures	CHI
WS07	VR	Education, interaction	Multi-device interaction	—	CHI
WS08	AR	Entertainment and gaming, social	Co-located asymmetric multi-user interaction	HMD mounted with a projector	UIST
WS09	AR	Interaction	Navigating application menus with foot taps	—	CHI
WS10	AR	Education, collaboration	Telepresence, annotation, shared spaces	—	CHI

HMD = head-mounted display; WS = workshop-ID.

the retrieval of 1,727 publications we subsequently sorted according to Google Scholar citation count. Starting from the most cited article, we applied the following criteria to identify suitable work:

— The article needs to have an implementation of an XR system to function as a basis for the creating dark patterns.
— There needs to be an accessible demo/video of the application to be later on used as descriptive material in the workshop to onboard the DD expert.
— The presented application had to describe a use case that incorporated features and interactions that require the user to make decisions and hence be subject to deception.
— We excluded design tools and conceptual work because, unlike articles that propose applications, it was not feasible to develop a base use case from these articles in a workshop setting.

After identifying each suitable article, referred to as a *seed* article, we reached out to one of its authors (if possible the first author), asking them to participate in our study. We contacted 29 potential XR experts of which 11 replied and 10 participated in our study.

Throughout recruitment, we aimed to maximize topic coverage with regards to XR application areas, and for an equal distribution between AR and VR seed articles (see Table 1). Seed articles were subsequently used in workshops to create a base scenario (see Section 3.2).

DD Expert Recruitment. Since DD patterns are getting more and more attention in the academic field and originated from industry practices [7], we included academic researchers as well as practitioners of design as participants. We further expected to complement the researchers' perspectives on dark patterns and manipulative designs by involving industry practitioners.

Regarding *academic DD experts*, we used the same method as described previously regarding XR expert recruitment; within the same time frame, and in the same venues, we search the abstract, title, and keywords using the following search query: `["dark pattern"] OR [deceptive AND` `design] OR [ethical AND design]`. This resulted in the retrieval of 26 publications. After sorting the articles according to Google Scholar citations, we reached out to 23 authors (typically the first author) from the top of the sorted list; 9 academic DD experts responded and 8 participated in our study.

Regarding *industry DD experts*, our criteria was two-fold: (1) the DD expert needs to be currently active in industry, for instance, as a UI designer or as a user experience designer (e.g., according to their public Linkedin profile), and (2) they need to have regularly published articles, books, or talks on DD, dark patterns, or ethical design in the past 5 years. For this latter condition, we searched for publications mentioned in relevant forums (e.g., r\assholedesign, r\darkpatterns), websites detailing and discussing dark patterns, Slack channels, Discord channels, Medium posts, Facebook groups, books, and LinkedIn. We identified and reached out to eight industry DD experts; two participated in our study.

Paring XR Experts with DD Experts. For each workshop a DD expert was paired with an XR expert based on the availability they indicated. For this, we used an online calendar tool for time management and coordinating meetings.

3.2 Workshop Design

We conducted 10 speculative co-design workshops, where we adapted the design thinking method of *brainwriting* [23]. Workshops were conducted remotely over Zoom video conferencing; Miro, an online collaborative whiteboard platform, was also used to facilitate collaboration between participants. Before each session, we sent the respective seed article to both participants. We asked both to make themselves familiar with the concept or, in the case of the XR expert who was one of the article's authors, to recap the article's content. Each session was scheduled (and typically lasted) for 90 min and was attended by a DD expert, an XR expert, and two researchers; where one researcher conducted the session (moderator) and another researcher functioned as a backup (backup moderator), in case the moderator faced difficulties, such as Internet connectivity issues.

Each session started with the two researchers introducing themselves, from this point onward the backup moderator remained quiet, and the main conductor lead the session. The participants were introduced to study goals and an outline of what they would expect during the workshop. Subsequently, the DD expert and the XR expert introduced themselves. After this point the sessions were recorded. The remainder of the workshop consisted of the four phases plus an interjection (see Figure 1):

Phase 1—Establishing the Primary Use Case (10 min). After both participants established a common understanding about functionalities of the seed article's application, they collaboratively created a primary use case. The primary use case provided details about the application's user(s), their motivation and goals, their actions and activities, the application's role in helping user(s) reach their goals, and the context of use.

Interjection—Familiarization with Dark Patterns (<6 min). Participants were provided with Brignull's definition of *dark patterns*: "Deceptive design patterns (also known as 'dark patterns') are tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something" [7], with an emphasis on a beneficiary as a crucial aspect of a dark pattern. Participants were subsequently provided with four examples from Brignull's dark pattern catalog [7] adapted for an XR setting. We chose to present the pattern types *Misdirection*, *Confirmshaming*, *Friend Spam*, and *Disguised Ads* as there were publications mentioning or hinting toward those four types adapted to an XR setting [12, 57].

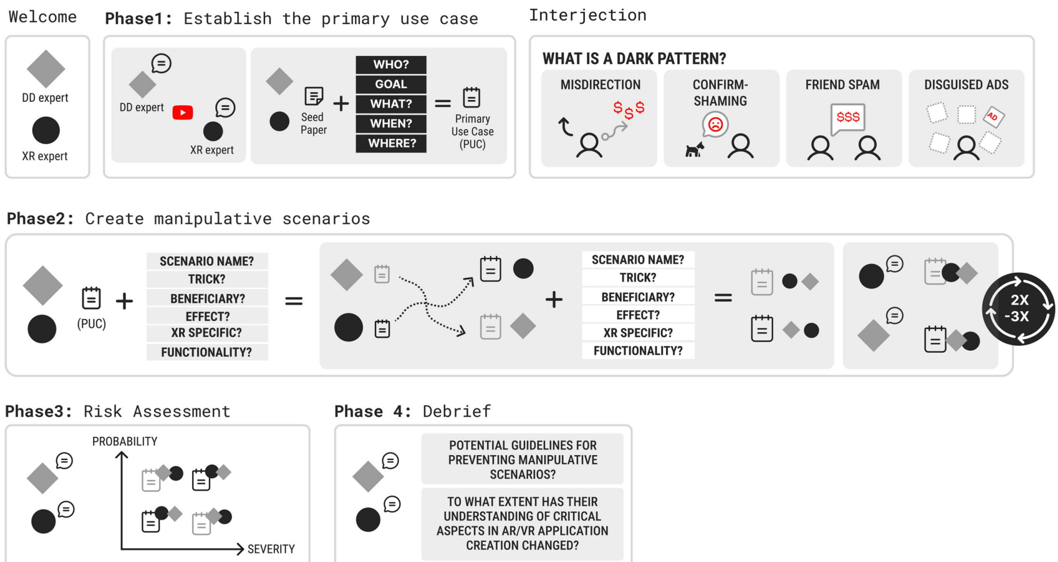

Fig. 1. Schematic workshop design. A figure describing the workshop design as a simplified schema. Three blocks are shown, grouping the workshop phases ‘welcome’, ‘phase1’ and ‘interjection’ in the first line, ‘phase 2’ in the second line, and ‘phase 3’ and ‘phase 4’ in the third line.

Phase 2—Co-Designing Manipulative Scenarios (30 min). This phase was split into three sub tasks, where participants focused on co-designing scenarios containing dark patterns, based on the primary use case (see Phase 1). First, each participant individually created a scenario (3 min), resulting in two scenarios. Then, participants swapped scenarios and worked on each other’s drafts by adding text (3 min). Finally, participants presented the combined scenarios and discussed, iterated, and amended them if needed (3 min per each of the two scenarios). This phase was repeated twice, resulting in four co-designed scenarios per workshop. In one instance this phase was repeated three times as participants saved time during the introduction and Phase 1, resulting in six co-designed scenarios.

Phase 3—Risk Assessment (10 min). Participants collaboratively risk assessed all co-designed scenarios they created in Phase 2, with regards to the severity and the probability of each scenario in a time frame considering the next 5–10 years. The probability-severity matrix consisted of five discrete values per axis. The values for the probability axis were *not probable*, *somewhat improbable*, *neutral*, *somewhat probable*, and *very probable*. The values for the severity axis were *insignificant*, *minor*, *major*, *severe major*, and *catastrophic*.

Phase 4—Debrief Discussion (< 15 min). Finally, the participants discussed potential guidelines for preventing or mitigating dark patterns in XR applications and exchanged lessons learned from the workshop. Participants received a $50 Amazon gift card for participating in a workshop.

3.3 Resulting Scenarios

Table 6 in Appendix provides a detailed list of the co-created scenarios and relates them to the respective workshops. In general, the co-designed scenarios kept the majority of properties inherited from their primary use case and seed article, such as if the application was AR or VR, used specific hardware, or featured specific interaction modalities. The full list of seed articles is also detailed in Table 1 and Section 3.1.

Table 2. Overview of Our Participants Based on Self-Reported Data; The Workshop Roles (Role) Were Assigned Based On a Participant's Background

WS	Role	Relation to XR	Exp XR	Relation to DD	Exp DD
WS01	DD	Consumer	None	Researcher, ethical design specialist	7–10
	XR	Consumer, researcher, engineer	4–6	Either none or not a conscious/ deliberate role	None
WS02	DD	Consumer, enthusiast, educator	1–3	Designer, researcher, ethical design specialist, coach/trainer	4–6
	XR	Researcher, engineer	1–3	None	None
WS03	DD	Enthusiast	None	Ethical design specialist, set up [a well-known dark patterns website]; published four papers	>10
	XR	Consumer, researcher	4–6	None	None
WS04	DD	Researcher	1–3	Researcher	7–10
	XR	Researcher	4–6	None	None
WS05	DD	Consumer	None	Researcher	1–3
	XR	Consumer, researcher, enthusiast	4–6	Researcher	<1
WS06	DD	None	None	Ethical design specialist	>10
	XR	Researcher	4–6	None	None
WS07	DD	Researcher	None	Researcher	1–3
	XR	Consumer, producer, researcher, engineer	4–6	None	None
WS08	DD	Researcher	>10	Researcher	1–3
	XR	Consumer, designer, researcher	4–6	Researcher	1–3
WS09	DD	Researcher, enthusiast	1–3	Researcher	1–3
	XR	Consumer, producer, researcher, engineer, enthusiast	7–10	None	None
WS10	DD	Writer, researcher, technologist	>10	Researcher, ethical design specialist	1–3
	XR	Consumer, researcher, engineer	4–6	None	None

WS = workshop-ID.
Exp XR/DD denotes the years a participant was working in the respective field of XR or DD.

Strategies. In total, our participants created 42 scenarios in 10 workshops and applied the following strategies to remodel the primary use case into scenarios describing manipulative application behavior: Seven repurposed features of the seed article application, six applied personalization of content and behavior, five abused or misused data, another five incorporated gamification elements, four described the creation or usage of an information asymmetry, four relied on perceptual manipulation, and two evoked and used emotion. Nine scenarios were created with tactics that did not fit other clusters, for example, impersonation or content placement. While some scenarios combined multiple tactics, others relied on a single manipulative strategy. In addition, four scenarios described aspects of hacking without specific manipulative strategies.

Environments. Regarding shared, virtual, and physical spaces, eight scenarios explicitly mentioned a shared virtual space, three a shared hybrid space (visual and physical), and two a shared physical space. The remaining 19 scenarios did not explicitly mention details in this regard.

Interaction Modalities. Out of the 42 scenarios, 12 were predominantly using visual modalities, 8 relied on manipulations involving tactile or force feedback and input, 5 required whole-body

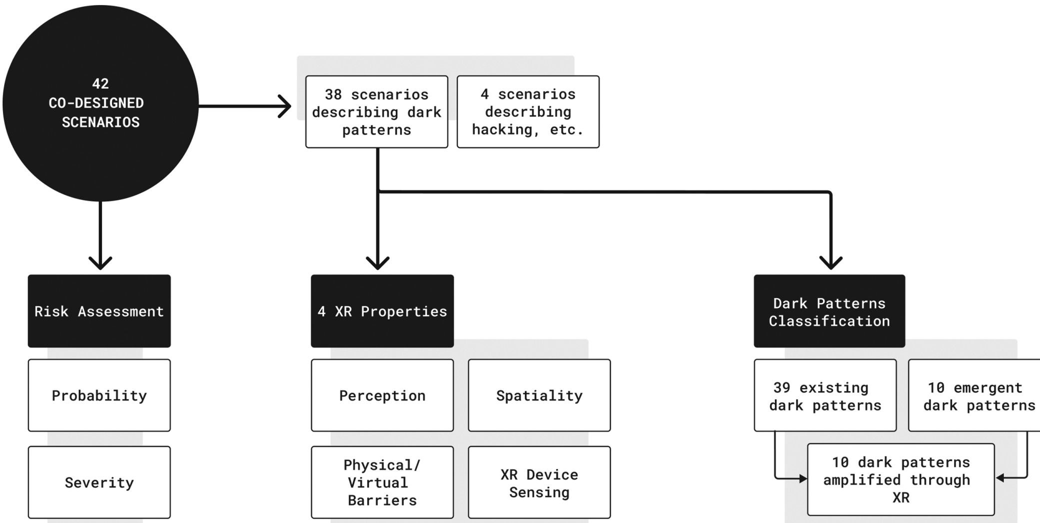

Fig. 2. Analysis of the 42 co-designed scenarios in relation to the synthesized results. The arrows depict the flow of analysis, black boxes the type of contribution, and white boxes the detailed contributions. This figure details the analysis of the 42 scenarios by splitting the data into the subsets 'risk assessment', '4 XR properties', and 'dark patterns classification'.

movement, and 1 relied on aural and vocal interaction. Further, five scenarios described multi-modal interaction. The remaining 11 scenarios did not specify the required or provided modalities.

3.4 Data Analysis

In total, over 10 co-design workshops, 42 co-designed scenarios were created by participants, of which we classified 38 as describing dark patterns and 4 as describing hacking or application misuse (WS01_01, WS07_4, WS08_3, WS10_4). The latter were excluded from dark pattern-related analysis (see Figure 2).

Two researchers followed the reflexive thematic analysis [19] approach and jointly analyzed the co-designed scenarios over Zoom and Miro. The researchers iterated over the dataset six times. The findings were presented to the rest of the authors to seek feedback and ensure clarity and intelligibility of the concepts and the process after every iteration. The following paragraphs present an overview of our analysis structured according to our research questions:

RQ1: What existing or new dark patterns could emerge in XR application scenarios? For this analysis, we used a combination of deduction and induction and initially discarded four co-designed scenarios because they described hacking or application misuse instead of dark patterns, i.e., the scenarios lacked a beneficiary or manipulative character. Therefore, this analysis was performed on a dataset consisting of 38 co-designed scenarios. Regarding the deductive approach, and to identify which, if any, of the existing dark patterns were contained in the co-designed scenarios, a code book was created based on Mathur et al.'s work (see Table 3 in [55]), consisting of 84 existing dark patterns. For each dark pattern, we went back to the source and extracted its definition/description. Subsequently, two researchers synchronously read and coded each co-designed scenario with as many dark patterns from the code book as possible. To further identify emerging and amplified dark patterns, three researchers clustered existing dark patterns regarding their manipulative mechanism expanding on Gray et al.'s manipulation strategies [31] and assigned them to the identified XR properties. In both iterations, a synchronous coding process enabled the coders to resolve disagreements through discussion. Regarding the inductive approach, whenever the researchers identified a dark pattern description in a co-designed scenario that was missing from the code book, a new dark pattern candidate was generated. Section 5 details our findings for this RQ.

RQ2: Which XR-specific properties would enable or potentially amplify dark patterns in XR applications? To answer this question, an inductive approach was used. The process consisted of two researchers reading each co-designed scenario and for each identified dark pattern noting which, if any, XR-specific properties were being leveraged to enable said dark pattern. This resulted in a collection of XR-specific properties, which were subsequently grouped into four overarching categories. Section 4 provides the results for this RQ.

RQ3: What aspects determine the probability and perceived severity of dark patterns in XR design? This question was answered using an inductive approach and thematic clustering of participants' quotes based on the identified codes. This resulted in four aspects influencing probability rating and four aspects considered for severity ranking. Section 6 details the findings for this RQ.

4 XR-Specific Properties Used by Dark Patterns

XR-specific properties are affordances or integral features leveraged by XR applications. In our workshops, participants used concepts of AR and VR applications taken from publications of the XR HCI community (see Table 1) as a basis to inform the creation of scenarios. We focused on future XR **head-mounted display (HMD)** scenarios rather than existing smartphone use cases to investigate which properties would enable or amplify future dark patterns in XR applications.

Our analysis (see Section 3.4) resulted in the four XR-specific property clusters *spatiality*, *perception*, *physical/virtual barriers*, and *XR device sensing* (see Figure 3).

In the following sections, we differentiate between the application of properties for *causing* versus *amplifying* a manipulation. The former are XR properties that are essential for achieving a scenario. The latter are XR properties that worsen the consequences of a manipulation or increase its effectiveness by amplifying the cause of the manipulation. In total, 22 scenarios described the application of a single XR property, 12 scenarios combined 2 or more XR properties, and 8 scenarios did not apply XR properties to *cause* a manipulation. Additionally, four scenarios applied XR properties for *amplifying* an existing dark pattern, nine scenarios further amplified a manipulation caused by XR properties, and four scenarios did not describe any XR properties. The following sections provide a more detailed description for each XR property cluster.

4.1 Spatiality

This cluster summarizes properties related to *unbound content*—as opposed to having content limited to a 2D screen—which addresses 3D registration of virtual content, distance and position, and *interaction* with a strong focus on whole body interaction. In total, 12 scenarios used properties of this cluster to *cause* manipulations and 2 used them to *amplify* manipulations. Popular manipulations included *steering*, *positioning*, or *rerouting* users, for example, through placing content strategically. Such mechanisms include placing content out of reach or too close to the user; forcing users to react to static or dynamic virtual objects to avoid a collision inducing strong haptic feedback; buttons that move or change their size shortly before a user reaches them; or static content blocking a user's view or path and requires them to walk around or destroy it.

Some scenarios paired elements of spatiality with user specific abilities like reaction time or predicted focus points. In the scenarios, this combination leads to personalized manipulations and increased their anticipated effectiveness for a specific user, by making it impossible to avoid

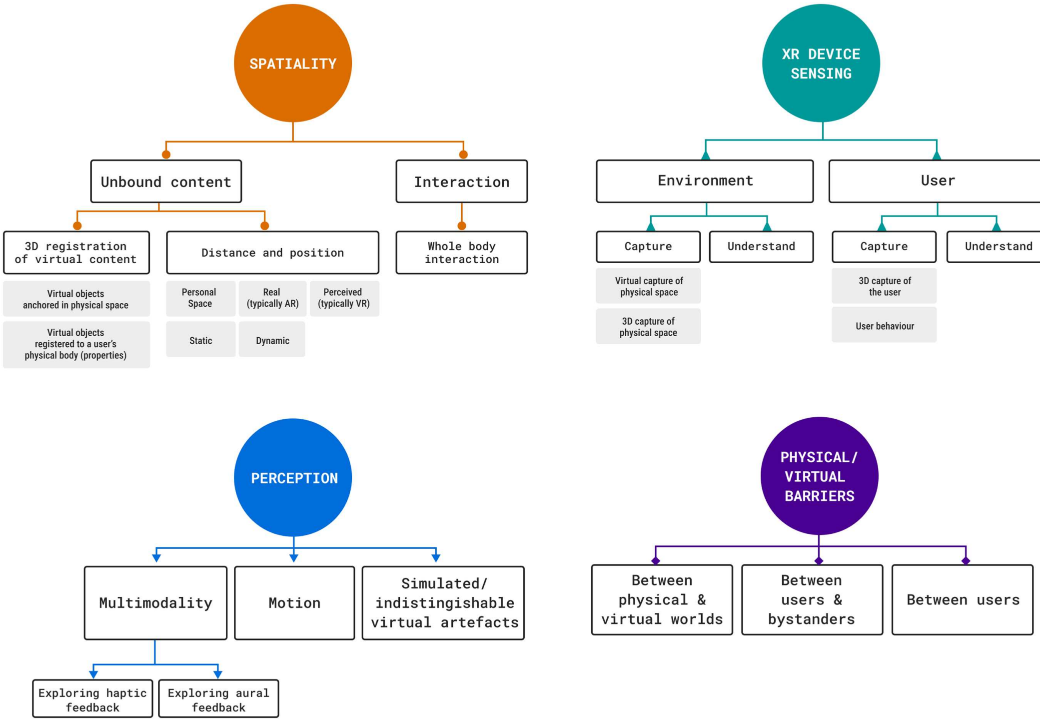

Fig. 3. XR properties identified in the scenarios clustered in the four topics perception, spatiality, physical/ virtual barriers, and XR device sensors. Each colored circle represents a thematic cluster regarding identified XR properties that could be leveraged to create or amplify dark patterns (see Table 3). A graph depicting 4 clusters that provide more detail regarding the identified XR properties. The four clusters are depicted as tree diagrams with 1 to 3 levels of detail.

or activate content. For example, the scenario *Spatial ads* (WS10_3) places advertisements in the predicted visual focus area of the user, impeding them from completing their current task. Hereby, virtual objects were either anchored in a physical/virtual space or registered to a user's body properties or position.

Regarding whole body interaction, manipulations were not only described regarding a user's physical abilities or properties but also *utilized intrinsic meaning* of a gesture, posture, or movement [37]. For example, the scenario *Puppy Crushing* (WS09_4) requires a user to hurt or kill a simulated virtual puppy to decline an offer by actively stepping on it.

4.2 Perception

We grouped properties related to multimodality, motion, and simulated indistinguishable virtual artifacts under the umbrella term *perception*. In total, nine scenarios described manipulations assigned to this cluster, with additional four using them to amplify manipulations.

Motion was operationalized as *punishment* in the scenarios proposed by our participants. For example, the scenario *Discourage user action in immersive experiences* (WS05_6), developers programmed a game in a way that it punished user choices which were not favored by the beneficiary. Some options provided by the respective application were programmed to cause simulator sickness after a user selected them, whereas others did not. In this scenario, it caused a user to either undo their choices or to select the latter options if they were again presented with a similar choice dialogue in the future. In contrast to this subtle and rather indirect punishment, participants also described direct force or tactile feedback of an intensity ranging from subtle nagging to being painful, provided directly to a user's head through the HMD as a consequence of their actions.

Guiding a user's *attention* was a common theme among scenarios applying additional modalities, for example via *directional* haptic, aural, or visual feedback. Our participants utilized both feedback induced by the worn XR devices as well as other physical elements of an interface, such as noise or wind caused by a drone (WS03_3). However, multimodal feedback was also described as a means to *overwrite* a user's senses. This manipulation type was often combined with the *physical/virtual barriers* property (see next section), as it hindered users to check if the sensed input was resulting in a nonconforming perception. The scenario *Texture Manipulation/False Textures* (WS03_1) describes in greater detail, how a VR e-commerce application renders fabrics to create a perception of superior quality compared to their physical counterpart. This deception leads users to purchase low-quality fabrics for a high price. However, our participants also identified the lack of multimodal feedback as a potential lever to increase sales for a beneficiary: In the scenario *Pay more to experience more* (WS05_5), a VR game comes with a free trial that allows users to experience and get used to the full multimodal feedback potential of their hardware. After the free trial runs out, the game reduces the bandwidth of multimodal output rendering the application as a poorer experience to make users purchase this functionality as upgrade or subscribe to a respective service to restore the previously experienced quality of haptic feedback.

Finally, our participants detailed simulated indistinguishable artifacts in the context of causing moral dilemmas a user must overcome to decline an offer. Scenarios reporting on this property mentioned photorealistic renderings as well as realistic behavior and sound. The scenario *Puppy crushing* (WS09_4), for example, adds overly detailed visual and aural effects on top of a required kick toward a photorealistic render of a puppy to decline an offer. The scenario describes this setting as an impactful appeal to a user's morality.

4.3 Physical/Virtual Barriers

This cluster addresses barriers between physical and virtual worlds, between users and bystanders, and between users sharing the same experience. In this case, such a barrier was described as a direct result of users wearing HMDs. While none of the scenarios describe such barriers as amplifiers, 10 scenarios detail them as causes for manipulations. Barriers between physical and virtual worlds were often applied when a user was unable to verify simulated content and feedback against physical world conditions or perceived reality.

In *Selling a cat in a bag* (WS06_1), a VR e-commerce application renders tactile and visual replicas of physical fabrics. This scenario describes how such renderings are generated to be more or less appealing to a user's preferences even though the physical fabric has slightly different features. However, the user has no means of validating the presented content, therefore, the manipulation remains unnoticed. In a similar line, *Switch-out and haze* (WS01_1) addresses the barrier between users and bystanders and describes a situation in which player 1 and player 2 compete against each other in a shared physical space. While player 1 is fully immersed in a VR game they access through an HMD, player 2 interacts with player 1 and the game environment through operating a 2D screen device such as a smartphone. Manipulative bystanders, however, switch places with player 2 and start to bully player 1. Due to the VR HMD, player 1 does not realize that player 2 has been substituted.

Finally, barriers between users made use of the ability to show or hide information about other users or the application's context while sharing the same experience. This information imbalance was established by either requiring the purchase of data to learn about a rival's weaknesses in a game setting (WS01_2) or create the need to purchase virtual goods to support party members by leaking information based on the context (WS08_1).

As a final remark regarding barriers, we want to highlight that such barriers could not only be used to create a discrepancy between the physical and the virtual world. As WS01_2 and WS08_1 indicated, it is also possible to create a discrepancy between virtuality and an application's underlying or prevailing reality. We will further discuss this point in Section 8.3.

4.4 XR Device Sensing

Sensor data from XR devices originating from optical, depth perception, eye tracking, or location sensors were applied in scenarios describing *capturing* and *understanding* information about the *environment* as well as *users*. Despite the reliance on sensing and the further processing of captured/recorded data not being XR-specific (e.g., such sensing can also be found in Internet of Things applications or smartphones), our participants mentioned the range and richness of collected information, which was sometimes combined with real-time interpretation, as being unique to XR. This multitude of worn sensors provides access to a breadth of personal/environmental information which was leveraged in the scenarios. In fact, scenarios containing sensor-specific elements always mention recording in combination with inference, further processing, and deriving further insights to enable user manipulation. In total, eight scenarios applied this property as the cause for a manipulation, whereas seven mention it as amplifier. Additionally, scenarios in this category often described XR device sensing as an *enabler for refining* manipulations of spatiality, perception, or physical/virtual barriers. For example, *Sneak and Purchase* (WS01_2) describes a VR game that dynamically adapts its difficulty to a user's weaknesses until they get stuck. This leaves them with having to either overcome the challenges that are personalized to exploit their specific physical and perceptional limits or purchase respective power-ups and items. Other scenarios detail the misuse of visual data for building more robust user profiles through learning about their habits and abilities (WS07_2), interpolating information restricted by a user's privacy settings (e.g., locating a user based on computer vision information even though GPS data tracking is switched off (WS02_2)), or semantically understanding a user's private space or current tasks to target ads (e.g., WS10_2).

5 Evidenced Dark Patterns in the XR Scenarios

To see if our participants described new dark patterns or if they reapplied and adapted existing ones in their co-designed scenarios, we performed a thematic analysis from which we excluded 4 of the 42 scenarios because they did not describe dark patterns (see Section 3.4). We found that from the remaining 38 co-designed scenarios, 34 (89.5%) could be assigned to more than one dark pattern listed in Mathur et al.'s pattern catalog [55] and the other 4 scenarios (10.5%) to only one of the listed dark patterns. This ambiguity led us to further exclude implementation-specific details from dark pattern definitions by clustering them regarding their manipulative mechanism (adapted from Gray et al. [31] and further detailed in Section 3.4). We then related the identified clusters to the XR-specific properties (see Section 4) to single out amplifications and adaptions our participants performed when incorporating the known dark patterns into XR scenarios. The dark patterns we identified in the co-designed scenarios are a subset of Mathur et al.'s consolidated pattern catalog [55] that we used as a basis for our analysis. The known patterns we identified in our scenarios stem from the following original submissions: Gray et al. [30, 31], Brignull et al. [7], Conti and Sobiesk [22], Chatelier et al. [15], Norwegian Consumer Council [29], Greenberg et al. [33], Bösch et al. [10], Zagal et al. [78], and Mathur et al. [54].

If a dark pattern was applied in a co-designed scenario but the original definition lacked a manipulative characteristic or our participants did not utilize XR properties, we did not further classify the respective dark pattern. Table 3 lists, defines, and relates the resulting eight manipulative mechanisms to the XR-specific properties identified in Section 4.

Table 3. Our Manipulative Mechanism Taxonomy Based on a Subset of Mathur et al.'s Consolidated Pattern Catalog [55] Related to the XR-Specific Properties Spatiality, Perception, XR Device Sensing, and Physical/Virtual Barriers (Section 4)

Manipulative mechanism	**Definition**	**N**	**Matching dark patterns from Mathur et al. [55]**	***Spatiality***	***Perception***	***XR Device Sensing***	***Physical/Virtual Barriers***
Persuading	Persuading users to perform actions, for example through nagging [31] or shaming	5	Confirmshaming [7], blaming the individual [15], nagging [31]	•	•	•	
Imbalancing options	Includes dark patterns and strategies to promote options or their impact over others, such as making them seem more attractive, requiring less effort, being easier to select, or resulting in more pleasant and beneficial consequences	15	Obfuscation [22], interface interference [31], obstruction [31], price comparison prevention [7], framing [29], rewards and punishment [29], comparison obfuscation [15], attention diversion [15]	•	•	•	
Baiting	Luring users into a situation through the presentation of desirable options, features, or content	4	Bait and switch [33], Bait and change [15], Bait and switch [7]	•	•	•	•
Lying	Presenting untrue content, features, properties, and so on to deceive a user	4	Misrepresenting [30], trick [22]	•		•	
Directing attention	Shift or steer a user's attention away from their current points of interest, tasks, actions, and so forth.	14	Misdirection [7], misdirection [54], distraction [22], attention grabber [33], controlling [30], interruption [7]		•	•	
Disguise	Disguising content, actions, or data collection; activities a system performs and masks from the user	11	Shadow user profile [10], disguised data collection [33], forced continuity [7], sneak into basket [7], disguised ads [7], impersonation [78], camouflaging advertising [15], hidden coins [7]		•	•	
Requiring a detour	Rerouting a user (physically or mentally) to make them pass by options, locations, and so forth.	2	Milk factor [33]	•		•	
Exploiting the drive to succeed	Utilizing introduced or existing obstacles a user encounters while pursuing a goal (see *forced action* [31])	11	Nickle and diming [30], forced action [31], grinding [78], monetized rivalries [78], pay to skip [78], pre-delivered content [78]			•	•
Not classified	—	3	Roach motel [7], making personal info public [33]	•			•

N = number of occurrences in our data set.

Our clustering revealed that existing dark patterns as commonly described in literature are often based on the same manipulative mechanism but their definitions differ in details related to their exact implementation. This limits their identifiability as novel or existing dark pattern in the context of a new technology and increases the ambiguity of classifications and complexity of potential regulations. For example, *Framing* [29] and *Attention Diversion* [33] both describe that some choices are highlighted to influence a user's decision. However, Attention Diversion focuses on the graphical UI elements, such as color and button size, whereas Framing is defined as "focusing on the positive aspects of one choice, while glossing over any potentially negative aspects" [29]. As it is debatable if both definitions describe the same or distinct patterns, we refrain from reporting on our findings as dark patterns if similar designs have been published in prior work. Instead, we fall back to reporting *manipulative mechanisms* and, if Mathur et al.'s consolidated dark pattern catalog does not list a fitting description of a reported dark pattern, we describe them as *emergent patterns*.

Finally, to increase compatibility with existing works that aim to establish technology-agnostic taxonomies for identifying and describing dark patterns, we compared our observations to Gray et al.'s preliminary ontology [32]. This ontology differentiates between **high-level (HL)** patterns that "are the most abstracted form of knowledge, including general strategies that characterize the inclusion of manipulative, coercive, or deceptive element" [32], and **meso-level (ML)** patterns that "bridge high- and low-level forms of knowledge and describe an angle of attack or specifc approach to limiting, impairing, or undermining the ability of the user to make autonomous and informed decisions or choices" [32]. The "most situated and contextually dependent form of knowledge, including specific means of execution that limits or undermines user autonomy and decision making" [32] connote the third level: low-level patterns. However, our subset of dark patterns and their application in the co-designed scenarios was not congruent with the proposed three-tier ontology (see Tables 4 and 5). For example, based on the co-designed scenarios, we identified *nagging* [31] as a manipulative mechanism that utilizes similar levers like *blaming the individual* [15] and *confirmshaming* [7]. Consequently and under consideration of their original definitions, we clustered the three dark patterns under the manipulative mechanism *persuading*. However, persuading, as defined in our work, overlaps with Gray et al.'s HL patterns *social engineering*, *nagging*, and *interface interference*, and further includes Gray et al.'s ML patterns *personalization*, *emotional or sensory manipulation*, and *shaming*. We consider this overlap of Gray's HL and ML patterns in our clustering to stem from the speculative character of the co-designed scenarios and the fact that they describe potential XR dark pattern variants rather than existing implementations. However, this overlap also indicates that XR dark patterns could utilize chaining of manipulative mechanisms and combine multiple XR properties to create more effective manipulations (see Sections 7, 8.2, and 9.2).

5.1 Reusing and Adapting Dark Patterns for XR

In the co-designed scenarios, our participants described several adapted versions of existing dark patterns that were enriched with XR-specific properties to function in spatial interfaces. Therefore, our participants often adjusted dark patterns of the clusters *directing attention* and *imbalancing options*. An example is the adaption of dark patterns assigned to *imbalancing options* to a 3D setting as described in the scenario *Preference toward one choice or another* (WS05_4): In addition to using visual features, the dark pattern applied directional haptic feedback to make a user reposition themselves or at least turn their head. By directing a user's attention toward specific interface areas through utilizing both visual and haptic stimulation, the scenario describes how a user would focus more on such highlighted options and therefore prefer them over others. Other aspects our participants utilized often to ease or hinder access to certain choices were distance and positioning. The scenario *Location manipulation* (WS09_1) describes an application that positions options preferred by the beneficiary closer to a user, whereas unwanted ones require the user to

Table 4. Our Manipulative Mechanism Taxonomy in Relation to Gray et al.'s Preliminary Ontology [32]

Manipulative Mechanism	**Definition**	**Relation to Gray et al. [32]**
Persuading	Persuading users to perform actions, for example through nagging [31] or shaming	HL: *Social engineering, nagging, interface interference* ML: *Personalization, emotional or sensory manipulation, shaming*
Imbalancing options	Includes dark patterns and strategies to promote options or their impact over others, such as making them seem more attractive, requiring less effort, being easier to select, or resulting in more pleasant and beneficial consequences	HL: *Interface interference*
Baiting	Luring users into a situation through the presentation of desirable options, features, or content	ML: *Bait and switch*
Lying	Presenting untrue content, features, properties, etc. to deceive a user	—
Directing attention	Shift or steer a user's attention away from their current points of interest, tasks, actions, and so forth.	HL: *Interface interference, forced action* ML: *Manipulating visual choice architecture, attention capture*
Disguise	Disguising content, actions, or data collection; activities a system performs and masks from the user	HL: *Sneaking*
Requiring a detour	Rerouting a user (physically or mentally) to make them pass by options, locations, and so forth.	—
Exploiting the drive to succeed	Utilizing introduced or existing obstacles a user encounters while pursuing a goal	HL: *Forced action* ML: *Gamification*

As our taxonomy is based on a not exactly congruent set of dark patterns, the identified clusters overlap with but are not the same as Gray et al.'s three-level ontology [32]. HL = high-level pattern, ML = meso-level pattern.

physically walk or stretch. Dark patterns assigned to the cluster *directing attention* predominantly used perceptual properties to steer a user's attention, such as haptic feedback, sound, and motion.

5.2 XR as an Amplifier of Known Dark Patterns

In general, scenarios described a manipulation as being more effective when a combination of properties was used or several manipulative mechanisms were chained together.

Based on our data, some scenarios amplified existing manipulative mechanisms through presenting them as *super-realistic*. *Persuading*, for example, applied elements of perception and spatiality to present realistic UI elements. To further amplify the effectiveness, the scenario Puppy Crushing (WS09_4) also made use of *intrinsic gestural meaning* and *psychological realism* in addition to a required *whole-body motion*—the user needed to step on a super-realistic, animated puppy to decline an offer. This example shows how leveraging embodiment and realism allows XR to utilize new mechanisms for raising the difficulty of making a decision. While similar dark patterns originating from 2D frequently apply visually or textually modified UI elements or preselected options (e.g., interface interference [31]), users can be forced to act with their hands and whole body in XR. This might cause psychological discomfort and a feeling of greater responsibility for both the action as well as the consequences [68, 75].

Additionally, the intrinsic meaning of a gesture, for example, stepping on someone or something, can evoke associations with unwanted or disliked feelings, such as superiority, contempt, or spite. This could further increase a user's discomfort and hesitation to perform such actions as it might contradict their current state of mind and ethical attitude. A comparable issue on 2D screens might have a milder impact since the potential lack of immersion, presence, and embodiment makes it easier for users to distance themselves from their actions and the resulting consequences [68, 75].

Another theme for amplification was *personalizing the manipulation* through the application of XR device sensor data. For example, scenarios detailing elements of *lying* and *baiting* utilized a combination of overwriting users' visual and tactile senses (perception and physical/virtual barriers) to construct experiences that could not be verified against the physical world. Those manipulations were then amplified through the application of user-related sensor data to personalize the experience. The scenario WS03_2, for example, describes how a VR e-commerce application adapts the presentation of weight and texture of clothes or fabric to a user's personal preferences to make them purchase an item. Finally, amplifications were utilizing that the *user is surrounded by the UI* rather than experiencing applications separated through a screen, and that *their whole body could directly be affected.* This was described, e.g., in scenario WS05_3 where the application forced users to experience ads after colliding with them. Similarly, WS04_1 positions advertisements strategically, i.e., in a user's focal point, so that it is difficult or impossible to avoid them (directing attention, exploiting the drive to succeed, imbalancing options). Finally, WS05_6 purposefully causes simulation sickness or inflicts physical pain through strong haptic feedback to a user's head to punish and manipulate their choices (imbalancing options).

5.3 Emergent Patterns of Manipulative Mechanisms in XR

Based on our workshop and Mathur et al.'s pattern catalog [55], we identified in total 15 co-designed scenarios that we summarized as 10 potentially new emergent patterns of manipulative mechanisms. Table 5 provides an overview as well as a description for each emergent patterns that are further detailed in the following paragraphs. As those emerging patterns are also adaptable to other technologies, we further highlight the aspects our participants perceived as XR-specific when creating the respective scenarios.

Persuading. Moral dilemma (1) was caused and utilized by combining elements of perception, spatiality, and XR device sensing to create realistic and context-aware appeals to a user's morality. The goal was to persuade a user to perform certain actions favored by a beneficiary. As we further discuss in Section 8.3, this pattern utilizes multiple XR-specific features like whole-body interaction combined with the intrinsic meaning of gestures, emotional manipulation, context-awareness, and photorealism to appeal to a user's morality (WS02_4, WS09_4).

Imbalancing Options. The emergent pattern *strategic content placement* (2) learns gaze behavior and predicts a user's focus points to strategically place content in a way that a user cannot avoid it (WS10_3). This increased the likelihood of accidentally activating such content. Such dark patterns require precise eye-tracking, contextual awareness and unlimited display space as it is available in XR devices. Similarly, *spatial imbalance of options* (3), as previously described, increased the physical effort required for a user to make a choice unfavored by a beneficiary. This effort was further personalized and dynamically adapted regarding a user's physical or mental abilities, preferences,

Table 5. Ten Emergent Dark Patterns We Identified as Being Novel Based on Mathur et al.'s Pattern Catalog [55]

Nr.	Name	Description	Related HL pattern [32]	Related ML pattern [32]
Persuading				
1	*Moral dilemma*	A system uses superrealism, context-awareness and/or multimodality to appeal to a user's morality.	Social engineering interface interference	Personalization emotional or sensory manipulation
Imbalancing options				
2	*Strategic content placement*	A system learns a user's gaze behavior and places content strategically to make it difficult to avoid.	Social engineering interface interference	Personalization manipulating visual choice architecture
3	*Spatial imbalance of options*	A system places options strategically to make users favor (easy to reach) or unfavor (difficult to reach) their selection.	Interface interference	Manipulating visual choice architecture
Baiting				
4	*Immersive bait and snap*	A system baits a user with disguised ads and provides a multimodal experience to prolong the time a user spends on it.	Sneaking	Bait and switch
Directing attention				
5	*Forced data capture*	A system causes a user to relocate or turn to specific positions to allow sensors the recording of missing information.	Forced action sneaking	Attention capture —
Disguise				
6	*Hiding behind the novelty factor*	A system performs hidden activities and blames it on the user's inexperience with the medium.	Sneaking	—
7	*Personal shadow profiles*	A system secretly interpolates information from accessible sources to circumvent privacy restrictions.	Sneaking	—
Exploiting the drive to succeed				
8	*Monetizing camaraderie*	A system displays selected information and emotionally pressures a user to purchase items for their party members to proceed with the game.	Social engineering forced action Interface interference	Personalization gamification emotional or sensory manipulation
9	*Combatants' weaknesses purchase*	A system adapts its difficulty based on a user's weaknesses, hindering their advances, to make them purchase power-ups, items, or weakness profiles from other users.	Social engineering forced action	Personalization gamification
10	*Realism surcharge*	A system limits the multimodal feedback of a user's hardware after a free trial, forcing them to purchase upgrades or sign-up to a service to regain the previous experience.	Forced action	Forced registration

Related HL or ML pattern originate from Gray et al.'s preliminary ontology [32]. The emergent dark patterns might also be applicable in other technology than XR.

habits, and surroundings and does not only require XR devices' sensing abilities but also a spatial and 3D display that is able to place content in physical distances (WS09_1, WS09_2, WS04_1).

Baiting. Immersive bait and snap (4) utilized XR's multimodality to subtly bait users, keep them focused on an interruption that behaves similar to their current content, and extend their time spent on it, such as advertisements. As such, *immersive bait and snap* predominantly applies to XR as it exploits spatiality and perception, as well as the psychological impact that 3D immersive content could have (WS05_3, see moral dilemma).

Directing Attention. A user's attention could be targeted toward areas of their surroundings that were not recorded before, for example, to collect spatial data. To direct respective XR device sensing a user could be prompted to turn their head. Such a (5) *forced data capture* manipulation can also be caused by another user sharing the same physical or virtual space (WS10_1) and result in the creation of personal shadow profiles (7).

Disguise. Three scenarios *hid behind the novelty factor* (6) of XR to disguise application's activities or content (WS06_2, WS06_3, WS06_4). This was paired with blaming the user for hidden costs, requiring extra payment for the XR feature (XR surcharges) after an application had been experienced, or utilizing the unfamiliarity with XR and the resulting user's insecurity and potential clumsiness when moving through and interacting with the application to add items to their virtual shopping cart. Furthermore, WS02_2 described the disguised interpolation of position and tracking data from video feeds and images using computer vision and image recognition algorithms if a user's privacy setting restricted the direct access to GPS tracking. As a consequence, the application generated *personal shadow profiles* (7) based on alternative data sources and cues that allowed to infer a user's current location. *Personal shadow profiles* does apply to all wearable computing devices that log user-related data but is enriched with visual, 3D positional, and ambient data in XR usage contexts.

Exploiting the Drive to Succeed. Manipulative mechanism variants emerging from applications directed toward gaming, for example, benefited from *monetizing camaraderie* (8) (as opposed to monetized rivalries [78]). They did so by making use of physical/virtual barriers between the physical and virtual world, users, and the resulting information imbalance: A player was prompted with status information from party members and emotionally pressured into purchasing items and power-ups to proceed and succeed with a game. *Combatant's weaknesses purchase* (9), however, recorded and analyzed a player's weaknesses and strengths, adapted its difficulty accordingly to hinder a player in proceeding and pressuring them into buying power-ups and items. Additionally, other players could purchase this information to gain advantage. Finally, in *realism surcharge* (10), a game baited a user with the full span of multimodal feedback by utilizing their XR hardware capabilities to their maximum. After a free trial, however, the hardware functionality was reduced and required a user to purchase or subscribe to a respective service to restore the previous quality of multimodality.

6 Probability–Severity Assessments of Scenarios

Finally, we asked our participants to anticipate the risk of their scenarios for the next 5–10 years to learn about aspects that need to be considered in potential dark pattern mitigation strategies and regulations. In our study, risk was expressed through probability, i.e., the likeliness of the scenario (1 = somewhat improbable, 9 = very probable) to be applied in commercial XR applications, and the severity (1 = insignificant, 9 = catastrophic) regarding the impact of a scenario for an undefined affected entity. We did not further specify the context of assessment or the perspective our participants should take as we were also interested in their opinion about relevant questions

Fig. 4. The scatter plot depicts the combined risk assessment of the co-designed scenarios across all workshops. The highlighted scenarios (black) WS10_1, WS10_2, and WS10_4 were duplicated as their risk was assessed based on how well the beneficiary knew the manipulated party (WS10_2: well-known: less severe; stranger: more severe), the application domain (WS10_4: consumer app: less severe, less probable; military: more severe, more probable), and the sensitivity of data (WS10_1: general privacy that lead to spending more money: less severe; health data: more severe). The color coding relates the scenarios to the identified dark pattern clusters in Section 5. Patterns with the label *not classified* (white) did not utilize XR-specific properties or the original definition lacked a manipulative mechanism (see Section 5). This figure shows the probability-severity matrix combined with the participants' rating. Probability lies on the y-axis with highest probability on the top, severity on the x-axis with highest severity on the right. The participants' ratings accumulate in the center of the top row which correlates with a very probable - major severity rating.

to ask when evaluating their scenarios' risks. Both dimensions consisted of a 9-point scale as we let participants place scenarios between the five discrete values of each axis (see Section 3.2). If participants rated the probability and severity differently in a single scenario depending on the affected party, application domain, or type of data, we encouraged them to duplicate it, add the affected user group, and place it in two different sections of the probability-severity matrix (see Figure 4).

6.1 Strategies for Assessing Risks

When assessing the risk of their scenarios, participants followed a general three-step approach. First, they agreed on a perspective they would take, such as a user-centric (harm) or technology-centric (trust in technology) perspective. They further agreed on the minimum and maximum ranks of the axis. For severity, our participants often used an implicit scale for type of harm ranging from a dark pattern annoying a user, over financial harm up to causing physical harm or death. While assessing the risk, some pairs of participants additionally switched perspectives and observed probability and severity from a company's or stakeholder's perspective. After reaching an agreement regarding the observational perspective as well as both the maximum and minimum of the scales, the participants took turns in picking scenarios. They discussed each and placed them on the matrix. The scenarios were arranged in relation to each other which sometimes led our participants to reassess a risk if additional aspects were emerging from the discussions.

6.2 Factors Influencing the Probability Ranking

Our participants discussed four major aspects regarding the probability of their scenarios to be applied in commercial XR applications in the next 5–10 years: technological feasibility, reference implementations and situations, profitability of the underlying business model, and existing or emergent regulations.

- *Technological feasibility and required effort.* When discussing the probability of scenarios to become implemented, our participants always considered required effort or ease of implementation. If a manipulation was easy to be programmed given the current advances in technology (WS01), they ranked a scenario's probability very high. In the same line, they discussed if some scenarios required specific hardware adaptions (WS05, WS08) or if the resulting manipulation is easy to be scaled up (WS09).
- *Reference implementations and situations.* Participants often related the described scenarios to already existing situations, such as aggressive ads on mobile phones (WS05), advertisement strategies in Pokémon Go (WS09), or abuse or misuse of social media (WS01). If a comparable situation already existed, they increased the probability rating for their scenario.
- *Profitability of the underlying business model.* Closely related to effort, our participants considered the profitability of the underlying business model in case a company was the beneficiary of a scenario, and assessed the financial feasibility (WS01, WS04, WS05), reputational impact or trust (WS01, WS06, WS08), and the persistence of a potential user base (WS01, WS06).
- *Existing or emergent regulations.* Finally, participants also speculated that the probability of encountering severe dark patterns decreases over time as *"companies will always try to take advantage of new technology"* (WS06) in the beginning but "legislation will pretty quickly pick [obvious and severe manipulations] up" (WS06) and regulate them.

6.3 Factors Influencing the Severity Ranking

To assess the severity of the scenarios, we asked participants to rank them on a 9-point scale ranging from insignificant (1), to minor (3), to major (5), to severe (7), and finally catastrophic (9). If participants could not reach a clear decision, they placed the scenarios between two options. In total, they discussed the four mutual influential topics *type of harms and consequences*, *application's context*, *detectability and already existing mitigation strategies*, and *user-related aspects*.

- *Types of harm and consequences.* Participants mentioned the greater context and scale of severity which needs to be specified in order to assess the seriousness of dark patterns, such as impact on the individual, society, democracy, or environment, because "there is a fair bit of choice" (WS06). Additionally, harms which are perceived as catastrophic for some individuals might not be perceived as a severe issue for others (WS09, WS08). Furthermore, our participants often related back to an implicit spectrum ranging from "being just annoy[ed]" (WS05) over financial loss and property damage, social and mental harm (i.e., trust, reputation, stigmatization), to physical harm, death, and "destroying the world" (WS06) as the most severe outcomes of a scenario. They further discussed if a manipulation affects a user beyond the application, for example, through ruining relationships or impacting ones social status, or if it takes immediate effect (WS04).

— *Application's context.* Our participants agreed on the *context-dependency* of an application as a relevant factor for severity. For example, as pointed out in WS10, targeted ads as described in WS10 might be harmless most of the time. However, under specific circumstances, they could cause serious consequences for individuals. In the example discussed in WS10, a father learned about the pregnancy of his daughter due to targeted ads, which lead to severe consequences for the daughter.
— *Detectability and existing mitigation strategies.* Our participants anticipated subtle manipulations that remained unnoticed by the user as more severe compared to obviously executed ones as the latter remained undetected (WS09). Further, well-known dark patterns with existing counter measures or regulations were rated as being less severe (WS07).
— *User-related aspects.* Finally, participants considered user-related properties as influential factors informing the scenarios' severity, such as the availability of a fall-back option, i.e., an equivalent and safe application or service that could be used instead (WS06, WS07), how many people are being reached through the manipulation including up- and downscaling as well as distribution (WS03, WS07), and targeting the vulnerability of specific user groups (WS03, WS04, WS08).

7 Summary of the Findings

This section summarizes our findings which we detailed above. Further, we provide answers to our three research questions. We will discuss and consider implications of our findings in the Sections 8 and 9.

RQ1: What existing or new dark patterns could emerge in XR application scenarios? To examine if the co-designed scenarios incorporated existing or new types of dark patterns, we assigned Mathur et al.'s collection of dark patterns [55] to the scenarios whenever possible. As the resulting assignments revealed an ambiguity of dark patterns regarding their descriptions, we further clustered these based on their manipulative mechanisms [31] to more robustly identify emergent dark patterns in XR. This activity resulted in the eight manipulative mechanisms persuading, imbalancing options, baiting, lying, directing attention, disguising content, actions or data collection, requiring a detour, and exploiting the drive to succeed (see Table 3). Those mechanisms were repeatedly applied in the co-designed scenarios and respective existing dark patterns were often combined to cause a user manipulation.

In total, we identified 39 dark patterns previously reported in related work and listed in Mathur et al.'s pattern collection [55] as well as 10 emergent dark patterns listed in Table 5 that were unique to our dataset. While some of those 10 emergent dark patterns displayed the usage of XR-specific properties (see RQ2), they were still grounded in the eight manipulative mechanisms also applied by already existing dark patterns. We conclude that XR dark patterns are predominantly variants of known dark patterns that leverage both XR properties and manipulative mechanisms that were already proven to work well in 2D media. However, as the following paragraph describes, adaptions were made to reapply known dark patterns to XR.

RQ2: Which XR-specific properties would enable or potentially amplify dark patterns in XR applications? We identified the four XR-specific property clusters—*spatiality*, *perception*, *physical/virtual barriers*, and *XR device sensors* (see Figure 3)—that were frequently applied to create the manipulations described in the co-designed scenarios. The properties were applied following different *tactics*:

- Manipulations based on *spatiality* frequently steered, positioned, or rerouted users by placing content strategically. Such content was additionally manipulated dynamically, e.g., by changing position or size shortly before a user activates it. Furthermore, collisions with virtual elements were paired with strong haptic feedback to cause a user's discomfort.
- *Perceptual manipulations* operationalized motion and haptic feedback as punishment, for example, by causing simulator sickness or providing unreasonable strong feedback to a user's head. Directing a user's attention or manipulating tactile perception was also frequently applied as a tactic, as well as providing hyper-realistic content indistinguishable from the physical surroundings (see Section 8.3).
- Manipulations leveraging *physical/virtual barriers* created a mismatch between physical and virtual properties, bystanders and/or users by displaying wrong information, emphasizing or diminishing properties, or showing users different subsets of information without the option to verify or complete them.
- Finally, manipulations based on *XR device sensing* primarily made use of the richness of information combined with real-time processing, evaluation and interpretation. As such, interpolating information a user restricted with dedicated privacy settings was a potential tactic, as well as personalizing manipulations based on a user's ability, habits, and the surrounding environment.

We further observed that XR-specific properties were combined to cause a more complex but difficult to avoid manipulation. A frequent tactic was personalizing such malicious designs based on a user's bodily or perceptual ability by placing UI elements slightly out of reach, exactly in the user's focal point, or making them impossible to avoid based on a user's personal reaction time. However, we also learned that manipulative mechanisms were not only amplified through the usage and combination of XR-specific properties but also chained together to increase their effectiveness or efficiency.

RQ3: What aspects determine the probability and perceived severity of dark patterns in XR design? With regard to the risk assessment, probability rankings were influenced by the anticipated technological feasibility and required effort including the ease of upscaling a manipulative implementation. Further, existing implementations or existing situations that could be referenced impacted the ranking, as well as the profitability of the underlying business model. Finally, our participants evaluated probability regarding existing and emerging regulations, and the extent to which such activities could be legally prohibited.

Severity rankings were anticipated by taking harms and their consequences for specific user groups or perspectives (e.g., societal impact) into account. Further, their ranking was based on an application's context, the detectability of such manipulative behavior including existing mitigation strategies, and user-related aspects such as the availability of fall-back options, the amount of people being influenced by a respective manipulation, ease of distribution, and user vulnerability.

We conclude that to rate a dark pattern's risk and create applicable regulations, it is necessary to consider both business-related aspects as well as the reach of consequences regarding users and their context. We further discuss potential implications in Section 9.2.

8 Discussion

In addition to the seven scenarios described by Eghtebas et al.'s work [25], our 42 scenarios depict diverse settings in which XR's unique properties can be applied to manipulate users to their expenses. However, like other work that followed speculative approaches, those scenarios suffer from the Collingridge dilemma [44], which refers to the fact that we cannot predict if XR dark patterns will manifest and impact users as envisioned in the dark pattern scenarios unless they are implemented and put to use. We addressed this with a two-fold approach: firstly, we based our workshop on applications that were previously implemented and, therefore, at least possible to come into existence from a technical point of view. Secondly, by asking our experts to assess the probability and severity of the dark pattern scenarios to estimate their potential applicability. The following sections consider our findings against related work. We first summarize our insights regarding the growing risk of dark patterns in XR, before reflecting on the implications and challenges it poses to current and future everyday XR users in Section 9.

8.1 The Growing Risk of XR Dark Patterns

Across 10 workshops, 20 experts in XR and DD co-created and reflected on in total 42 scenarios, exploiting *spatiality*, *perception*, *XR device sensors* and *physical/virtual barriers* to manipulate users into undesired or unintended outcomes.

As our findings and prior work investigating preceptual manipulations [16, 25, 73] and dark patterns in AR [77] demonstrate, DD is not only applicable to XR but also has the potential to thrive on such platforms. XR has unique affordances compared to prior computing technologies, as our identified four XR properties attest to, having the capacity to both sense and understand our environment, and react to this—all while portraying virtual elements with an ever-increasing degree of perceptual and psychological realism (perception, spatiality) and the ability to overwrite a user's sensory input (physical/virtual barriers).

Those four XR-specific properties (RQ2) and the eight manipulative mechanisms (RQ1) align well with Eghtebas et al.'s three overarching themes (situatedness of information, altering perception, and ubiquitous sensing) [25] in potential XR dark patterns research. While XR has the potential to overcome the constraints of physical displays and instead be enacted in 3D across both virtual and real worlds anywhere and anytime, our scenarios also picked up some of the concerns voiced by Greenberg et al.'s work that investigates dark patterns in proxemic interactions on public displays [33] (e.g., WS08_3 *Forced Shouldersurf*). We, therefore, imagine future work investigating dark patterns in XR to increasingly focus on the ubiquitous characteristics of emerging XR applications combined with synchronous and asynchronous multi-user and multi-location settings. Furthermore, our analysis is an important first step in understanding which XR-specific properties can be leveraged to enable or chained together to amplify dark patterns. We consider our respective contributions as a solid basis for future work that could counteract the emergence of dark patterns in XR. Further, our scenarios spanned a range of probabilities and severities, evidencing notable harms in the process including financial loss, social and mental harm, and physical harm.

8.2 Old Manipulations in New Bottles: XR as a Facilitator and Amplifier of Existing Dark Patterns

As we could observe regarding RQ1, the majority of the co-designed manipulations constituted adaptations of known dark patterns that existing works have previously identified. For example, common patterns around the manipulative mechanisms *imbalancing options* or *baiting*, that typically exploit or manipulate the presentation of the underlying UI to effect behavior change, are equally applicable as such interfaces move from physical displays toward virtual spatial presentations [33].

Notably however, XR offers the potential to *amplify* the reach, scope, and efficacy of existing dark patterns in making use of spatiality, perception, XR device sensing, and physical/virtual barriers. For example, *persuasion* and *personalization* (see Table 3) both benefit from XR's ability to understand the user's actions (e.g., based on their behavior) and context to ultimately render realistic and highly personalized spatial virtual elements anywhere around the user without the separating features of a visible and tangible display. The co-designed scenarios described similar effects when known dark patterns that exclusively rely on visual aspects in 2D screens were transferred into 3D space and arguably became more effective through the additional spatial factor. For example, dark patterns that utilized visual imbalance induced through increased or reduced text sizes, contrast, or positioning on a 2D page (e.g., interface interference [32]) can additionally leverage a user's physical reach as described in the emergent pattern type *spatial imbalance of options*. Consequently, as we pointed out in Section 5, this adaptability combined with chaining imposes new challenges for ontologies and taxonomies required to foster transparent and interdisciplinary exchange. We will further discuss those aspects in Section 9.2.

8.3 "Puppy Crushing": Emergent Dark Patterns and the Potential to Exploit Affordances of XR

We also found evidence for 10 emergent patterns that XR can support (see RQ1). While there are some debatable overlaps with known dark patterns, in our view some of these patterns were sufficiently unique to XR to constitute separate consideration, in particular leveraging perceptual realism, attention tracking, and context-awareness to create novel archetypes (see RQ2). One example that perhaps best exemplified XR's unique capacity here was presented in the scenario *Puppy Crushing*, a moral dilemma that utilized photorealism, context-awareness, multimodality, and the intrinsic meaning of gestures to appeal to the user's paternal instinct and morality to manipulate decision making—having the user crush the virtual puppy to decline an offer. From the perspective of the manipulator, at best the user would refuse the action and implicitly accept the offer made; at worst, they would step on the puppy, associating visceral sensations of anguish regarding their actions with declining the offer, potentially influencing future behavior in the process.

8.4 XR Unlocks Common Dark Patterns in Physical, Psychological, and Virtual Realities

As we also pointed out in Section 4.3, it is important to align our understanding of *reality* in the context of anticipating and discussing dark patterns for XR. As our dataset showed, there are dark patterns that could apply to both *physical* reality (i.e., what *is* true), *psychological reality* (i.e., what we *believe* is true) as well as the *underlying* or *prevailing reality* [68] of an application (i.e., what is true *in the application's context*). Consequently, dark patterns could occlude each or only one of the potential realities to create a discrepancy between users, bystanders, and their individual reality. Given the vast background of work hinting toward VR having less stringent requirements regarding realism, we hypothesize that such barriers might be easier created and disguised in VR applications. This also means that differentiating between dark patterns affecting physical reality, psychological reality, the underlying or prevailing reality, or all of them might be necessary when it comes to risk assessment and regulations.

9 Implications and Future Work

A post-workshop critique raised by one of our participants addresses the perspective HCI often takes regarding dark patterns, as they are not something that accidentally come into existence but are rather motivated and pushed by companies to increase their financial benefit. As such, the root cause, namely users and application providers having different interests in technology, would not be solved by simply creating guidelines or embedding education around ethical aspects of technology design—at least not as a standalone solution. We, therefore, want to raise the question as to whether educating designers and users through publishing guidelines, as well as regulating already adapted technology and applications, is enough to mitigate or prevent the manifestation of dark patterns—or if we as a research community need to take different actions.

The following implications address concerns we see in this regard and offer potential solutions for our research community that is actively engaging in innovating XR hardware and software, as well as investigating aspects of DD and ethical creation and use of (XR) technology.

9.1 Toward Safe and Ethical Disclosure of XR Dark Patterns

We posit that the XR, HCI and DD communities need to urgently consider how we ethically investigate, evaluate, and report on the emergent vulnerabilities and potentially abusive uses of XR technology [31, 34]. We see particular benefit in examining how ethical principles in vulnerability disclosure could be applied to advances in XR. The security community has utilized a breadth of approaches here, ranging from concepts around responsible disclosure (i.e., informing the developers to give them a chance to fix the issue before publishing the vulnerability) toward full disclosure (i.e., announcing the vulnerability publicly right away). By its very nature, HCI research tends toward full disclosure, and this article is no exception. It is our view that as a research community we need to anticipate the misuse and abuse of XR devices if we are to successfully understand, detect, and mitigate said abuse. However, the full disclosure approach poses risks—that in conceptualizing these misuses we accelerate their eventual creation and real-world deployment, leaving activists with tools like "naming and shaming" [7] to contain the damage afterward.

Consequently, identifying and formalizing the principles around ethical, safe disclosure of XR vulnerabilities is of significant importance to this nascent research field. We believe this is of particular importance for works that knowingly or inadvertently expand our potential DD toolset (e.g., introducing novel perceptual manipulations) without considering for how such a capability might be misused, what ethical risks are posed, and so on. One recommendation that should be considered, previously put forward by Tseng et al. [73] is for the inclusion of standardized/tagged disclosure statements regarding the risks posed, such that as a community we can begin to build a knowledge base of potential vulnerabilities and accelerate collaborative efforts toward countermeasures. One potential format is to require authors of article submissions to include a statement on the potential negative societal impacts of the submitted work, as successfully applied in the NeurIPS community.[4] While this does pose the risk of accelerating the exploitation of said vulnerabilities, we see it inevitable that malicious actors will seek to leverage DD given the anticipated adoption of everyday AR- and VR-based metaverses, and our community has the potential to head-off significant manipulative designs before they see the light of day (e.g., [57]).

9.2 Toward Technology-Agnostic DD Classifications

The existence of XR dark patterns also provokes significant new research challenges around their further monitoring, reporting, identification, and formalization [32]—especially in regard to emerging technologies like XR and artificial intelligence. Building upon our preliminary investigation, our research aims to provide first insights into detecting the presence or use of XR dark patterns, preventing their use and designing mitigations that diminish their effectiveness or utility. Addressing such challenges will require a multi-disciplinary effort across the XR, HCI, and DD communities, to better understand who would seek to exploit XR dark patterns, and how—from individuals to

[4]NeurIPS Ethics Guidelines require authors of submissions to submit a statement on "Potential Negative Societal Impacts" (available at https://nips.cc/public/EthicsGuidelines, accessed 15 September 2022).

institutions to state actors. Consequently, future research should foster interdisciplinary efforts to address how dark patterns could manifest in any emerging technology and how they can be exploited to manipulate the behavior of individuals, groups, communities, and societies—from influencing consumer decisions to employer–employee manipulations. Current efforts in unifying existing dark patterns taxonomies and ontologies, for example, as proposed by Mathur et al. [55] or Gray et al. [32] are essential but arguably too narrow as they do not consider XR specific properties like spatiality. As our insights revealed, it might be a valuable addition to extend such works to incorporate amplifiers like the chaining of multiple manipulations, and personalization regarding physical and mental user characteristics. Further, building on works from Greenberg et al. [33], Eghtebas et al. [25], and revisiting the scenarios created during our workshops, we foresee a required shift in perspectives when researching dark patterns toward ubiquitous experiences, in which the UI becomes part of the real (or virtual) environment and escapes the bounds of existing digital display, finally resulting in the user leaving the spectator role and becoming part of the UI. In such experiences, bodily harm as a direct consequence of manipulations induced by dark patterns becomes possible [73] and needs to be considered in classificatory and regulatory frameworks.

Further, technology-agnostic regulations building on such frameworks might require to be based on a standardized ranking system, e.g., by utilizing probability-severity rankings as applied in our work. However, such rankings are likely to diverge across technological domains as several aspects of our participants' ranking were tied to respective technology readiness levels (see Section 6 and RQ3). This hints toward a required periodical reevaluation of dark patterns regarding their risk levels and an adaption of required regulations as they and their application context might change over time.

9.3 Toward Responsible Practices

Finally, we support and add to the arguments voiced by, e.g., Gray et al. [31], Narayanan et al. [59], and Nelissen and Funk [60] that designers and developers need to more strongly consider ethical approaches to application design. As practice-based research into XR design indicates [3, 39, 40, 41], design ethics are potentially ill-considered and lack visibility in the design process. This leads us to the hypothesis that ethical anticipation and evaluation of XR applications are not sufficiently considered within today's XR design practices in education, research, and industry. Therefore, we see the creation, validation, and evolution of *designerly*, meaningful tools and methods for responsible and critical XR design as a mandatory future work [42]. With respect to RQ3 in which we investigated aspects that should be considered when assessing the risk of XR dark patterns, our participants provided first insights in how to approach such an assessment. Those insights could be used as a basis to create tools and policies for a dark pattern risk evaluation. However, as our approach was speculative and conducted in a research setting, those aspects need to be thoroughly evaluated in a larger scale in practice and based on implemented dark patterns in XR or any other technology.

In addition, there is also a responsibility to then engage with other non-technical mechanisms for protecting users. In particular, proposing voluntary guidelines that XR platforms could adopt regarding the disclosure and use of XR dark patterns could inform app store policies and prevent the majority of XR dark patterns from reaching consumer devices. In time, we might expect more mandatory measures to take form, for example working toward legislation that can limit what XR dark patterns can and cannot do, and their (non-)consensual use. We also see addressing XR dark patterns as a key component in ensuring mental privacy and free will, key *perceptual rights* [61] that should be preserved as society transitions toward acceptance and adoption of XR. As a starting point, our experts' assessment approaches detailed in Section 6 might serve as initial input for creating such designerly tools for XR and other technologies. However, it is also important to include already existing courses of action based on actual practice in both industry and corresponding design communities, for example [7, 28, 58].

Lastly, as DD are likely to be implemented based on company policy and business models, our goal should not only be to provide pragmatic ethical guidance to XR creators and platforms—we as a research community should also investigate ways to establish and support designers' ethical agency early in the creation process. This includes exploring additional levers in industrial contexts to safeguard against DD, such as providing creators with mechanisms of agency and action to recognize respective company-induced business goals, and supporting whistleblowers in being able to point out, and act against, malicious practices when encountered.

10 Limitations

We aimed for broad topic coverage when recruiting participants and for a representative sample of seed articles. However, we were limited by who replied to our call for participation and the venues we used as recruitment platforms. Focusing on the most cited publications might have resulted in basing the developed scenarios on a particular perspective on XR. Further, contacting the first authors could result in an over-representation of junior researchers due to the widely adopted practice of listing them in the first positions.

Along the same line, some of our participants were either familiar with, were previously affiliated with, or had collaborated with some of the authors on projects included in our dataset. We managed those conflicts by reassigning the role of the workshop moderator among the authors when possible. However, as participants mainly interacted with each other rather than the workshop moderator, we argue this bias is negligible.

While our scenarios uncovered diverse potential future XR dark patterns and respective application domains, some potential and known areas were underrepresented. Future studies should broaden their focus to include aspects like fake news, harassment, and exclusion to complement our initial insights.

Finally, even though we attempted to balance DD experts from academia with DD experts from industry, only a few DD experts with an industry background responded positively. However, we want to emphasize that including DD experts from the industry with proficiency in dark patterns research is necessary to broaden the perspective on potential risks paired with realistic scenarios. Along the same line, future studies could extend their recruitment efforts to also invite XR experts from the industry to participate.

11 Conclusion

As the XR adoption rate by the masses increases, so does the imminent appearance of XR dark patterns. However, we lack insights in how purposefully DD might manifest in XR and how it will affect users. In our work, we approached XR dark patterns through a series of 10 speculative design workshops with a pair of experts from XR and DD academia and practices. In total, they created and discussed 42 dark pattern scenarios, which described 39 existing and adapted dark patterns as well as 10 potentially emergent dark patterns. We further described eight manipulative strategies our participants implemented to leverage four XR-specific properties (perception, spatiality, physical/virtual barriers, and XR device sensors). Finally, the risk assessment of the dark pattern scenarios regarding probability and severity revealed both the applicability of XR dark patterns as well as the necessity of interdisciplinary approaches to regulate and mitigate XR dark patterns in the near future. We therefore urge XR, HCI, and DD communities to consider how to further investigate, evaluate and report on the abusive potential of XR.

Appendix A: List of Dark Pattern Scenarios

Table 6. Short Descriptions of the 42 Dark Pattern Scenarios Created by Our Participants

ID	Scenario title	Description
WS01_1	Switch-out and haze	In a game where a VR HMD-wearing user is interacting with a non-HMD-wearing user, manipulative bystanders swap the non-HMD-wearing user with someone else to make fun of the HMD-wearing user.
WS01_2	Sneak and purchase	In a VR game, the application records users' weaknesses and uses this information in two ways to increase sales: (1) selling the information to other users, for competitive advantage and (2) adapting the game to users' weaknesses, requiring purchase to progress.
WS01_3	Physical harm	In a co-located VR game, the game leverages perceptual manipulation and spatial interaction to encourage users to physically harm each other (e.g., by encouraging users to apply force and speed in their gestural interactions). This benefits businesses in the medical domain (e.g., fixing broken teeth).
WS01_4	Shaming + gatekeeping	In a VR game, a user agrees to share their failures (e.g., videos of their worst moments in the game) on social media to advance in the game. The company uses this content to advertise the game and to encourage out-of-game engagement.
WS02_1	Displaying things to buy in close proximity	An AR application for cyclists displays discounts and offers of nearby shops in addition to traffic-safety-related information to increase the shops' sales.
WS02_2	Location give-away	An AR application for cyclists collects and interprets visual data about their environment to locate them despite the locating services being turned off to complete user profiles, reroute them wrt. nearby shops, or further interpreting the data without user consent.
WS02_3	Unnecessary pop-ups	In an AR application for cyclists, the application sends unrelated pop-ups to interrupt their current activity trying to get them to take a retour increase cognitive load to, for example increase sales for customers of this service.
WS02_4	Textual / verbal misconceptions	In an AR app for cyclists' traffic safety, a voice assistant or conversational aspect questions a cyclists' decisions in real time to make them think about it.
WS03_1	Texture manipulation/false textures	In a VR e-commerce application, textures that are more desirable (potentially by that specific user) instead of the "real" feel of the material are rendered, aiming to manipulate users into buying that product.

(Continued)

Table 6. Continued

ID	Scenario title	Description
WS03_2	False weight/adaptive weight	In a VR e-commerce application, the company adapts the physical weight of virtually rendered fabrics based on user preferences to increase sales.
WS03_3	Attention redirection/distracting from flaws	In a VR e-commerce application, the application utilizes (physical) properties of the interface to distract a user or redirect their attention (e.g., toward unseen areas of the virtual boutique or away from some flaws in the object's design) to increase sales.
WS03_4	Best version of the clothes on the user	In a VR e-commerce application, the application displays virtual renders of clothing fitted perfectly to a user's body, to increase sales; while in reality there is no capacity to adapt the product to the user body shape.
WS04_1	XR-vision ads	In an AR eye health-related training application, the application covertly moves relevant content away and brings other irrelevant content (e.g., ads) to the front. The ads could also use particular "tricks" to be more visually appealing/easier for the user to focus on.
WS04_2	Up-sell	In an AR eye health-related training application, the application uses time pressure (fake or less time than is reasonable) to up-sell additional eye exercises.
WS04_3	Fake friends	In an AR eye health-related training application, the application displays fake "eye exercise profiles" of their (potentially faked) friends to spark rivalry and increase the time spent on the application.
WS04_4	Privacy violation	In an AR eye health-related training application, to increase users' engagement with the app, new "spy" scenarios prompt users to look into spaces where other users are engaged in eye treatment exercises.
WS05_1	Wrong notification	In a VR application with haptic facial feedback, the application uses unprompted or unexpected notifications to redirect a user's attention to ads.
WS05_2	Direct attention to in-experience ads	In a VR application with haptic facial feedback, the application uses strong haptic or tactile feedback (force, heat) to redirect a user's attention to something unrelated (e.g., ads).
WS05_3	Immersive add	In a VR application with haptic facial feedback, immersive and "fun to experience" ads use proximity to a user's position/face, positioning in the user's FOV, anchoring to the user's body, and haptic facial feedback to require users to actively dodge the experience to make them difficult to avoid.

(Continued)

Table 6. Continued

ID	Scenario title	Description
WS05_4	Preference towards one choice or another (for example simulation games)	In a VR simulation game with haptic facial feedback, the application utilizes directional facial force feedback to direct users to wanted choices.
WS05_5	Pay more to experience more	In an VR application with haptic facial feedback, the application developers limit the full force feedback potential of the VR headset after a free trial period to trick users to purchase said feedback ability (e.g., in a subscription model) as additional feature to increase sales or the time a user spends on the application (in case time is used as currency).
WS05_6	Discourage user action in immersive experiences	In an immersive VR experience with haptic facial feedback, application developers punish unwanted decisions a user takes by providing strong haptic feedback or causing simulator sickness to steer users toward wanted actions.
WS06_1	Selling a cat in a bag	In a VR e-commerce application, the application renders tactile and visual appearance of material in a way that users more likely chose a more expensive option to increase sales.
WS06_2	Secret price increase	In a VR e-commerce application, the application does not display prices, other sourcing information (e.g., origin), or provide respective filter mechanisms but preferably presents expensive choices to the user to increase sales.
WS06_3	Shaming through novelty	In a VR e-commerce application, the company charges an extra VR fee to increase sales.
WS06_4	Sneaky product adds	In an VR e-commerce application, the application sneaks unwanted products into a user's shopping cart based on their currently explored items (e.g., specific cleanser) by exploiting the complexity of navigating a 3D space to increase sales.
WS07_1	Not a freeware	In a VR application, the system offers several functions without clearly stating which are free to increase accidental purchases by undiscerning users (e.g., children).
WS07_2	AR data harvesting	A VR application requires users to sign up and secretly records and shares data regarding their surroundings (including other people) and behavior which is then analyzed using machine learning to built a more robust user profile.
WS07_3	Expose people	A company publishes out-of-context user poses and movements recorded while using a VR sketching application to amuse bystanders.
WS07_4	Homework surveillance	A VR application is used by teachers/a university to monitor interactions and engagement of students while doing their homework using the application's data to grade their performance.

(Continued)

Table 6. Continued

ID	Scenario title	Description
WS08_1	Shared misinformation/manipulating inconsistent info	In a collaborative AR game, the application intentionally hides or displays information about one player to the other to either spark rivalry or put them under pressure to purchase features or items for/against each other.
WS08_2	Clickjacking (sort of)	In a collaborative AR game paring an HMD-wearing user with a projection-interacting player, the projection-interacting player can trigger UI elements of the HMD-wearing user without seeing the respective elements, thus accidentally purchasing items or activating ads.
WS08_3	Forced shouldersurf	In an AR application with a projection-based element, confidential information is projected to a known bystander and can be accessed by uninvolved third parties.
WS08_4	Prolong game-play	In a multi-user AR game (shared physical space), the application displays information about the about-to-win user to boost the about-to-lose player without both players realizing the unfair mechanics to intentionally prolong the time spent in the game.
WS09_1	Location manipulation	In an AR application operated via foot, the application places options or buttons which are beneficial for the company in a way that they are easier to access or activate (single-tap vs. multi-tap) to make users select respective options.
WS09_2	Button movement	In an AR-application operated via foot, the application manipulates button positioning or properties while the user is already aiming to activate an option to cause accidental triggers of costly options.
WS09_3	Make people look somewhere	In an AR application operated via foot, the application stops users via notification and strategically places buttons or content to make them change their direction or steer their visual attention toward physical objects or physical-world information.
WS09_4	Puppy Crushing	In an AR application operated via foot, the application covers options which are adverse to a company's benefit with a photo realistic puppy who needs to be forcefully kicked out of the way or crushed (with visual/aural feedback) for the user to decline an option.
WS10_1	Data misuse	In a collaborative AR learning application, a company uses the application to scan and semantically understand a user's surroundings by also prompting the user to move their visual attention; based on the data, the teacher recommends specific products to increase sales.

(Continued)

Table 6. Continued

ID	Scenario title	Description
WS10_2	Beneficiary: teacher	In a collaborative AR learning application which is used paired with a physical product, a teacher uses information about the learner's physical surroundings to recommend the purchase of additional gear or items or make the learner to change physical properties in their spaces.
WS10_3	Spatial ads	In a collaborative AR learning platform, the system learns user gaze behavior and interest to strategically position ads in between the task instructions.
WS10_4	Beneficiary: learner	In a collaborative AR learning application, a malicious party impersonates a learner to make the instructor to reveal potentially critical details about their knowledge and ways of teaching.

Acknowledgments

We thank our participants for their valuable time and for sharing both their work as well as their insights in our workshops and also thank Maria Gabriela López Garro for her visual support.

References

[1] Melvin Abraham, Pejman Saeghe, Mark McGill, and Mohamed Khamis. 2022. Implications of XR on Privacy, Security and Behaviour: Insights from Experts. In *Proceedings of the Nordic Human-Computer Interaction Conference (NordiCHI '22)*. ACM, New York, NY, 12 pages. DOI: https://doi.org/10.1145/3546155.3546691

[2] Richard Allen. 1995. *Projecting Illusion: Film Spectatorship and the Impression of Reality*. Cambridge University Press, Cambridge, United Kingdom.

[3] Narges Ashtari, Andrea Bunt, Joanna McGrenere, Michael Nebeling, and Parmit K. Chilana. 2020. Creating Augmented and Virtual Reality Applications: Current Practices, Challenges, and Opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, 1–13. DOI: https://doi.org/10.1145/3313831.3376722

[4] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, 1968–1979. DOI: https://doi.org/10.1145/2858036.2858226

[5] Stefano Baldassi, Tadayoshi Kohno, Franziska Roesner, and Moqian Tian. 2018. Challenges and New Directions in Augmented Reality, Computer Security, and Neuroscience - Part 1: Risks to Sensation and Perception. arXiv:1806.10557. Retrieved from http://arxiv.org/abs/1806.10557

[6] Dorothée Bentz, Nan Wang, Merle K. Ibach, Nathalie S. Schicktanz, Anja Zimmer, Andreas Papassotiropoulos, and Dominique J. F. de Quervain. 2021. Effectiveness of a Stand-Alone, Smartphone-Based Virtual Reality Exposure App to Reduce Fear of Heights in Real-Life: A Randomized Trial. *NPJ Digital Medicine* 4, 16 (2021), 9 pages. DOI: https://doi.org/10.1038/s41746-021-00387-7

[7] Harry Brignull. 2022. Deceptive Design - User Interfaces Crafted to Trick You. Retrieved June 29, 2022 from https://www.deceptive.design/

[8] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "I Am Definitely Manipulated, Even When I Am Aware of It. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In *Proceedings of the Designing Interactive Systems Conference 2021 (DIS '21)*. ACM, New York, NY, 763–776. DOI: https://doi.org/10.1145/3461778.3462086

[9] Elise Bonnail, Wen-Jie Tseng, Mark Mcgill, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2023. Memory Manipulations in Extended Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, Article 875, 20 pages. DOI: https://doi.org/10.1145/3544548.3580988

[10] 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.

[11] Harry Brignull. 2011. *Dark Patterns: Deception vs. Honesty in UI Design*. Interaction Design, Usability. Retrieved from https://alistapart.com/article/dark-patterns-deception-vs-honesty-in-ui-design/

[12] Lauren Buck and Rachel McDonnell. 2022. Security and Privacy in the Metaverse: The Threat of the Digital Human. In *Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality*. ACM, New Orleans, LA, 4 pages.

[13] Kent Bye, Diane Hosfelt, Sam Chase, Matt Miesnieks, and Taylor Beck. 2019. The Ethical and Privacy Implications of Mixed Reality. In *Proceedings of the ACM SIGGRAPH 2019 Panels*. ACM, New York, NY. DOI: https://doi.org/10.1145/3306212.3328138

[14] Peter Casey, Ibrahim Baggili, and Ananya Yarramreddy. 2021. Immersive Virtual Reality Attacks and the Human Joystick. *IEEE Transactions on Dependable and Secure Computing* 18, 2 (Mar. 2021), 550–562. DOI: https://doi.org/10.1109/TDSC.2019.2907942

[15] Régis Chatellier, Geoffrey Delcroix, Estelle Hary, Camille Girard-Chanudet, Faget Pauline Marie Leroux, and Stephanie Chapelle. 2019. Shaping Choices in the Digital World. From Dark Patterns to Data Protection: The Influence of UX/UI Design on User Empowerment. *IP Reports Innovation and Foresight* 6 (Jan. 2019), 50. Retrieved from https://www.cnil.fr/sites/cnil/files/2023-06/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf

[16] Kaiming Cheng, Jeffery F. Tian, Tadayoshi Kohno, and Franziska Roesner. 2023. Exploring User Reactions and Mental Models Towards Perceptual Manipulation Attacks in Mixed Reality. In *Proceedings of the 32nd USENIX Security Symposium*. USENIX Association, Anaheim, CA, USA, 18 pages.

[17] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D. Wilson. 2017. Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, 3718–3728. DOI: https://doi.org/10.1145/3025453.3025753

[18] Danielle Keats Citron and Daniel J. Solove. 2022. Privacy Harms. *B.U.L. Review* 102 (2022), 793.

[19] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. In *Qualitative Psychology*, J.A. Smith (Ed.), SAGE Publications, London, 222–248.

[20] Competition and Markets Authority. 2022a. Evidence Review of Online Choice Architecture and Consumer and Competition Harm. *Competition & Markets Authority (CMA)* 157 (2022), 261 pages. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1069423/OCA_Evidence_Review_Paper_14.4.22.pdf

[21] Competition and Markets Authority. 2022b. Online Choice Architecture: How Digital Design Can Harm Competition and Consumers. *Competition & Markets Authority (CMA)* 155 (2022), 76 pages. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1066524/Online_choice_architecture_discussion_paper.pdf

[22] Gregory Conti and Edward Sobiesk. 2010. Malicious Interface Design: Exploiting the User. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. ACM, New York, NY, USA, 271–280. DOI: https://doi.org/10.1145/1772690.1772719

[23] Rikke Friis Dam and Teo Yu Siang. 2020. *Learn How to Use the Best Ideation Methods: Brainstorming, Braindumping, Brainwriting, and Brainwalking*. Interaction Design Foundation. Retrieved from https://www.interaction-design.org/literature/article/learn-how-to-use-the-best-ideation-methods-brainstorming-braindumping-brainwriting-and-brainwalking

[24] Arindam Dey, Mark Billinghurst, Robert W. Lindeman, and J. Edward Swan. 2018. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI* 5 (2018), 37.

[25] Chloe Eghtebas, Gudrun Klinker, Susanne Boll, and Marion Koelle. 2023. Co-Speculating on Dark Scenarios and Unintended Consequences of a Ubiquitous (ly) Augmented Reality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. ACM, New York, NY, 2392–2407. DOI: https://doi.org/10.1145/3563657.3596073

[26] European Data Protection Board. 2022. *Guidelines 3/2022 on Dark Patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them*. European Data Protection Board. Retrieved from https://edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf

[27] 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). https://eur-lex.europa.eu/eli/reg/2022/2065#document1

[28] Trine Falbe, Tiziana D'Agostino, Maria Arango Kure, Jarijn Nijkamp, Raffaella Isidorim, and Lexi Namer. 2022. Ethical Design Network — Share, Inspire, Empower. Retrieved December 21, 2022 from https://ethicaldesignnetwork.com/.

[29] Forbrukerradet. 2018. *Deceived by Design. How Tech Companies Use Dark Patterns to Discourage Us from Exercising Our Rights to Privacy*. Consumer Council of Norway. Retrieved from https://storage02.forbrukerradet.no/media/2018/06/2018-06-27-deceived-by-design-final.pdf

[30] Colin M. Gray, Shruthi Sai Chivukula, and Ahreum Lee. 2020. What Kind of Work Do "Asshole Designers" Create? Describing Properties of Ethical Concern on Reddit. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. ACM, New York, NY, 61–73. DOI: https://doi.org/10.1145/3357236.3395486

[31] 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 (CHI '18)*. ACM, New York, NY, 1–14. DOI: https://doi.org/10.1145/3173574.3174108

[32] Colin M. Gray, Cristiana Santos, and Nataliia Bielova. 2023. Towards a Preliminary Ontology of Dark Patterns Knowledge. In *Proceedings of the Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. ACM, New York, NY, Article 286, 9 pages. DOI: https://doi.org/10.1145/3544549.3585676

[33] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. Dark Patterns in Proxemic Interactions: A Critical Perspective. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, 523–532. DOI: https://doi.org/10.1145/2598510.2598541

[34] Jan Gugenheimer, Wen-Jie Tseng, Abraham Hani Mhaidli, Jan Ole Rixen, Mark McGill, Michael Nebeling, Mohamed Khamis, Florian Schaub, and Sanchari Das. 2022. Novel Challenges of Safety, Security and Privacy in Extended Reality. In *Proceedings of the Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. ACM, New York, NY, Article 108, 5 pages. DOI: https://doi.org/10.1145/3491101.3503741

[35] Andreas Halbig and Marc Erich Latoschik. 2021. A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. *Frontiers in Virtual Reality* 2 (2021), 32 pages.

[36] Scott Hayden. 2020. *Mother Meets Recreation of Her Deceased Child in VR — Road to VR*. World to VR. Retrieved from https://www.roadtovr.com/mother-meets-recreation-of-deceased-child-in-vr/

[37] Caroline Hummels and Pieter Jan Stappers. 1998. Meaningful Gestures for Human Computer Interaction: Beyond Hand Postures. In *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, Nara, Japan, 591–596. DOI: https://doi.org/10.1109/AFGR.1998.671012

[38] Ross Johnstone, Neil McDonnell, and Julie R. Williamson. 2022. When Virtuality Surpasses Reality: Possible Futures of Ubiquitous XR. In *Proceedings of the Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. ACM, New York, NY, Article 6, 8 pages. DOI: https://doi.org/10.1145/3491101.3516396

[39] Veronika Krauß, Alexander Boden, Leif Oppermann, and René Reiners. 2021. Current Practices, Challenges, and Design Implications for Collaborative AR/VR Application Development. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, Article 454, 15 pages. DOI: https://doi.org/10.1145/3411764.3445335

[40] Veronika Krauß, Michael Nebeling, Florian Jasche, and Alexander Boden. 2022. Elements of XR Prototyping: Characterizing the Role and Use of Prototypes in Augmented and Virtual Reality Design. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, Article 310, 18 pages. DOI: https://doi.org/10.1145/3491102.3517714

[41] Veronika Krauß. 2022. Exploring Dark Patterns in XR. In *Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality (CHI EA'21)*. ACM, New York, NY, 1–2. Retrieved from https://wenjietseng.com/assets/pdf/SSPXR22_submissions/SSPXR22_paper_8.pdf

[42] Veronika Krauß, Jenny Berkholz, Lena Recki, and Alexander Boden. 2023. Beyond Well-Intentioned: An HCI Students' Ethical Assessment of Their Own XR Designs. In *Proceedings of the 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Sydney, Australia, 95–68. DOI: https://doi.org/10.1109/ISMAR59233.2023.00020

[43] Chiara Krisam, Heike Dietmann, Melanie Volkamer, and Oksana Kulyk. 2021. Dark Patterns in the Wild: Review of Cookie Disclaimer Designs on Top 500 German Websites. In *Proceedings of the 2021 European Symposium on Usable Security (EuroUSEC '21)*. ACM, New York, NY, 1–8. DOI: https://doi.org/10.1145/3481357.3481516

[44] Olya Kudina and Peter-Paul Verbeek. 2019. Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy. *Science, Technology, & Human Values* 44, 2 (2019), 291–314. DOI: https://doi.org/10.1177/0162243918793711

[45] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'Dark Pattern' in Home Robots. In *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Daegu, South Korea, 374–381.

[46] Kiron Lebeck, Tadayoshi Kohno, and Franziska Roesner. 2016. How to Safely Augment Reality: Challenges and Directions. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications (HotMobile '16)*. ACM, New York, NY, 45–50. DOI: https://doi.org/10.1145/2873587.2873595

[47] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. 2017. Securing Augmented Reality Output. In *Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Jose, CA, USA, 320–337.

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

[49] Kai Lukoff, Alexis Hiniker, Colin M. Gray, Arunesh Mathur, and Shruthi Sai Chivukula. 2021. What Can CHI Do About Dark Patterns? In *Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. ACM, New York, NY, Article 102, 6 pages. DOI: https://doi.org/10.1145/3411763.3441360

[50] Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. 2020. Towards the Identification of Dark Patterns: An Analysis Based on End-User Reactions. In *Proceedings of the 11th Indian Conference on Human-Computer Interaction (IndiaHCI '20)*. ACM, New York, NY, 24–33. DOI: https://doi.org/10.1145/3429290.3429293

[51] Jung Yeon Ma and Jong Soo Choi. 2006. Augmented Reality as Perceptual Reality. In *Interactive Technologies and Sociotechnical Systems*. Hongbin Zha, Zhigeng Pan, Hal Thwaites, Alonzo C. Addison, and Maurizio Forte (Eds.). Springer, Berlin, 7–10.

[52] Lara Maister, Mel Slater, Maria V Sanchez-Vives, and Manos Tsakiris. 2015. Changing Bodies Changes Minds: Owning Another Body Affects Social Cognition. *Trends in Cognitive Sciences* 19, 1 (2015), 6–12.

[53] Steve Mann. 2002. Mediated Reality with Implementations for Everyday Life. *Presence Connect* 1 (2002), 2002.

[54] 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. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32.

[55] 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 (CHI '21)*. ACM, New York, NY, Article 360, 18 pages. DOI: https://doi.org/10.1145/3411764.3445610

[56] Mark McGill. 2021. *The IEEE Global Initiative on Ethics of Extended Reality (XR) Report–Extended Reality (XR) and the Erosion of Anonymity and Privacy*. Technical Report. IEEE Global Initiative on Ethics of Extended Reality (XR), 1–24 pages.

[57] Abraham Hani Mhaidli and Florian Schaub. 2021. Identifying Manipulative Advertising Techniques in XR Through Scenario Construction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, Article 296, 18 pages. DOI: https://doi.org/10.1145/3411764.3445253

[58] Lexi Namer, Ehtical Design Network, and Kate Every. 2022. Ethical Design Resources. Retrieved December 21, 2022 from https://www.ethicaldesignresources.com/

[59] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. 2020. Dark Patterns: Past, Present, and Future: The Evolution of Tricky User Interfaces. *Queue* 18, 2 (2020), 67–92.

[60] L. G. M. Nelissen and Mathias Funk. 2022. Rationalizing Dark Patterns: Examining the Process of Designing Privacy UX Through Speculative Enactments. *International Journal of Design* 16, 1 (Apr. 2022), 75–92. DOI: https://doi.org/10.57698/v16i1.05

[61] NeuroRights Foundation. 2019. *The NeuroRights Foundation – New Human Rights for the Age of Neurotechnology*. The Neurorights Foundation. Retrieved from https://neurorightsfoundation.org/mission

[62] 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 *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, 1–13. DOI: https://doi.org/10.1145/3313831.3376321

[63] 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 (EuroUSEC '22)*. ACM, New York, NY, 64–78. DOI: https://doi.org/10.1145/3549015.3554213

[64] Sharif Razzaque, David Swapp, Mel Slater, Mary C. Whitton, and Anthony Steed. 2002. Redirected Walking in Place. In *Proceedings of the Workshop on Virtual Environments 2002 (EGVE '02)*. Eurographics Association, Goslar, DEU, 123–130.

[65] Michael Rietzler, Florian Geiselhart, Jan Gugenheimer, and Enrico Rukzio. 2018. Breaking the Tracking: Enabling Weight Perception Using Perceivable Tracking Offsets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, 1–12. DOI: https://doi.org/10.1145/3173574.3173702

[66] Franziska Roesner and Tadayoshi Kohno. 2021. Security and Privacy for Augmented Reality: Our 10-Year Retrospective. In *Proceedings of the VR4Sec: 1st International Workshop on Security for XR and XR for Security*. Online, 5 pages.

[67] Franziska Roesner, Tadayoshi Kohno, and David Molnar. 2014. Security and Privacy for Augmented Reality Systems. *Communications of the ACM* 57, 4 (2014), 88–96.

[68] Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, Dalila Szostak. 2020. The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality* 1 (2020), 512449.

[69] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. 2008. Analyses of Human Sensitivity to Redirected Walking. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology (VRST '08)*. ACM, New York, NY, 149–156. DOI: https://doi.org/10.1145/1450579.1450611

[70] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. 2010. Estimation of Detection Thresholds for Redirected Walking Techniques. *IEEE Transactions on Visualization and Computer Graphics* 16, 1 (Jan. 2010), 17–27. DOI: https://doi.org/10.1109/TVCG.2009.62

[71] Zihao Su, Faysal Hossain Shezan, Yuan Tian, David Evans, and Seongkook Heo. 2022. Perception Hacking for 2D Cursorjacking in Virtual Reality. In *Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality (CHI EA'21)*. ACM, New Orleans, LA, USA, 5 pages.

[72] Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan Mcguire, David Luebke, and Arie Kaufman. 2018. Towards Virtual Reality Infinite Walking: Dynamic Saccadic Redirection. *ACM Transactions on Graphics* 37, 4 (Jul. 2018), 67:1–67:13. DOI: https://doi.org/10.1145/3197517.3201294

[73] Wen-Jie Tseng, Elise Bonnail, Mark McGill, Mohamed Khamis, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2022. The Dark Side of Perceptual Manipulations in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, Article 612, 15 pages. DOI: https://doi.org/10.1145/3491102.3517728

[74] Mark Warner and Debra Fischer. 2019. *Senators Introduce Bipartisan Legislation to Ban Manipulative 'Dark Patterns'*. Debra Fisher. Retrieved from https://www.fischer.senate.gov/public/index.cfm/2019/4/senators-introduce-bipartisan-legislation-to-ban-manipulative-dark-patterns

[75] Graham Wilson and Mark McGill. 2018. Violent Video Games in Virtual Reality: Re-Evaluating the Impact and Rating of Interactive Experiences. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, New York, NY, 535–548. DOI: https://doi.org/10.1145/3242671.3242684

[76] Marty J. Wolf, Frances Grodzinsky, and Keith Miller. 2016. Augmented Reality All around Us: Power and Perception at a Crossroads. *SIGCAS Computers and Society* 45, 3 (Jan. 2016), 126–131. DOI: https://doi.org/10.1145/2874239.2874257

[77] Carlos Bermejo Fernandez Xian Wang, Lik-Hang Lee and Pan Hui. 2023. The dark side of augmented reality: exploring manipulative designs in AR. *International Journal of Human-Computer Interaction* (2023), 1–16. https://doi.org/10.1080/10447318.2023.2188799

[78] José P. Zagal, Staffan Björk, and Chris Lewis. 2013. Dark Patterns in the Design of Games. In *Proceedings of the Foundations of Digital Games 2013*. Society for the Advancement of the Science of Digital Games, Chania, Crete, Greece, 39–46.

Received 14 February 2023; revised 1 December 2023; accepted 11 March 2024