What Makes a Dark Pattern... Dark?

Design Attributes, Normative Considerations, and Measurement Methods

Arunesh Mathur
amathur@princeton.edu
Princeton University

Jonathan Mayer
jonathan.mayer@princeton.edu
Princeton University

Mihir Kshirsagar
mihir@princeton.edu
Princeton University

ABSTRACT

There is a rapidly growing literature on dark patterns, user interface designs—typically related to shopping or privacy—that researchers deem problematic. Recent work has been predominantly descriptive, documenting and categorizing objectionable user interfaces. These contributions have been invaluable in highlighting specific designs for researchers and policymakers. But the current literature lacks a conceptual foundation: What makes a user interface a dark pattern? *Why* are certain designs problematic for users or society?

We review recent work on dark patterns and demonstrate that the literature does not reflect a singular concern or consistent definition, but rather, a set of thematically related considerations. Drawing from scholarship in psychology, economics, ethics, philosophy, and law, we articulate a set of normative perspectives for analyzing dark patterns and their effects on individuals and society. We then show how future research on dark patterns can go beyond subjective criticism of user interface designs and apply empirical methods grounded in normative perspectives.

1 INTRODUCTION

Recent scholarship has called attention to dark patterns, user interface designs that researchers deem problematic. The preponderance of academic literature on dark patterns has curated collections of objectionable user interface designs [3, 21] and highlighted the frequency of dark patterns in specific contexts, such as privacy settings [4], online gaming [62], and online shopping [35]. Related work has also traced the lineage of dark patterns to discrete trends in business and social science [37].

Meanwhile, legislators and regulators are taking notice. In the U.S., Congress is considering legislation that would restrict dark patterns [58], and California [40] recently enacted legislation (via a ballot initiative) that defines and prohibits dark patterns associated

with privacy consent. In the EU, regulators are beginning to target dark patterns for enforcement actions [18, 42]. Dark patterns are also a persistent source of public frustration, such as on Twitter (#darkpattern and @darkpatterns) and a Reddit subreddit (r/assholedesign) with over 1.5 million members.

The current academic discourse about dark patterns is built on a shaky foundation, however. While the literature has rich descriptive contributions, calling attention to user interface designs that the authors (understandably) find objectionable, the contours and normative underpinnings of dark patterns have received less attention. What, exactly, makes a user interface design a dark pattern? And what makes a dark pattern normatively problematic?

This gap in the literature has significant consequences. Within the Human–Computer Interaction (HCI) community and related areas of computer science, publications are in limited dialogue, and there is a disconnect between research results and possible government responses. Meanwhile, outside of computer science, critics have begun to dismiss dark patterns research as academic handwringing about aggressive marketing and sharp dealing—the online equivalents of practices that are unsavory but commonplace for offline businesses [7, 27]. Absent a clearly articulated rationale or standard for taking action against certain user interface designs, legislators and regulators on both sides of the Atlantic will continue to struggle with possible market interventions [8].

This paper aims to provide analytical clarity and normative foundations for dark patterns research in HCI, other areas of computer science, and related disciplines. We begin with a review of the rapidly growing literature on dark patterns, identifying key definitions and taxonomic concepts. We demonstrate that dark patterns research is not grounded in a singular, unified concern or definition. Rather, dark patterns scholarship—similar to privacy scholarship—is motivated by a family of thematically related considerations. We illustrate this point by synthesizing dark patterns definitions, types, and taxonomies from recent scholarship into a pair of themes.

Next, we discuss how topics in related disciplines can enrich the dark patterns discourse in the HCI community, including sludges in behavioral economics, manipulation in philosophy and ethics, and market manipulation in law. We draw out the perspectives and normative considerations that these other disciplines offer, and we use these to develop a set of normative lenses for evaluating dark patterns.

Finally, we describe how future research can use HCI methods to examine dark patterns through each normative lens. We demonstrate this path forward for future work by applying our recommended approach in a case study. We conclude with a discussion of how the normative perspectives and measurement methods we develop in this paper can be of use to HCI practitioners and policymakers.

Our primary contributions are as follows.

- We review the dark patterns literature in HCI and related areas of computer science, identify diverse definitions and taxonomies, and synthesize cross-cutting themes (Section 2).
- We connect the themes in dark patterns scholarship to representative related work in other disciplines (Section 3).
- We describe a set of discrete normative perspectives, grounded in the representative related work, that could guide future dark patterns research (Section 4).
- We propose that future dark patterns research use methods that align with those normative perspectives, and we demonstrate our proposed approach with a case study (Section 5).

2 A COMPARISON OF DARK PATTERN DEFINITIONS, TYPES, AND ATTRIBUTES

We begin with a scoping review of the dark patterns academic literature, which is predominantly (but not exclusively) in the area of HCI.[1] To facilitate this survey, we compiled a dataset of papers by searching the ACM Digital Library, arXiv, and Google Scholar for work that referenced "dark patterns" or similar terms.[2] We then filtered the dataset, retaining papers that a) discuss dark patterns in the context of user interface design, and b) have been published in an academic venue.[3] The resulting dataset included a total of 20 papers, published in HCI conferences and journals (e.g., CHI, NordiCHI, and DIS), security and privacy venues (e.g., PETS and NSPW), law journals, and a psychology journal.[4]

Next, we augmented our dataset with legislation and regulatory materials related to dark patterns.[5] We ran a search for "dark pattern(s)" on WestLaw for federal and state materials in the U.S., and we identified international materials from references in publications. Our search yielded the Deceptive Experiences to Online Users Reduction (DETOUR) Act [58], the California Privacy Rights Act (CPRA, also known as Proposition 24 in the November 2020 election) [40], a report by France's data protection agency (CNIL) [42], and a report by the Norwegian Consumer Council (NCC/Frobrukerrådet) [18].

We then coded each document in the dataset for four distinct properties. First, we examined dark pattern definitions. We coded a document as introducing a definition if it included a definition, explicitly or implicitly, and the definition was either entirely original to the work or went beyond a verbatim reference to prior work.

---

[1]We refer to this analysis as a scoping review rather than a systematization of the literature, because our goal is to evaluate key definitions and concepts in publications rather than reported evidence.

[2]The complete set of search queries was: "anti-pattern(s)," "dark pattern(s)," "deceptive design pattern(s)," "FoMo design(s)," and "manipulative design pattern(s)." We selected these queries based on the terminology in dark patterns scholarship we had already identified, and we believe they captured a representative—if not fully complete—dataset of related work.

[3]We did not include papers that only incidentally discussed dark patterns. We included the initial unpublished Brignull work [3], a concurrent publication by Conti and Sobiesk [9] that predates the popularization of the "dark patterns" term, and a forthcoming paper by Luguri and Strahilevitz [32] because those contributions have generated significant discussion.

[4]The dataset is available at https://github.com/citp/dark-patterns. We compiled the dataset in December 2020.

[5]We included introduced legislation and official agency publications in our dataset. Our search additionally encompassed press releases, hearing materials, and statements from individual members of multi-member commissions; we reviewed this content to ensure we had not missed materials for the dataset.

If a document introduced a definition, we extracted the definition. The second property we coded was whether a document offered a typology or taxonomy of dark patterns. If it did, we extracted the categories of dark patterns, individual dark patterns, and associated descriptions. The third property that we coded was whether a document discussed any normative considerations, either explicitly or implicitly. If it did, we extracted the normative concepts. Finally, we coded each document's use of user studies. If a paper included a new user study, we extracted the goals of the study, the type of study, and any measurement metrics.

In the following subsections, we use our dataset to compare dark patterns definitions (Section 2.1), types (Section 2.2), and attributes (Section 2.3) in prior work. We trace the origins and evolution of various dark patterns definitions, types, and attributes. We show that the literature—which we emphasize is valuable—has a number of inconsistencies and contradictions. We argue that these are a result of the literature's attempt to capture dark patterns in one succinct definition or list of types. Instead, we derive a pair of themes related to choice architecture that succinctly reflect the dark patterns discourse (Section 2.4): modifying the user's choice set and manipulating the information flow to the user.

2.1 Defining Dark Patterns

In 2010, when Brignull first introduced the term "dark patterns" on the website darkpatterns.org [3], he described the user interface designs as "tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something." Brignull's initial work set a flurry of academic research into motion that attempted to define and describe dark patterns. We identified a total of 19 such definitions in our dataset, and from these definitions, we inductively surfaced four facets of dark pattern definitions.

The first facet describes characteristics of the user interface that can affect users. Some definitions describe dark pattern user interfaces as design "tricks" (Brignull [3] and Waldman [57]), while others describe them as "misleading" interfaces (Bösch et al. [4] and CNIL [42]). Some documents define dark patterns with a set of alternative characteristics. For instance, Mathur et al. [35] state that dark pattern designs are "coercing, steering, or deceiving," and Gray et al. [20] note that dark patterns have "obnoxious, coercive, or deceitful" properties.

The second facet of definitions is the mechanism of effect for influencing users.[6] Collectively, the dark pattern definitions specify 13 distinct ways that user interfaces could influence users. Some definitions describe dark patterns as subverting user intent (Brignull [3], Lewis [31], Bösch et al. [4], Westin and Chiasson [59], Waldman [57], and Day and Stemler [11]) or subverting user preferences (Bösch et al. [4], Mathur et al. [35], Luguri and Strahilevitz [32], CNIL [42], DETOUR Act [58], and CPRA [40]). Some documents state that dark patterns "trick" users (Bösch et al. [4], Westin and Chiasson [59], and Maier and Harr [34]), while others define dark patterns as undermining user autonomy (DETOUR Act [58] and CPRA [40]). Many definitions specify multiple mechanisms of effect on users. For instance, Westin and Chiasson [59] define dark patterns to

---

[6]We recognize that the distinction between characteristics of the user interface and mechanisms of effect on users may, in some instances, be more grammatical than substantive. We parse this distinction because some readers may find it meaningful.

Table 1: A classification of various "dark pattern" definitions in academic literature, law, and policy. Documents are ordered by date.

|		Academic Publications															Government Materials			
		Brignull [3]	Conti & Sobiesk [9]	Zagal et al. [62]	Lewis [31]	Bösch et al. [4]	Gray et al. [21]	Mathur et al. [35]	Luguri & Strahilevitz [32]	Lacey & Caudwell [29]	Utz et al. [56]	Westin & Chiasson [59]	Waldman [57]	Day & Stemler [11]	Gray et al. [20]	Maier & Harr [34]	NCC [18]	CNIL [42]	DETOUR Act [58]	CPRA [40]
Characteristics of the User Interface	Coercive						◐								◐					
	Deceptive					●	◐								◐					
	Malicious		●																	
	Misleading					●										●				
	Obnoxious														◐					
	Seductive															●				
	Steering							◐									●			
	Trickery	●												●						
Mechanisms of Effect on Users	Attack users		◐																	
	Confuse users								◐											
	Deceive users							●												
	Exploit users		◐																	
	Manipulate users		◐						◐			●	●							
	Mislead users															●				
	Steer users													●						
	Subvert user intent	●			●	●						●	●	●						
	Subvert user preferences						●	●	◐									●	●	●
	Trick users					●						●		●						
	Undermine user autonomy																		●	●
	Without user consent			●																
	Without user knowledge																	●		
Role of User Interface Designers	Abuse of designer knowledge						●			●		●				●				
	Designer intent		●	●			●		●						●			●	●	●
Benefits and Harms	Benefit to service		●					●			●				●					
	Harm to users			●			●						●				●			

● Required element of "dark pattern" definition ◐ Alternative element of "dark pattern" definition

"manipulate" users, subvert user intent, and "trick" users, whereas Conti and Sobiesk [9] state that dark patterns "attack," "exploit," or "manipulate" users.

The third facet of dark pattern definitions is the role of the user interface designer. Some definitions involve a designer abusing their domain-specific knowledge of human behavior (Gray et al. [21], Lacey and Caudwell [29], Westin and Chiasson [59], and Maier and Harr [34]). Other definitions state that designers intentionally deploy dark patterns to achieve a goal (Conti and Sobiesk [9], Zagal et al. [62], Gray et al. [21], Luguri and Strahilevitz [32], Gray et al. [20], CNIL [42], DETOUR Act [58], and CPRA [40]). We note that, as far back as 1998, Fogg articulated a very similar concept: user interface designers could use "persuasive technologies" to intentionally influence users [15].

The fourth facet of dark pattern definitions is the benefits and harms resulting from a user interface design. Some definitions describe a dark pattern as aiming to benefit an online service (Conti and Sobiesk [9], Mathur et al. [35], Utz et al. [56], and Gray et al. [20]). Other dark pattern definitions involve harm to users (Zagal et al. [62], Gray et al. [21], Waldman [57], and the NCC [18]).

We summarize the 19 dark pattern definitions along these four facets in Table 1. As the table shows, there is significant variation among the facets reflected in definitions. For instance, nine definitions do not involve any characteristic of the user interface, four definitions do not specify a mechanism of effect on users, eight definitions do not address the role of user interface designers, and ten definitions do not involve benefit or harm elements.
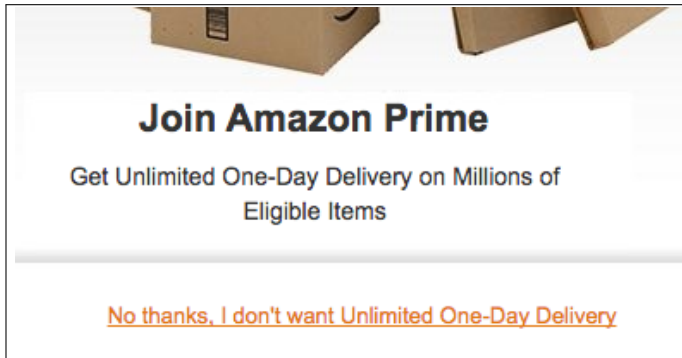
Figure 1: The Confirmshaming dark pattern on Amazon. The option to decline the offer uses framing to guilt the user into opting in.

There is also significant variation within each facet of the definitions. For example, Brignull [3] and Conti and Sobiesk [9] both have definitions that involve characteristics of user interfaces, but the former describes dark patterns as "tricks" while the latter notes that dark patterns are "malicious."

Another point of divergence is that, even when definitions share an element, the element may be required for one definition but not another definition. Both Gray et al. [21] and Mathur et al. [35], for example, take the position that deception is an element of a dark pattern. But deception is a required element for the former, while one of several alternative elements for the latter. Similarly, Bösch et al. [4] describes dark patterns as subverting user preferences, as does Luguri and Strahilevitz [32]. But the latter work additionally considers user interfaces that confuse or manipulate users to be dark patterns.

A final challenge for dark patterns definitions is lack of specificity in recurring terminology. Brignull [3] and Waldman [57], for example, describe dark patterns as involving "tricks." But what constitutes a trick? Similarly, Maier and Harr [34] note that dark patterns are "seductive" user interfaces. What makes a user interface seductive?

2.2 Types of Dark Patterns

Prior academic work on dark patterns has often gone beyond a definition and provided a taxonomy of user interface types, with a rich description of each dark pattern. Brignull's [3] initial taxonomy introduced 12 types of dark patterns. Bait-and-Switch, for example, is a user interface where "the user sets out to do one thing, but a different, undesirable thing happens instead." Confirmshaming is "the act of guilting the user into opting in to something." Figure 1 illustrates one instance of the Confirmshaming dark pattern. In concurrent work, Conti and Sobiesk [9] documented 11 categories of problematic user interfaces. Obfuscation, for instance, involves "hiding desired information and interface elements." Trick involves "misleading the user or other attempts at deception, such as spoofed content or interface elements."

Several subsequent projects have focused on dark patterns in specific categories of applications or systems. Zagal et al. [62] described seven types of dark patterns used by video game developers.

The authors introduced dark patterns such as Pay to Skip, where a developer charges users to solve in-game challenges and creates an uneven playing field for paying and non-paying users. Grinding requires users to perform repetitive and tedious tasks in a game, which can keep users invested in playing. Lewis [31] focused on dark patterns in mobile apps. Temporal dark patterns in apps "request too much time, provide too little entertainment time, or result in users feeling like they've wasted their time." Monetary dark patterns cause "users to regret spending money, lose track of how much money they will spend, or [not] know how much money will be required in order to progress." Social Capital dark patterns cause "users [to] stand the risk of losing their social standing, or feel like they're interacting with software out of social obligation to fulfill a motivation." Greenberg et al. [22] considered systems that account for physical space and described a taxonomy of eight "proxemic" (i.e., physical proximity) interaction dark patterns. The Attention Grabber, for example, is where a user "happens to pass by the field of view of a strategically located system, and the system takes deliberate action to attract and keep that person's attention." The Milk Factor is where a system "forces the user to move through or go to a specific location in order to get a service." Finally, Lacey and Caudwell [29] focused on home robotics systems and whether "cuteness" can constitute a dark pattern.

Another set of studies has documented dark patterns that undermine user privacy. Bösch et al. [4] introduced a set of seven privacy dark patterns, including Bad Defaults, where "the default options are sometimes chosen badly in the sense that they ease or encourage the sharing of personal information." Hidden Legalese Stipulations is when "services hide stipulations in the policies which target the privacy of the user." The NCC report [18] on dark patterns described a set of five privacy dark patterns. Ease, for example, is "making the process toward [privacy-respecting] alternatives a long and arduous process." Framing is when a user interface "focus[es] on the positive aspects of one choice, while glossing over any potentially negative aspects, inclin[ing] many users to comply with the service provider's wishes." The CNIL report [42] emphasized how dark patterns can undermine data protection rights, documenting 18 distinct types of dark patterns. Impenetrable Wall, for instance, is a user interface that "blocks access to a service [with] a cookie wall or account creation [when that] is not necessary to use the service." Last Minute Consent involves "seeking consent for the collection of data at a specific moment where we know that the individual is in a weak position because [they are] in a hurry or impatient to finish."

Finally, Gray et al. [21] created a taxonomy of dark patterns from the ground up by collecting a corpus of dark patterns submitted by users on Twitter. The authors revamped Brignull's original taxonomy of dark patterns and introduced new dark patterns. Nagging, for instance, is when "the user's desired task is interrupted one or more times by other tasks not directly related to the one the user is focusing on." Forced Action is when "users are required to perform a specific action to access (or continue to access) specific functionality." In another study, Gray et al. [20] created a similar taxonomy by collecting and analyzing a corpus of user interface designs submitted by users on Reddit. The authors introduced Misrepresenting, for example, in which "designers provide ambiguous

and incorrect information in a direct way to trick users." Controlling is when "designers interrupt or restrict the user's task flow, explicitly directing the task flow."

Recent studies have also attempted to measure the prevalence of dark patterns—those that appeared in the taxonomies above—on the web and in mobile apps. Using automated web crawling methods, Mathur et al. [35] identified 1,800 dark patterns on shopping websites including dark patterns such as Sneaking, Social Proof, and Countdown Timers. Utz et al. [56] analyzed the cookie banner notices of a random sample of 1,000 of the most popular websites in the EU, finding that over 50% of the banners contained at least one dark pattern. The dark patterns that the authors identified included privacy-unfriendly defaults, hidden opt-out choices, and preselected checkboxes that enable data collection. In another measurement study focused on Android apps, Di Geronimo et al. [12] found that more than 95% of the 200 most popular apps contain at least one dark pattern from the Gray et al [21] taxonomy. Nouwens et al. [39] examined 680 of the most popular consent management platforms—services that help websites manage user privacy consent—on the top 10,000 websites in the U.K., finding that only 11.8% of websites do not contain any dark patterns. Soe et al. [47] manually collected and analyzed 300 data collection consent notices from Scandinavian and English news websites. The authors found that nearly every notice (297) contained at least one dark pattern.

Combining the diverse range of dark pattern taxonomies and types with the diverse range of definitions sharply illustrates the lack of conceptual consistency in prior work. The Bösch et al. [4] and CNIL [42] definitions, for example, provide that a dark pattern must be misleading. But there is nothing inherently misleading about many of the dark patterns described in the literature, such as Conti and Sobiesk's [9] Coercion and Forced Work patterns, Gray et al.'s [21] Forced Action dark pattern, or Mathur et al.'s [35] Obstruction pattern. These user interface designs can be entirely truthful with users.

What's more, there is some conceptual inconsistency even within papers—projects can describe dark pattern types that do not appear to fit within the provided definition. Brignull [3] and Bösch et al. [4], for example, describe dark patterns as "tricks." It is difficult to see where the trick is, however, in Brignull's Confirmshaming or Bösch et al.'s Forced Registration—these user interface designs are often entirely (and frustratingly) transparent to users. Similarly, Zagal et al. [62] describe dark patterns as user interface designs that likely lack consent. The Grinding dark pattern discussed in the paper, however, inherently relies on the user's continuing consent in playing the game. Gray et al. [21] define dark patterns as "deceptive functionality." But there is nothing inherently deceptive about the paper's Nagging dark pattern.

In sum, the dark pattern definitions and taxonomies in prior work have been exceedingly valuable for surfacing descriptive insights and calling attention to problematic practices. But the dark patterns literature currently lacks a clear and consistent conceptual foundation.

2.3 Attributes of Dark Patterns

A recent paper by Mathur et al. [35] attempted to take a step beyond dark pattern definitions and taxonomies, offering a set of shared higher-level attributes that could descriptively organize instances of dark patterns in the literature. We discuss the Mathur et al. attributes below, and we propose an additional attribute.

2.3.1 Asymmetric. Asymmetric dark patterns impose unequal burdens on the choices available to the user. The choices that benefit the service are feature prominently while the options that benefit the user are typically tucked away behind several clicks or are obscured from view by varying the style and position of the choice. Asymmetric dark patterns are particularly common in consent interfaces. For instance, the Trick Questions dark pattern can impose a cognitive burden on choices that withhold consent by using confusing language and double negatives. The Confirmshaming dark pattern can use emotion to burden choices, associating guilt with certain choices and not others. Privacy Zuckering burdens choices with user interface friction, hiding privacy-respecting settings behind obscure menus.

2.3.2 Covert. Covert dark patterns push a user toward selecting certain decisions or outcomes, but hide the influence mechanism from the user. Some dark patterns achieve this property by exploiting cognitive biases, while others use color and style to steer users. Dark patterns that fit within Brignull's [3] Misdirection category are often covert. For instance, a shopping website might add a "free" gift to the user's shopping cart (e.g., a magazine subscription) to inflate the apparent shopping cart discount in a way that is not readily apparent. Websites may leverage the decoy effect cognitive bias, in which an additional choice—the decoy—makes certain other choices seem more appealing. Users may fail to recognize the decoy's presence is merely to influence their decision making.

2.3.3 Deceptive. Deceptive dark patterns induce false beliefs in users through affirmative misstatements, misleading statements, or omissions. Deceptive dark patterns are particularly common on shopping websites. For instance, the Countdown Timer dark pattern involves a phony deadline for a deal or offer, with a prominent ticking timer showing that the deadline is imminent. Mathur et al. [35] observed that over 40% of all the timers they discovered on shopping websites either reset on expiration or were inconsequential to the advertised offer.

2.3.4 Information hiding. Information hiding dark patterns obscure or delay the presentation of necessary information to users. One example is the Sneaking category of dark patterns. Another is the Hidden Subscription dark pattern, which does not reveal to users that the purchase they are making is a recurring subscription. Similarly, the Hidden Costs dark pattern hides relevant cost information until right before the user completes the transaction and is unlikely to change their course of action.

2.3.5 Restrictive. Restrictive dark patterns reduce or eliminate the choices presented to users. For example, the Forced Action dark pattern on a website might require users to agree to both the terms of use as well as receiving marketing emails before creating an account. Similarly, the Hard to Cancel dark pattern makes it easy for users to sign up for subscription services online, but does not give the user an online option to cancel that same service.

2.3.6 Disparate treatment. The attributes from Mathur et al. [35] describe most dark patterns that appear in most taxonomies, but do

Table 2: Dark patterns attributes grouped by how they modify the user's choice architecture.

Choice Architecture	Attribute	Description
Modifying the decision space	Asymmetric	Unequal burdens on choices available to the user
	Restrictive	Eliminate certain choices that should be available to users
	Disparate Treatment	Disadvantage and treat one group of users differently from another
	Covert	Hiding the influence mechanism from users
Manipulation the information flow	Deceptive	Induce false beliefs in users either through affirmative misstatements, misleading statements, or omissions
	Information Hiding	Obscure or delay the presentation of necessary information to users

not cover a category of dark patterns that disadvantage and treat one group of users differently from another. These dark patterns are commonly found in games, as described by Zagal et al. [62]. For example, the Pay to Skip interface on gaming sites allows users with more resources to gain an edge over users who cannot afford to pay. Other interfaces that systematically disadvantage one group of users often operate covertly so that users may not know why they are shown a limited set of choices [23]. To account for such interfaces we add disparate treatment as an attribute of dark patterns. Table 2 summarizes the complete set of attributes.

2.4 Dark Patterns as a Thematic Research Area

The set of attributes described above exhibit commonalities. Dark patterns that are asymmetric, covert, or restrictive, or that involve disparate treatment, attempt to influence user decisions by **modifying the set of choices** available to users. Dark patterns that are deceptive or information hiding influence user decisions by **manipulating the information** that is available to to users. Ultimately, both of these themes reflect how dark patterns **modify the underlying choice architecture** for users.

As we show in Table ??, these themes—choice architecture, subdivided into user decision space and information flows—cohere the current dark patterns literature. We are able to provide a complete and cohesive descriptive taxonomy of dark patterns scholarship, from these themes, to attributes, to specific dark pattern user interface designs that are described in the prior work.

Importantly, *themes* are what bind the current dark patterns scholarship together. We can only descriptively characterize the field at a high level of generality, and even then, it reflects a pair of related but distinct concepts. If we attempt to describe dark patterns with a greater degree of specificity—which would be essential for any actionable definition—the dark patterns literature immediately fragments further.

The structure of the dark patterns discourse has a parallel to Solove's seminal perspective on privacy discourse [48]. In the same way that Solove argued that privacy is a Wittgensteinian family of concepts that are distinct, but resemble each other, our descriptive deconstruction and reconstruction of the dark patterns literature reveals a set of interrelated thematic concerns. An important consequence of Solove's insight for privacy literature was that scholars could avoid being mired in distilling a unified definition of privacy,

and instead focus on the normative considerations that motivate the different concepts at play.

We now turn to that task for dark patterns scholarship. Using our thematic observations about choice architecture as a starting point, we review related work in other disciplines. We then develop discrete normative perspectives that can help researchers more effectively articulate and study the problems that dark patterns pose for individuals and society.

3 CHOICE ARCHITECTURE IN BEHAVIORAL ECONOMICS, PHILOSOPHY, AND LAW

While the HCI community has begun to develop a dark patterns literature, other disciplines—including psychology, economics, philosophy, and law—have grappled with similar issues that result from modifications to users' choice architecture in contexts beyond user interfaces. Unlike HCI, however, these disciplines have deeply examined the normative implications that arise from modifying choice architecture. The HCI community should engage with the substantial scholarship in these disciplines to move beyond its current descriptive mode of thinking and to surface the normative implications of dark patterns. We synthesize representative scholarship from these disciplines below.

3.1 Nudge and Sludge in Behavioral Economics

The field of behavioral economics is the study of psychology as it relates to the economic decision-making processes of individuals and organizations. Behavioral economists focus on the limits of human rationality, and how informational and cognitive capabilities influence these limits. In particular, they examine how human decision-making under uncertainty deviates from classical economic models because of how gains and losses are valued [55].

While the intellectual foundations for this approach were built in the late 1970s, the last 15 years has seen a surge in work that applies the lessons from behavioral economics to achieve various public policy goals. During this time, Thaler and Sunstein [54] conceived of "nudges," which they defined as "private or public initiatives that steer people in particular directions but that also allow them to go their own way." Numerous types of nudges have found their way into public life. One common example is that setting the default choice to enroll users in a retirement savings plan boosts participation by overcoming inertia [28].

Table 3: Types of dark pattern interfaces classified along how they modify the user's choice architecture.

| Prior Work | Dark Pattern Type | Choice Architecture | | | | | |
| | | Modify the Decision Space | | | | Manipulate the Info. Flow | |
		Asymme.	Restri.	Dis. Treat.	Covert	Decept.	Info. Hiding
Brignull [3]	Bait and Switch					●	
	Confirmshaming	●					
	Disguised Ads					●	
	Forced Continuity					◐	●
	Friend Spam					●	
	Hidden Costs					◐	●
	Misdirection				●		
	Price Comp. Prevention		●				●
	Privacy Zuckering	◐	◐				◐
	Roach Motel		●				
	Sneak into Basket					◐	●
	Trick Questions	●			●		
Conti & Sobiesk [9]	Coercion		●				
	Confusion	●			●		
	Distraction	◐			●		
	Forced Work	◐	◐				
	Interruption	◐	◐				
	Manipulating Navigation	◐			◐		
	Obfuscation	◐			◐		
	Restricting Functionality	◐	◐				
	Trick				◐	◐	
Zagal et al. [62]	Grinding		●				
	Impersonation					●	
	Monetized Rivalries			●			
	Pay to Skip			●			
	Playing by Appointment		●				
	Pre-Delivered Content			●		◐	●
	Social Pyramid Schemes		●				
Greenberg et al. [22]	Attention Grabber	◐			●		
	Bait and Switch					●	
	Captive Audience	◐	●				
	Disguised Data Collection				●		◐
	Milk Factor		●				
	Never Forget				●		
	Personal Info. Public		●				
	Unintended Relationships				●		
Bösch et al. [4]	Address Book Leeching					●	
	Bad Defaults	●					
	Forced Registration		●				
	Hidden Legal. Stip.						●
	Immortal Accounts	◐	◐				
	Privacy Zuckering	◐	◐				◐
	Shadow User Profiles		●		●		

Table 3: Types of dark pattern interfaces classified along how they modify the user's choice architecture.

| Prior Work | Dark Pattern Type | Choice Architecture | | | | | |
| | | Modify the Decision Space | | | | Manipulate the Info. Flow | |
		Asymme.	Restri.	Dis. Treat.	Covert	Decept.	Info. Hiding
Gray et al. [21]	Forced Action		●				
	Interface Interference	◑			◑		◑
	Nagging	●					
	Obstruction		◑		◑		
	Sneaking					◑	●
NCC [18]	Default Settings	●					
	Ease	●					
	Framing	●			◑		
	Rewards & Punishment	●	◑				
	Forced Action		●				
CNIL [42]	Attention Diversion	◑			◑		
	Bait & Change					●	
	Blaming the Individual	●					
	Camouflaged Advertising					●	
	Chameleon Strategy	◑					●
	Comparison Obfuscation		●				●
	Default Sharing	●					
	False Continuity					◑	●
	Hard to Adjust Privacy	●					
	Impenetrable Wall		●				
	Improving Experience				◑		◑
	Just You and Us				◑		◑
	Last Minute Consent		◑		◑		
	Obfuscating Settings	●					
	Repetitive Incentive	◑					◑
	Safety Blackmail					●	
	Trick Question	●			●		
	Wrong Signal						●
Mathur et al. [35]	Forced Action		●				
	Misdirection	◑			◑	◑	
	Obstruction		●				◑
	Scarcity				◑	◑	◑
	Sneaking					◑	●
	Social Proof				◑	◑	
Lacey & Caudwell [29]	Cuteness of Robots				●		
Gray et al. [20]	Automating the User		●				
	Two-Faced	●			●		
	Controlling	◑	◑		◑		
	Entrapping		●		●		
	Nickle and Diming	●	◑			◑	●
	Misrepresenting					●	●

● Required "dark pattern" attribute   ◑ Optional "dark pattern" attribute

Initially, Thaler and Sunstein focused on the beneficial uses of nudges to further normative goals that increased individual and societal welfare [49]. More recently, they have [50, 53] denounced "sludges"—nudges that induce excessive or unjustified friction and "cost time or money; make life difficult to navigate; may be frustrating, stigmatizing or humiliating; and that might end up depriving people of access to important goods, opportunities and services." Ultimately, nudges or sludges are just tools to achieve (or detract from) particular policy goals and do not supply the normative rationale for what goals to adopt.

3.2   Manipulation in Philosophy and Ethics

For decades, philosophy and ethics scholars have examined what it means to manipulate someone. Scholars agree that to manipulate an individual is to subvert their decision-making and to deny them authorship over their decisions [17, 44, 51, 60, 61]. Manipulation is distinct from techniques like coercion and persuasion. Coercing someone is forcing or threatening them to take an action, while persuading someone involves a forthright appeal to their conscious decision-making capacities.

There is, however, a debate among these scholars on whether manipulation operates as a hidden force. Susser et al. [51], for example, argue that manipulation is covert influence, meaning that the manner of the influence is not apparent to the individual being influenced. But others, like Wood [61] and Noggle [38], argue that manipulation can be overt, citing examples where "irresistible incentives" lure individuals into selecting certain choices over others [44].

From a normative standpoint, scholars have also raised numerous concerns about how manipulation can be exploitative, impoverish individuals, unfair, and—most important of all—undermine individual autonomy [17, 51]. Manipulation can cause individuals to falsely believe they are making their own choices and are in control of their own destiny, and manipulation that targets individual vulnerabilities can amplify these effects.

3.3   Market Manipulation in Law

Legal scholars Hanson and Kysar [24] were the first to describe a theory of market manipulation. They described market manipulation as the process of marketers and sellers exploiting individual non-rational behavior and cognitive biases for profit. For example, the framing effect bias explains variations in how users perceive a level of risk differently depending on the manner in which risk information is presented. Manufacturers exploit this cognitive bias in consumers to undermine the severity of product safety warnings to further their profits.

Calo [5] argues that digital marketplaces and online stores make market manipulation even more effective because of the scale and sophistication with which these marketplaces can target users. For example, unlike brick-and-mortar stores, digital marketplaces can specifically learn about and optimize manipulative techniques to individual users, thereby increasing their average effect across the entire population. Calo and Rosenblat [6] also argue that newer forms of marketplaces such as the sharing economy present at least as many—if not more—opportunities for digital market manipulation. Unlike traditional commerce marketplaces, the sharing economy marketplaces can manipulate both the sellers (e.g., Airbnb hosts) and the buyers (e.g., travelers).

In terms of the normative considerations of digital market manipulation, Calo [5] argues that it can increase economic harm and reduce market efficiency. Market manipulation techniques that rely on personal data can lead to a loss of privacy for users and also undermine their autonomy.

## 4   NORMATIVE PERSPECTIVES ON DARK PATTERNS

The previous section demonstrates how choice architecture problems in other disciplines have surfaced normative considerations that help analyze these problems. Unlike those disciplines, the dark patterns literature has largely focused on the descriptive aspects of dark patterns. Underlying these descriptive aspects, however, are a set of nascent normative concerns that attempt to explain *why* dark patterns should concern us. We draw out these normative considerations from the literature and organize them into four normative lenses. We acknowledge that the considerations that these lenses offer will not appeal equally to everyone concerned about dark patterns. Rather, our goal here is to set the stage for researchers to develop a common language to discuss problematic practices.

4.1   Individual Welfare

The individual welfare normative lens views dark patterns on the basis of whether they diminish individual consumer welfare. Under this lens, a dark pattern is any interface that modifies the choice architecture to benefit the designer at the expense of the user's welfare. The dark patterns literature has highlighted the impact to individual welfare as the primary normative concern about dark patterns. When researchers argue that a dark pattern goes against the user's "best interests," "harms" the user, or creates a "negative experience" for users [18, 20, 20, 31, 57, 62], they focus on some aspect of individual welfare. We discuss three kinds of individual welfare that can be diminished through dark patterns.

4.1.1   Financial loss. The most straightforward welfare consequence to users is when a dark pattern causes them to suffer a financial loss. Shopping and travel websites have a wide variety of interfaces that push users into spending more than they originally anticipate. Examples of such interfaces include those that profit from users by adding products into shopping carts without users' consent (Sneak into Basket) and those that mislead users into believing they are signing up for a one-time offer or free trial when in reality they are signing up for recurring fees (Hidden Subscription). Across the web, advertisements that masquerade as non-advertising content (Disguised Ads) can also mislead users into purchasing products they may not have otherwise purchased [36].

Concerns about financial loss echo throughout the dark patterns literature. Conti and Sobiesk [9] focus on "malicious interfaces" that "are employed for a variety of reasons that are often linked to direct or indirect acquisition of revenue [such as] selling a product or service, increasing brand recognition..." Zagal et al. [62] and Lewis [31] argue that dark patterns cause "negative experiences" for users and tag those dark patterns that "are a means for companies to extract extra money from users" as monetary dark patterns. Brignull
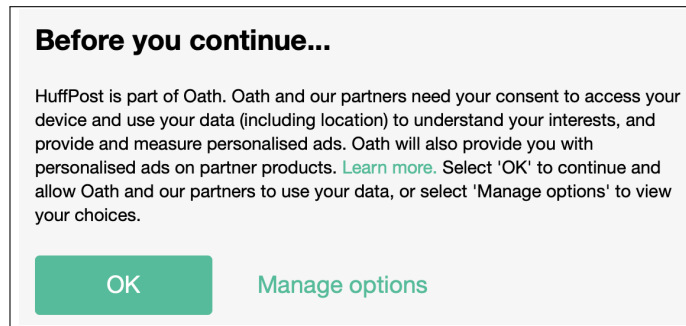
Figure 2: A dark pattern in the cookie banner notice on Huff-Post. The interface provides a positive option ("OK") to consent to tracking and data collection but the negative option is hidden behind a sequence of seven screens starting from "Manage Options."
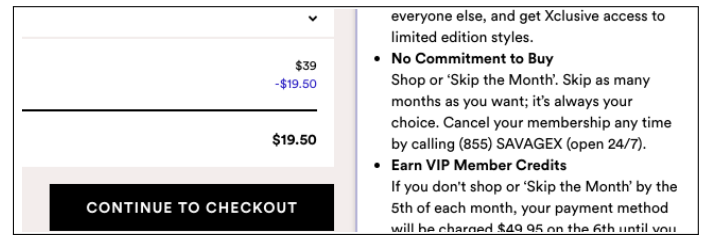

Figure 3: The Obstruction (Hard to Cancel) dark pattern on Savage X Fenty. While users can sign up for a membership online, they can only cancel their membership by placing a phone call to customer care.

defined dark patterns as tricks that mislead users into "buying or signing up for something [that they didn't mean to]."

4.1.2 Invasion of privacy. Another welfare consequence to users is when a dark pattern undermines their privacy, for example, by undermining notice-and-choice privacy regimes.[7] Examples of such interfaces include privacy-invasive defaults that expose user data (Bad Defaults), privacy-respecting choices that are hard to access (Privacy Zuckering, Interface Interference, and Obstruction), and the use of fear or other emotion-laden language to drive users away from making privacy-respecting choices (Blaming the Individual, Confirmshaming, and Framing).

The dark patterns literature is ripe with concerns about privacy as an aspect of individual welfare. Through their "malicious interfaces" framing, Conti and Sobiesk [9] argue that dark patterns extract revenue from the user by gathering their personal information. Bösch et al. [4] argue that "[i]nstead of privacy-friendly solutions, [the proponents of dark patterns] aim for systems that purposefully and intentionally exploit their users' privacy—for instance motivated by criminal reasons or financially exploitable business strategies." The NCC's report on dark patterns [18] states that "[t]he practice of misleading consumers into making certain choices, which may put their privacy at risk, is unethical and exploitative." Waldman [57] argues that "[dark patterns] confuse users by asking questions in ways non experts cannot understand, they obfuscate by hiding interface elements that could help users protect their privacy, they require registration and associated disclosures in order to access functionality, and hide malicious behavior in the abyss of legalese privacy policies." A whole host of subsequent research considers how dark patterns diminish privacy [33, 39, 42, 47, 56, 59]. As an aside, it is important to note that while we take a welfarist view on privacy here, there are other views that describe privacy as a public good, a human right, or an aspect of individual autonomy.

4.1.3 Cognitive burden. The last welfare consequence to users is when a dark pattern causes them to expend unnecessary time,

energy, and attention. Imposing such a cognitive tax on users, especially when they have competing claims on their attention, can push users to select the easiest available choice, which suits the service at the expense of the user. Examples of such interfaces include those that obstruct users from canceling the online services they are subscribed to by requiring them to phone in only during certain hours (Hard to Cancel) and those that repeatedly prompt the user to accept certain choices (Nagging). In the case of Instacart—a popular grocery delivery application in the U.S.—the Nagging dark pattern continually beeped delivery workers' phones until they agreed to to accept lower-paying jobs [13].

The concern about the cognitive burden of an interface has not been cited an explicit normative concern in the dark patterns literature. If the concern is raised, it is in the context of specific dark patterns such as Nagging, Hard to Cancel, and Hidden Legalese Stipulations (and other information overloading dark patterns). As one example, in the context of cookie consent dialogs, Soe et al. [47] argue that cookie consent dialogs without a negative option to deny consent "also introduce additional cognitive burden on the user." Figure 2 shows once instance of a cookie consent dialog with a hard to exercise deny option.

Evaluation. The challenge for the individual consumer welfare perspective is that it is difficult to describe the baseline standard we should have in mind when examining potential dark patterns. Individual preferences and desires are difficult to measure for a variety of reasons, including because they are unstable, vary across individuals, and may require large-scale measurement. Critics also point out that all marketing involves shaping user preferences to varying degrees and so this lens says little about how we can distinguish benign or tolerable practices from ones that deserve sanction. Another challenge is that this perspective does not account for the dynamics of competitive markets where users can punish bad actors for mistreating them by abandoning them or imposing some reputational cost. Finally, this perspective does not account for potential welfare benefits that may flow from allowing companies to shape the decision space or manipulate information flow, or the costs of imposing liability for all manners of marketing practices.

4.2 Collective Welfare

The collective welfare lens takes a broader normative perspective on consumer welfare compared to the individual welfare lens, and views dark patterns on the basis of whether they diminish collective

---

[7]Readers may also find Hartzog's *Privacy's Blueprint: The Battle to Control the Design of New Technologies* [25] as a valuable examination of how user interface design can impact privacy.

welfare. Under this lens, a dark pattern is any interface that modifies the choice architecture to benefit the designer at the expense of collective welfare. Unlike individual welfare, which has been the primary normative focus of the dark patterns literature, collective welfare has received limited attention. We discuss four kinds of collective welfare associated with society and markets that can be diminished through dark patterns:

4.2.1 Competition. An efficient and competitive marketplace enables innovation, results in lower prices for products, and helps consumers align with products that match their preferences. However, dominant providers in particular can abuse their position of monopoly power and diminish competition by making it look like consumers independently selected their product rather than from a series of dark patterns that disfavor offers from competitors. Dark patterns such as a pre-selected checkboxes (Preselection) and use of emotions and fear (Framing) can allow providers to tie-in unrelated products to unfairly gain market share despite having an inferior tied product. Dark Patterns from the Obstruction category (Figure 3) can assist providers in creating high switching costs for users to potentially superior competitors while simultaneously creating a high barrier of entry for others. A prominent example is the Microsoft antitrust case,[8] where Microsoft used the Obstruction dark pattern to prevent users from uninstalling Internet Explorer like other programs in Windows 98.

The dark patterns literature has only provided limited commentary about competition concerns emerging from dark patterns. An exception is the work of Day and Stemler [11], which raises the primary concern that dark patterns "erode users' ability to act rationally, which empowers platforms to extract wealth and build market power without doing so on the merits." Day and Stemler argue that dark patterns harm competition because "when the design is meant to generate switching costs without an offsetting benefit to users," and that the key in making this judgement should be "whether the design was meant to enhance addiction and manipulate usage while providing consumers with a qualitatively worse product."

4.2.2 Price transparency. Price transparency enables consumers to make informed decisions and thus creates an efficient marketplace. Providers do not, however, always have the incentive to compete on prices. For example, oil companies were using a variety of names and descriptions for the same product until the FTC stepped in in 1971 to issue rules requiring standardized octane ratings for fuel to prevent consumer confusion.[9] A more contemporary instance is the use of different product names and attributes to sell the same product, as exemplified by the flat screen TV market.

Dark patterns hide the true costs of products from consumers and prevent them from comparison shopping. Dark patterns like Hidden Costs and Price Comparison Prevention impede consumers from making informed decisions. As a result, not only does opaque pricing cause potential financial loses for consumers—they might have selected a cheaper alternative if they were able to compare and contrast competing prices—but it also reduces competition in the marketplace. Economists argue that under certain conditions

providers gain no benefit from exposing the opaque prices of their competitors [19], and instead, they find it more profitable to employ and copy the same practices.

4.2.3 Trust in the market. Dark patterns can undermine consumer trust in markets and hurt companies who engage in legitimate and honest practices. As these patterns proliferate, users who become aware of them may become skeptical of and resistant to interface elements that *look* like dark patterns. For example, studies have shown that users have become resistant and wise to Scarcity dark patterns [46]. But this increased skepticism in users may lead them to miss out on genuine deals from honest retailers and hurt the business of those companies. Through their study, Luguri and Strahilevitz [32] demonstrate that web services making use of egregious dark patterns can suffer repercussions, as users are more likely to report negative emotions after interacting with them. While this might suggest dark patterns are ultimately self-defeating, the problem is that the bad actor may not be the only one paying the price.

4.2.4 Unanticipated societal consequences. More generally, design choices can also lead to consequences that are not originally anticipated by the designer. For example, interfaces employed by companies like Facebook make users turn over private data to allow the company to amass large databases with detailed profiles about individuals that it can use for advertising. But that information in turn can also be repurposed to create products that can undermine societal values. A prominent example here is how Cambridge Analytica harvested Facebook user data and initiated a political disinformation campaign to impact the U.S. presidential election in 2016. As a result, a collective consumer welfare perspective requires evaluating potential unintended or side effects of design choices.

Evaluation. This attention to collective welfare consequences also explains how while certain dark patterns may not present a problem to a particular user's welfare, they can impose a cumulative cost on other members of society that is intolerable. An example is the Social Pyramid Schemes in gaming apps that encourage users to allow the service to spam their friends and acquaintances or nudge them to sign up. The users may understand and accept the trade-off, but there is a significant cost to society as a whole from tolerating such practices. The collective welfare approach depends on having a mechanism to balance the different costs and benefits that result from design choices. Should designs deliver the great good for the greatest number of people? Or should design protect the interests of more vulnerable communities of users?

4.3 Regulatory Objectives

The regulatory objectives lens uses democratically created rules and standards to view when dark patterns cause individual and collective harms such as diminishing the individual's financial welfare and undermining fair market competition respectively. Under this lens, a dark pattern is any interface that modifies the choice architecture to interfere with or undermine specific regulatory objectives. The lens takes the operating legal regime as a given in making this judgement about an interface.

---

[8]United States v.Microsoft Corp., 253 F.3d 34, 65 (D.C. Cir. 2001).
[9]Statement of Basis and Purpose, Posting of Minimum Octane Numbers on Gasoline Dispensing Pumps, 36 Fed. Reg. 23871, 23882 (1971).

Generally speaking, the underlying regulations that define the standards that any interface needs to meet do not have to be described with a great deal of specificity. For example, interfaces that are deceptive because they omit material information or rely on false information can run afoul of well established consumer protection laws prohibiting such practices. As a result, regulators do not have a single standard to decide what constitutes a problematic practice—they tend to take a case-by-case approach to address distinct problems. Sometimes authorities also pass specific laws to regulate certain practices like the EU did with the General Data Protection Regulation (GDPR) [14] to handle privacy.

In the U.S., the Federal Trade Commission (FTC) prominently uses it authority under Section 5(a) of the Federal Trade Commission Act to determine whether a company has acted unfairly or deceptively.[10] The FTC evaluates whether a practice is deceptive from the perspective of a "reasonable consumer"—meaning, does the practice leave a false or misleading impression on a typical consumer? The Commission also has the authority to prohibit non-deceptive practices that have the effect of treating consumers unfairly. The FTC has a long history of taking action to "halt some form of seller behavior that unreasonably creates or takes advantage of an obstacle to the free exercise of consumer decision-making," even without proof of financial loss [1]. In such cases, the FTC determines whether the act or practice "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition."[11] State attorneys general have analogous powers under state law to prohibit unfair or deceptive practices.

In recent years, however, the FTC has focused on only providing remedies for "significant injuries" like financial losses. The FTC's reasoning is based on the prevailing "efficient market hypothesis" among U.S. regulators, which posits that because users value choice and independent decision-making, companies will compete to provide products and services that do not take advantage of consumers [43]. Consequently, companies that fail to match that demand will not succeed in this competitive marketplace. Industry self-regulatory standards to police harmful conduct can also supplement this reliance on the markets. But if there are instances of market failures, the authorities turn to public and private enforcement actions to reign in the companies' conduct.

Authorities can also pass regulations that protect users by including mandatory disclosures about material facts or a cooling off period to allow users the time to evaluate information in a less pressured setting. There is considerable variation in how different legal regimes approach the problem. In the EU, for example, policymakers are predisposed to pass detailed laws and regulations defining permissible and impermissible conduct, whereas U.S. regulators tend to take a principles-based approach.

Aside from unfair and deceptive standards, legislators in the U.S. have proposed the DETOUR Act [58] to regulate broad swaths of dark patterns. The DETOUR Act proposes a standard for dark patterns and makes it unlawful for a large online operator to "to design, modify, or manipulate a user interface with the purpose

---

[10] 15 U.S.C. § 45(a)(1).
[11] 15 U.S.C. § 45(n).

or substantial effect of obscuring, subverting, or impairing user autonomy, decision-making, or choice to obtain consent or user data." Among other things, the bill requires online sites to be more transparent about how they use design features to influence user behavior. Relatedly, the recently passed CPRA [40] makes it illegal for businesses for obtain certain privacy consent through the use of dark patterns, which the law defines as "user interfaces designed or manipulated with the substantial effect of subverting or impairing user autonomy, decision-making, or choice, as further defined by regulation."

The dark patterns literature has offered some initial commentary on whether dark patterns are lawful. Conti and Sobiesk [9] note that "malicious interfaces" can "obfuscat[e] legally mandated but undesirable information from the user." Mathur et al. [35] argue that the deceptive dark patterns they discovered are unambiguously unlawful under Section 5 of the FTC Act and various state laws in the U.S. Luguri and Strahilevitz [32] offer a detailed commentary about the lawfulness of dark patterns under Section 5 along with other federal frameworks. However, only recently has the literature begun evaluating dark patterns against specific regulatory objectives. The majority of this discussion has resulted from studies [33, 39, 47, 56] examining whether the consent management platforms that have emerged on websites as a result of the GDPR are compliant with the law. Nouwens et al. [39] consider consent management platforms compliant with the GDPR if the consent they aim to collect through their interface is explicit, is easy to accept or deny, and contains no preselected checkboxes. Failing to meet these criteria makes the consent interface a potentially unlawful dark pattern. Soe et al. [47] similarly provide a list of eight criteria that would make a consent management platform interface a dark pattern under the GDPR.

Evaluation. The regulatory objectives perspective on dark patterns is more instrumental then normative—the most forceful normative arguments for implementing regulatory objectives are typically the normative arguments for those objectives, rather than compliance with regulation in the abstract. This perspective does not inherently advance a normative argument about why we should care financial losses, privacy harms, or cognitive burdens, beyond noting whether the law directs us to care about those values. This perspective does come with a significant advantage, though: fashioning regulation into measurable metrics for empirical research is usually much easier than adapting normative principles to research.

4.4 Individual Autonomy

In contrast to the welfare and regulatory perspectives we have discussed so far, the individual autonomy lens is a rights-based lens. Autonomy is the normative value that users have the right to act on their own reasons when making decisions [17, 51]. The autonomy perspective shares some common ground with the individual welfare perspective, but it does not make a cost-benefit decision to determine if an interface is problematic. Under this lens, a dark pattern is a user interface that undermines individual decision-making. A dark pattern that infringes on individual autonomy might modify choice architecture in a way that causes users to make choices that they would not have otherwise selected absent the modified

choice architecture.[12] Alternatively, a dark pattern might deny a user choice, obscure available choices, or burden the exercise of choice.

The vast majority of dark patterns attempt to undermine individual autonomy. One set of such dark patterns include those that enable addiction. Online services have incentives to maximize user engagement with their platform that is at odds with user autonomy. Dark patterns that cause digital addiction can lead to a variety of negative consequences including psychological (e.g., poor concentration), and physical (e.g., sleep disturbance) harms [2]. In fact, Schüll observed that online services have created techniques such as the infinite scroll user interface, or the YouTube autoplay feature in line with those developed by the gambling industry—many of which employ a variable-rewards schedule approach to keep users engaged [31, 45]. Lootboxes—variable reward containers that give users advantages over others in video games—have also come under scrutiny because of their ability to continually engage kids and teen gamers [26].

Concerns about autonomy echo through nearly the entire dark patterns literature. Brignull [3] defined dark patterns as "tricks that make you do things that you didn't mean to..." The NCC's report [18] argues that dark patterns "deprive [users] of their agency." The DETOUR Act [58] and the CPRA [40] both describe dark patterns as "subverting or impairing user autonomy." Westin and Chiasson [59] argue that dark patterns "trick and manipulate users into taking an action they would've other-wise been unlikely to take." Further, with regards to determining whether dark patterns are ethical or not, Westin and Chiasson argue that "it is the loss of individual autonomy promoted by dark patterns that we should carefully consider."

Evaluation. The autonomy perspective is appealing, because it captures many of the concerns articulated in prior dark pattern definitions and taxonomies. One drawback of the autonomy perspective is that it sweeps broadly, dubbing all interfaces that interfere with decision making as dark patterns regardless of whether the outcome of that decision benefits the individual or society. Another challenge is that the autonomy perspective depends on a highly idealized version of human action that does not account for lived realities, where decisions are made with limited information and scarce resources for deliberation. Yet another challenge is line drawing: how do we distinguish between permissible burdens on autonomy (e.g., persuasive advertising that changes behavior) from dark patterns that violate autonomy? And, relatedly, how can we measure whether and the extent to which a user interface violates individual autonomy?

5 MEASUREMENT METHODS AND APPLICATIONS

The four normative lenses—individual welfare, collective welfare, regulatory objectives, and individual autonomy—offer us different perspectives with which we can analyze dark patterns. The

perspectives do not, however, tell us how to evaluate user interfaces in practice. In this section, we describe how HCI researchers can use established empirical methods to analyze dark patterns through these four normative lenses. Grounding empirical analysis in normative considerations enables researchers to surface the problematic aspects of dark patterns in a principled and defensible manner. Our goal is to encourage the HCI community to move beyond ad hoc and descriptive labels for dark patterns and focus future research on assessing how, exactly, dark patterns interact with specific normative concerns.

To set the stage, we first provide a brief description of various kinds of HCI research methods. We describe the breadth of information each method is capable of collecting and how that information can be subsequently analyzed. Next, we illustrate how the methods can be applied to study the normative lenses we previously described. We close the section by applying the methods in a case study to analyze the Trick Questions dark pattern.

5.1 Empirical Research Methods in HCI

The HCI community has championed the use of methods that elicit users' attitudes and behaviors to help design and evaluate user interfaces. Each one of these methods has its own merits and drawbacks, and—depending on how it is ultimately applied—provides researchers with a different set of insights about user interfaces. Note that this description of research methods is not meant to be comprehensive. Instead, we refer interested readers to standard HCI textbooks [30, 41].

One set of methods allows researchers to study user experience with a particular design by directly interacting with or observing users. This set of methods includes surveys, interviews, focus groups, and ethnography. Surveys are a structured set of questions that can be deployed to gather responses from a large sample of users. Interviews on the other hand are typically less structured and allow researchers to conduct an in-depth and back-and-forth discussion with users. Interviews enable a deeper inquiry because of their flexibility, but unlike surveys they are harder to scale. In focus groups, a researcher directs a discussion with a group of users to elicit individual and collective feedback about a design. Ethnography involves passively observe the behavior of individuals and groups in their own social contexts.

Lab and field studies are settings that allow researchers to examine actual user interaction with a user interface. In a lab study, researchers typically invite users into the lab and observe them interacting with a particular design. Researchers then ask users about their experience on interaction cues of interest, or after the study is complete in the form of a cognitive walkthrough. Lab measurement methods can be particularly rigorous, such as gaze tracking systems that enable researchers to understand user attention to specific components of user interfaces. An important drawback of the lab setting is that it often has limited external validity, because of sampling, response, and other biases. Lab insights may not generalize to how users interact with a design in the real world.

Instead, researchers may use field deployments to study user interaction *in situ*. Field deployments can include diary studies and experience sampling. Both these methods require users to complete a prompt of questions immediately after an event of interest

---

[12]This standard for a dark pattern that infringes on autonomy interests is strongly linked to the welfare perspective too, since the standard prioritizes outcomes consistent with consumer preferences. From an individual welfare perspective, consistency with individual preferences is desirable because preferences are a possible measure of welfare. From an individual autonomy perspective, consistency with individual preferences is desirable because preferences are a gauge of autonomous decision-making.

when using a particular design, outside of their interaction with researchers. These deployments can be particularly useful in studying the experience of sub-populations that may be otherwise hard to reach. Field deployments with measurement panels can provide valuable insights into how users behave in the presence of a particular user interface.

Experiments—the bedrock of the scientific method—enable generating cause-and-effect relationships. Typically in an experiment, researchers randomly assign users to the control or treatment conditions, and then measure some quantity of interest across those conditions. Statistical significance in the measure between the conditions reveals the effect of the treatment. As with the previous set of methods, experiments may be conducted in the lab or in the real world. Lab experiments provide a controlled setting, and they minimize the influence of external factors on the experiment. Field experiments on the other hand enable inquiry in a more realistic and everyday setting, and also help in recruiting a relatively more diverse sample of participants. Online marketplaces like Amazon Mechanical Turk and Prolific are popular venues to conduct such large-scale field experiments.

5.2 Applying HCI Methods to Analyze Dark Patterns

Each of the lenses we described in Section 4 lends itself to a different set of measurement methods, metrics, and goals. In the following sections, we describe how the measurement methods above can be used to analyze dark patterns. As the dark patterns literature has only recently begun employing user studies [9, 12, 32–34, 39, 56] to analyze user interfaces, we discuss these studies in the context of how they align with normative lenses.

5.2.1 What should we measure under each normative lens?

Individual welfare: The first normative lens describes how dark patterns diminish individual welfare. From this perspective, the goal of any measurement should be to measure the change in individual welfare that results from a dark pattern. If the individual welfare of concern is financial value, then a study should measure the financial loss suffered by users. An experiment comparing the dark pattern interface to a neutral design could measure the average monetary amount lost by users because of the dark pattern. The experiment conducted by Luguri and Strahilevitz [32] is one relevant example. The study measured the rate at which participants accepted a hypothetical insurance offering under the absence and presence of dark patterns, then calculated the amount of money lost by those who were subject to the dark pattern. On the other hand, if the individual welfare of concern is privacy, then a study should measure a welfarist conception of privacy. It could measure the financial value of the data that was disclosed due to the dark pattern, or it could measure the fraction of users whose preferences were not satisfied because of the dark pattern. These could be measured using a survey, field study, or experiment comparing the dark pattern interface to a neutral design. Finally, if the individual welfare of concern is cognitive burden, then a study could measure the cognitive burden imposed on users. A survey or experiment could employ various neurophysiological [16] or stated measures of cognitive load to compare dark patterns and neutral designs [52].

The survey conducted by Conti and Sobiesk [9] is an example of a user study that measured the level of frustration users felt with various kinds of dark patterns.

Collective welfare: The second normative lens describes how dark patterns diminish collective welfare. From this perspective, the goal of any measurement should be to measure the change in collective welfare that results from a dark pattern. If the collective welfare of concern is competition, then a study should measure the reduction in the competitiveness of the marketplace. Competition could be measured in several different ways. As one example, a study could first develop an understanding of individual behavior under a dark pattern when compared to a neutral design. It could then induce the broader effect of this behavior on competition in the marketplace, such as by using simulations and modeling. On the other hand, if the collective welfare of concern is marketplace trust, then a study should measure the loss of trust among the various market participants. A study could use a survey to measure consumers' perceptions of the trustworthiness of the marketplace in the absence and presence of a dark pattern.

Regulatory objectives: The third normative lens describes how dark patterns interfere with regulations. Under this lens, the goal of any measurement should be to assess whether a dark pattern complies with relevant regulation. For example, the GDPR requires that consent must be "freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." A study could measure compliance with the GDPR by examining whether users are able to express their privacy preferences and provide consent free of coercion in the presence of a dark pattern. The online experiments conducted by Nouwens et al. [39], Utz et al. [56], and Machuletz et al. [33] are relevant examples. All three studies analyzed whether designs commonly used by consent management platforms are compliant with the law. Among other metrics, each study measured the fraction of users who consented to have their data processed by a website under the presence of a dark pattern. The experiment conducted by Nouwens et al. [39] additionally measured the fraction of users whose preferences were satisfied by the consent interface. This additional dimension measures privacy from an individual welfare perspective.

Individual autonomy: The fourth normative lens describes how dark patterns undermine individual autonomy. From this perspective, the goal of any measurement should be to measure the extent to which a dark pattern interferes with users' ability to make independent decisions. Comparing a dark pattern to a neutral design, a study could use a survey or experiment to measure whether users perceive a loss of autonomy, whether users are aware of alternative choices, or whether users are unable to exercise their preferred choices. Again, the online experiments conducted by Utz et al. [56] and Machuletz et al. [33] are relevant examples. In measuring compliance with the GDPR, both experiments additionally measured autonomy. Machuletz et al. [33] measured whether users found it difficult to express certain choices through the consent interface. Utz et al. [56] measured the mental models that users had about the consent interface, including awareness of the set of choices offered by the interface.

5.2.2 What is an appropriate baseline for measurement?

Some of the measurements described in the previous section can be established on their own—without needing a baseline design to compare against. From the the regulatory objectives perspective, for example, measuring whether a dark pattern complies with a regulation might only require analyzing whether it complies with a clearly established rule or standard. Consider the FTC's endorsement guidelines, which regulate the disclosure of social media ads (the Disguised Ad dark pattern). The guidelines state that an undisclosed ad is problematic "if it misleads a *significant minority* of consumers." If a measurement reveals that a significant minority—for some threshold—has been misled, then the undisclosed ad does not comply with the guidelines. There is no need to evaluate the ad's design against a hypothetical alternative ad. As another example, consider the individual welfare and autonomy perspectives. A study might deem a design a dark pattern if a sufficient proportion of users, above some threshold, do not have their preferences or expectations satisfied by the user interface. The threshold could be a minority of users (as in the FTC's endorsement guidelines), or it could be a simple majority, or it could be a supermajority. Setting the threshold would require normative judgment, but again, would not require a comparison to alternative designs.

Another way to study a user interface without selecting a baseline is to compare the design to the space of nearly identical designs. Consider, for example, a consent form that involves difficult to understand prompts (the Trick Questions dark pattern). A study could compare the form to alternative versions of the form with slight differences in phrasing and presentation. If the initial form is an outlier in comparison to the space of similar forms (e.g., when comparing the proportion of users who agree or the proportion of users who have their preferences satisfied), that could be interpreted as evidence that the form is a dark pattern.

Many of the measurements described above, however, require a baseline alternative user interface for comparison.[13] In comparative research designs like experiments, envisioning this neutral control user interface—a design not containing the dark pattern—is an important consideration. There are important parallels here with how regulators have approached deceptive advertising. Some past enforcement actions have involved comparing a concerning ad to a "tombstone" control ad [10], which omits the potentially deceptive content.

So, how should HCI researchers pick an appropriate control design for dark pattern studies? We foresee at least three different options. First, the appropriate baseline may be specified by regulation itself, by what users reasonably prefer, or by what users might generally expect of the interface. Second, the appropriate baseline may be derived from the space of designs that are similar to the concerning user interface but that do not contain the possible dark pattern. In many instances, it will be possible to produce a slight variant of a user interface that simply omits the possible dark pattern. Third, the appropriate baseline may be selected from the space of similar designs used by other online services for similar choices. For example, if one online service's sign-up flow includes a

possible dark pattern component, but a competing service's sign-up flow does not include that component, then the competitor's user interface may be an appropriate baseline for comparison.

5.2.3 What is an appropriate threshold for identifying a dark pattern?

Regardless of whether an empirical study involves a baseline user interface, if a study aims to characterize designs as dark patterns (or not), it requires setting a threshold for what constitutes a dark pattern. Luguri and Strahilevitz [32] propose one possible threshold: if the uptake rate associated with a user interface is over doubled in comparison to an alternative user interface, then the user interface is a dark pattern. The rationale, they argue, is that this test means a user's uptake is more likely than not attributable to a dark pattern. While this is one possible approach, as noted in the prior section, there are many others. The threshold could be absolute, or relative via subtraction, or relative as a ratio, among other options. The threshold could be modest, or it could be a high bar for labeling a user interface a dark pattern. Setting the threshold requires careful normative justification.

5.2.4 Who is affected by the dark pattern?

Any user study ultimately needs to be representative of the users that are targeted by the dark pattern. This could be the general public or special sub-populations like children, older adults, or low-income individuals. A more difficult challenge concerns evaluating the effect of a dark pattern on third parties—users who do not interact with the interface directly. For example, addictive online services like games can have an detrimental effect on not just on the user, but also on their family and friends. Similarly, a number of growth hacking dark patterns rely on users giving easy access to their social network that might seem relatively trivial to them, but has a cumulative negative impact of dealing with unwanted communications that can be quite expensive for their friends and family. However, even if a dark pattern is targeted at the general public, it can have disproportionate impacts on certain sub-populations. Study designs should account for conducting such analyses.

5.2.5 What role can non-quantitative methods play?

While we have so far discussed quantitative methods, other HCI methods have an important role to play in analyzing dark patterns through the lenses we describe. Interviews and focus groups, like the ones conducted by Maier and Harr [34], could be used to analyze dark patterns that affect sub-populations that might be hard to reach otherwise. Future work could, for example, use ethnographic methods and observe designers to understand how user interfaces become dark patterns during development. Results from this work could shed light on multiple normative considerations.

5.2.6 How are dark patterns different from marketing?

Some commentators have questioned whether dark patterns are merely aggressive marketing practices, the sort of sharp dealing that has long existed (and been tolerated) for offline businesses [7, 27]. Researchers have responded by noting possible key differences between dark patterns and traditional marketing, including low cost, large scale, and unprecedented sophistication [35]. Researchers have especially highlighted the role of A/B testing and machine learning, which can easily generate and evaluate user interfaces in ways that may be problematic for users [35, 37]. Future work

---

[13]User interface designs that include multiple possible dark patterns pose an additional challenge. One approach would be to implement multiple baseline designs for comparison, examining each of the possible dark patterns separately. Another approach would be to combine the possible dark patterns for treatment and control purposes.

Figure 4: The Trick Questions dark pattern on Virgin Active. Opting out of marketing and promotions requires unticking the first checkbox and leaving the second checkbox ticked.

could involve measurements that are grounded in a normative perspective and specifically address how these factors compare to traditional marketing. For example, studies could compare online dark patterns to analogous offline marketing designs, or could simulate how machine learning might naturally converge on user interface designs that undermine autonomy or harm users.

5.3 Case Study: Trick Questions Dark Pattern

In the last section, we discussed how a wide variety of HCI methods could be used to analyze dark patterns. We now present a case study to concretely demonstrate how researchers might apply those methods to evaluate a user interface in a manner that is grounded in a normative perspective. The case study is based on the Trick Questions dark pattern, in which confusing language steers users into making certain choices. Figure 4 illustrates one instance of a Trick Questions user interface.

Each of the four normative lenses provides a different way of approaching a Trick Questions user interface. Starting with the individual welfare lens, a researcher might ask: does this user interface undermine individual welfare? The researcher might answer that question by conducting a randomized controlled lab experiment, examining the extent to which the interface respects user preferences. The researcher could generate an alternative baseline design—say, a user interface with unambiguous explanation of the user's choices—then compare the choices users make with the original design and the alternative design. If the choices with the alternative design are significantly closer to user preferences than the choices with the original design, the researcher could interpret that as evidence that the original design undermined individual welfare as a Trick Questions dark pattern. Alternatively, where user choices have a monetary component (e.g., making a payment), the researcher could compare the costs imposed on users between the original and alternative designs. Again, if there is a significant difference, the researcher could interpret that as evidence of a dark pattern from the individual welfare perspective.

Now consider the collective welfare lens, and assume the researcher's primary collective welfare concern is market trust. The researcher might again use a randomized controlled laboratory experiment, but measuring how users self-report trust in the relevant online services market before and after completing the experiment

task. If there is a significant drop in trust for the initial design but not the alternative design, that could be interpreted as evidence of a Trick Questions dark pattern that undermines collective welfare.

Turning to the regulatory objective lens, a researcher might seek to measure whether the user interface complies with a specific regulatory obligation, such as the GDPR's requirement that consent be "freely given, specific, informed and unambiguous." The researcher could conduct a survey, in which participants receive a screenshot of the user interface and are asked questions about what each option presented in the interface does. If only a small proportion of users correctly understand what the privacy options in the user interface do, that could be interpreted as evidence that the user interface is a Trick Questions dark pattern from the regulatory perspective.

Finally, consider the individual autonomy lens. The researcher might seek to understand whether users are aware that the user interface provides them with options, such that they can exercise independent decision-making. The researcher could again use a survey method, but focusing on whether participants even notice the available options in the user interface. If only a small proportion of users recognize that there are options available, that could be interpreted as evidence that the user interface involves Trick Questions from the individual autonomy perspective.

The example research approaches we have just sketched are far from exhaustive, and in many respects the approaches are relatively unsophisticated and naive. Our goal with this case study is to show that when approaching dark patterns measurement research, scholars can and should articulate an informed decision about the normative perspectives they are applying, then articulate a further informed decision about the research methods they will use to shed light on those normative considerations.

6 CONCLUSION

Our primary goal in this paper is to advance the dark patterns literature in the HCI research community. Our community can benefit from scholarship in other disciplines (Section 3), which we develop into a set of normative perspectives for assessing dark patterns (Section 4). We then show how empirical dark patterns research methods can be grounded in these specific normative perspectives (Section 5).

We encourage our colleagues to pursue this path in future work. Focusing on the normative foundations of dark patterns scholarship will motivate research problems and methods, and it will lend analytical rigor to understanding how dark patterns affect individuals and society. This approach will also benefit HCI practitioners by furthering our community's quest to develop ethical standards for user interface design. And by converging on principled methods for evaluating dark patterns, we can provide practitioners with a new toolkit for critically examining their own practices. This effort will also aid the HCI community in coming to terms with its responsibility for the societal consequences of its work. In particular, we have seen how insights about human interactions can used to exploit users. Our paper is not immune from that problem, but we trust that empowering researchers and regulators to examine the risks of dark patterns far outweighs the risk of describing mechanisms that a bad actor might seek to exploit.

Finally, grounding the measurement of dark patterns in normative perspectives will be invaluable for policymakers. Dark patterns have attracted serious legislative and regulatory attention in the U.S. and the EU. But policymakers have struggled with how to respond, since the issue is so amorphous and under-conceptualized. As a result, government actions on dark patterns have thus far remained limited to traditional consumer protection and privacy law. The HCI community can break through this logjam by developing metrics and methods for regulatory agencies to deploy.

ACKNOWLEDGMENTS

REFERENCES

[1] J. Howard Beales. 2003. The FTC's Use of Unfairness Authority: Its Rise, Fall, and Resurrection. https://www.ftc.gov/public-statements/2003/05/ftcs-use-unfairness-authority-its-rise-fall-and-resurrection.

[2] Pierre Berthon, Leyland Pitt, and Colin Campbell. 2019. Addictive De-Vices: A Public Policy Analysis of Sources and Solutions to Digital Addiction. *Journal of Public Policy & Marketing* 38, 4 (2019), 451–468. https://doi.org/10.1177/0743915619859852 arXiv:https://doi.org/10.1177/0743915619859852

[3] Harry Brignull. 2018. Dark Patterns. https://darkpatterns.org/. Accessed September 9, 2020.

[4] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 01 Oct. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proceedings on Privacy Enhancing Technologies* 2016, 4 (01 Oct. 2016), 237–254. https://doi.org/10.1515/popets-2016-0038

[5] Ryan Calo. 2013. Digital market manipulation. *Geo. Wash. L. Rev.* 82 (2013), 995.

[6] Ryan Calo and Alex Rosenblat. 2017. The taking economy: Uber, information, and power. *Colum. L. Rev.* 117 (2017), 1623.

[7] Federal Trade Commission. 2019. Competition and Consumer Protection in the 21st Century. https://www.ftc.gov/system/files/documents/public_events/1418273/ftc_hearings_session_12_transcript_day_1_4-9-19.pdf.

[8] Competition and Markets Authority. 2020. Online platforms and digital advertising market study. https://assets.publishing.service.gov.uk/media/5efc3faae90e075c4e144c69/Appendix_Y_-_Fairness_by_Design_Final_Version_v.8.pdf. Accessed September 9, 2020.

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

[10] Richard Craswell. 1996. Compared to What-The Use of Control Ads in Deceptive Advertising Litigation. *Antitrust LJ* 65 (1996), 757.

[11] Gregory Day and Abbey Stemler. 2020. Are Dark Patterns Anticompetitive? *Alabama Law Review* 72 (2020), 1–45.

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

[13] Josh Eidelson. 2019. Instacart Hounds Workers to Take Jobs That Aren't Worth It. https://www.bloomberg.com/news/articles/2019-07-15/instacart-hounds-workers-to-take-jobs-that-aren-t-worth-it. Accessed September 9, 2020.

[14] European Parliament and Council of European Union. 2018. Consent under the GDPR: valid, freely given, specific, informed and active consent. https://www.i-scoop.eu/gdpr/consent-gdpr/. Accessed September 9, 2020.

[15] Brian J Fogg. 1999. Persuasive technologie301398. *Commun. ACM* 42, 5 (1999), 26–29.

[16] Tycho K Fredericks, Sang D Choi, Jason Hart, Steven E Butt, and Anil Mital. 2005. An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. *International Journal of Industrial Ergonomics* 35, 12 (2005), 1097–1107.

[17] Brett Frischmann and Evan Selinger. 2018. *Re-Engineering Humanity.* Cambridge University Press. https://doi.org/10.1017/9781316544846

[18] Frobrukerrådet. 2018. Deceived by design: How tech companies use dark patterns to discourage us from exercising our rights to privacy.

[19] Xavier Gabaix and David Laibson. 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics* 121, 2 (2006), 505–540.

[20] 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* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 61–73. https://doi.org/10.1145/3357236.3395486

[21] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 534, 14 pages. https://doi.org/10.1145/3173574.3174108

[22] 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* (Vancouver, BC, Canada) *(DIS '14)*. ACM, New York, NY, USA, 523–532. https://doi.org/10.1145/2598510.2598541

[23] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring Price Discrimination and Steering on E-Commerce Web Sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. Association for Computing Machinery, New York, NY, USA, 305–318. https://doi.org/10.1145/2663716.2663744

[24] Jon D Hanson and Douglas A Kysar. 1999. Taking behavioralism seriously: The problem of market manipulation. *NYUL Rev.* 74 (1999), 630.

[25] Woodrow Hartzog. 2018. *Privacy's blueprint: the battle to control the design of new technologies.* Harvard University Press.

[26] Amy Hebert. 2019. Video games, loot boxes, and your money. https://www.consumer.ftc.gov/blog/2019/09/video-games-loot-boxes-and-your-money. Accessed September 9, 2020.

[27] Justin Hurwitz. 2020. Designing a Pattern, Darkly. *NCJL & Tech.* 22 (2020), 57.

[28] Eric J Johnson and Daniel Goldstein. 2003. Do defaults save lives?

[29] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'Dark Pattern' in Home Robots. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*. IEEE Press, Daegu, Republic of Korea, 374–381.

[30] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction.* Morgan Kaufmann, Boston.

[31] Chris Lewis. 2014. *Irresistible Apps: Motivational Design Patterns for Apps, Games, and Web-based Communities* (1st ed.). Apress, Berkely, CA, USA.

[32] Jamie Luguri and Lior Strahilevitz. 2020. Shining a Light on Dark Patterns. *Journal of Legal Analysis* 12 (2020), Forthcoming.

[33] Dominique Machuletz and Rainer Böhme. 2020. Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (2020), 481–498.

[34] Maximilian Maier and Rikard Harr. 2020. Dark Design Patterns: An End-User Perspective. *Human Technology* 16, 2 (2020), 170–199.

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

[36] Arunesh Mathur, Arvind Narayanan, and Marshini Chetty. 2018. Endorsements on Social Media: An Empirical Study of Affiliate Marketing Disclosures on YouTube and Pinterest. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 119 (Nov. 2018), 26 pages. https://doi.org/10.1145/3274388

[37] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. 2020. Dark Patterns: Past, Present, and Future. *Queue* 18, 2 (2020), 67–92.

[38] Robert Noggle. 1996. Manipulative actions: a conceptual and moral analysis. *American Philosophical Quarterly* 33, 1 (1996), 43–55.

[39] 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376321

[40] California Secretary of State. 2020. Qualified Statewide Ballot Measures. https://www.oag.ca.gov/system/files/initiatives/pdfs/19-0021A1%20%28Consumer%20Privacy%20-%20Version%203%29_1.pdf. Accessed September 9, 2020.

[41] Judith S. Olson and Wendy A. Kellogg. 2014. *Ways of Knowing in HCI.* Springer Publishing Company, Incorporated, USA.

[42] National Commission on Informatics and Liberty (CNIL). 2020. Shaping Choices in the Digital World. https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf. Accessed September 9, 2020.

[43] Robert Pitofsky. 1977. Beyond Nader: Consumer Protection and the Regulation of Advertising. *Harvard Law Review* 90, 4 (1977), 661–701. http://www.jstor.org/stable/1340281

[44] Joel Rudinow. 1978. Manipulation. *Ethics* 88, 4 (1978), 338–347. https://doi.org/10.1086/292086

[45] Natasha Dow Schüll. 2014. *Addiction by design: Machine gambling in Las Vegas.* Princeton University Press.

[46] Simon Shaw. 2019. Consumers Are Becoming Wise to Your Nudge. https://behavioralscientist.org/consumers-are-becoming-wise-to-your-nudge/. Accessed September 9, 2020.

[47] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. 2020. Circumvention by Design – Dark Patterns in Cookie Consent for Online News Outlets. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20).* Association for Computing Machinery, New York, NY, USA, Article 19, 12 pages. https://doi.org/10.1145/3419249.3420132

[48] Daniel J Solove. 2005. A taxonomy of privacy. *U. Pa. L. Rev.* 154 (2005), 477.

[49] Cass R. Sunstein. 2017. Nudges that fail. *Behavioural Public Policy* 1, 1 (2017), 4–25. https://doi.org/10.1017/bpp.2016.3

[50] Cass R. Sunstein. 2020. Sludge Audits. *Behavioural Public Policy* (2020), 1–20. https://doi.org/10.1017/bpp.2019.32

[51] Daniel Susser, Beate Roessler, and Helen Nissenbaum. 2020. Online Manipulation: Hidden Influences in a Digital World. *Georgetown Law Technology Review* 4 (2020), 45.

[52] John Sweller. 2018. Measuring cognitive load. *Perspectives on medical education* 7, 1 (2018), 1–2.

[53] Richard H. Thaler. 2018. Nudge, not sludge. *Science* 361, 6401 (2018), 431–431. https://science.sciencemag.org/content/361/6401/431.full.pdf

[54] Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness.* Penguin.

[55] Amos Tversky and Daniel Kahneman. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47, 2 (1979), 263–291.

[56] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)Informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19).* Association for Computing Machinery, New York, NY, USA, 973–990. https://doi.org/10.1145/3319535.3354212

[57] Ari Ezra Waldman. 2020. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology* 31 (2020), 105–109. https://doi.org/10.1016/j.copsyc.2019.08.025 Privacy and Disclosure, Online and in Social Interactions.

[58] Mark Warner and Debra Fischer. 2019. Senators Introduce Bipartisan Legislation to Ban Manipulative "Dark Patterns". https://www.fischer.senate.gov/public/index.cfm/2019/4/senators-introduce-bipartisan-legislation-to-ban-manipulative-dark-patterns. Accessed September 9, 2020.

[59] Fiona Westin and Sonia Chiasson. 2019. Opt out of Privacy or "Go Home": Understanding Reluctant Privacy Behaviours through the FoMO-Centric Design Paradigm. In *Proceedings of the New Security Paradigms Workshop* (San Carlos, Costa Rica) *(NSPW '19).* Association for Computing Machinery, New York, NY, USA, 57–67. https://doi.org/10.1145/3368860.3368865

[60] T. Martin Wilkinson. 2013. Nudging and manipulation. *Political Studies* 61, 2 (2013), 341–355.

[61] Allen Wood. 2014. Coercion, manipulation, exploitation. *Manipulation: Theory and practice* 1 (2014), 17–50.

[62] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark patterns in the design of games. In *Foundations of Digital Games 2013.* Society for the Advancement of the Science of Digital Games, Santa Cruz, CA, 8.