Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites

Draft: June 25, 2019

ARUNESH MATHUR, Princeton University, USA
GUNES ACAR, Princeton University, USA
MICHAEL FRIEDMAN, Princeton University, USA
ELENA LUCHERINI, Princeton University, USA
JONATHAN MAYER, Princeton University, USA
MARSHINI CHETTY, University of Chicago, USA
ARVIND NARAYANAN, Princeton University, USA

Dark patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions. We present automated techniques that enable experts to identify dark patterns on a large set of websites. Using these techniques, we study shopping websites, which often use dark patterns these to influence users into making more purchases or disclosing more information than they would otherwise. Analyzing ~53K product pages from ~11K shopping websites, we discover 1,841 dark pattern instances, together representing 15 types and 7 categories. We examine the underlying influence of these dark patterns, documenting their potential harm on user decision-making. We also examine these dark patterns for deceptive practices, and find 183 websites that engage in such practices. Finally, we uncover 22 third-party entities that offer dark patterns as a turnkey solution. Based on our findings, we make recommendations for stakeholders including researchers and regulators to study, mitigate, and minimize the use of these patterns.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; *HCI theory, concepts and models*; • **Social and professional topics** → **Consumer products policy**; • **Information systems** → *Browsers*.

Additional Key Words and Phrases: Dark Patterns; Consumer Protection; Deceptive Content

1 INTRODUCTION

Dark patterns [31, 47] are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might not make. Such interface design is an increasingly common occurrence on digital platforms including social media [45] and shopping websites [31], mobile apps [5, 30], and

Authors' addresses: Arunesh Mathur, Princeton University, 304 Sherrerd Hall, Princeton, NJ, 08544, USA, amathur@cs.princeton.edu; Gunes Acar, Princeton University, 320 Sherrerd Hall, Princeton, NJ, 08544, USA, gunes@princeton.edu; Michael Friedman, Princeton University, 35 Olden Street, Princeton, NJ, 08544, USA, mjf4@princeton.edu; Elena Lucherini, Princeton University, 312 Sherrerd Hall, Princeton, NJ, 08544, USA, elucherini@cs.princeton.edu; Jonathan Mayer, Princeton University, 307 Sherrerd Hall, Princeton, NJ, 08544, USA, jonathan.mayer@princeton.edu; Marshini Chetty, University of Chicago, Chicago, IL, 60637, USA, marshini@uchicago.edu; Arvind Narayanan, Princeton University, 308 Sherrerd Hall, Princeton, NJ, 08544, USA, arvindn@cs.princeton.edu.

XXXX-XXXX/2018/6-ART $15.00
https://doi.org/10.1145/1122445.1122456

video games [83]. At best, dark patterns annoy and frustrate users. At worst, dark patterns users can mislead and deceive users, e.g., by causing financial loss [1, 2], tricking users into giving up vast amounts of personal data [45], or inducing compulsive and addictive behavior in adults [71] and children [20].

While prior work [30, 31, 37, 47] has provided a starting point for describing the types of dark patterns, there is no large-scale evidence documenting the prevalence of dark patterns, or a systematic and descriptive investigation of how the various different types of dark patterns harm users. If we are to develop countermeasures against dark patterns, we first need to examine where, how often, and the technical means by which dark patterns appear, and second, we need to be able to compare and contrast how various dark patterns influence user decision-making. By doing so, we can both inform users about and protect them from such patterns, and given that many of these patterns are unlawful, aid regulatory agencies in addressing and mitigating their use.

In this paper, we present an automated approach that enables experts to identify dark patterns at scale on the web. Our approach relies on (1) a web crawler, built on top of OpenWPM [24, 39]—a web privacy measurement platform—to simulate a user browsing experience and identify user interface elements; (2) text clustering to extract recurring user interface designs from the resulting data; and (3) inspecting the resulting clusters for instances of dark patterns. We also develop a novel taxonomy of dark pattern characteristics so that researchers and regulators can use descriptive and comparative terminology to understand how dark patterns influence user decision-making.

While our automated approach generalizes, we focus this study on shopping websites. Dark patterns are especially common on shopping websites, used by an overwhelming majority of the American public [75], where they trick users into signing up for recurring subscriptions and making unwanted purchases, resulting in concrete financial loss. We use our web crawler to visit the ~11K most popular shopping websites worldwide, and from the resulting analysis create a large data set of dark patterns and document their prevalence. In doing so, we discover several new instances and variations of previously documented dark patterns [31, 47]. We also classify the dark patterns we encounter using our taxonomy of dark pattern characteristics.

We have five main findings:

- We discovered 1,841 instances of dark patterns on shopping websites, which together represent 15 types of dark patterns and 7 broad categories.
- These 1,841 dark patterns were present on 1,267 of the ~11K shopping websites (~11.2%) in our data set. Shopping websites that were more popular, according to Alexa rankings [9], were more likely to feature dark patterns. This represents a lower bound on the number of dark patterns on these websites, since our automated approach only examined text-based user interfaces on a sample of products pages per website.
- Using our taxonomy of dark pattern characteristics, we classified the dark patterns we discover on the basis whether they lead to an *asymmetry* of choice, are *covert* in their effect, are *deceptive* in nature, *hide information* from users, and *restrict* choice. We also map the dark patterns in our data set to the cognitive biases they exploit. These biases collectively describe the consumer psychology underpinnings of the dark patterns we identified.
- In total, we uncovered 234 instances of deceptive dark patterns across 183 websites. We highlight the types of dark patterns we discovered that rely on consumer deception.
- We identified 22 third-party entities that provide shopping websites with the ability to create dark patterns on their sites. Two of these entities openly advertised practices that enable deceptive messages.

Through this study, we make the following contributions:

- We measured the prevalence of dark patterns on 11K shopping websites. We will make this data set of dark patterns and our automated techniques available to researchers, journalists, and regulators to raise awareness of dark patterns, develop user-facing tools to combat these patterns, advance research in this space, and enable regulatory oversight [20].
- We contribute automated measurement techniques that enable expert analysts to discover new or revisit existing instances of dark patterns on the web. As part of this contribution, we make our web crawler and associated technical artifacts available on GitHub which can be used to conduct longitudinal measurements on shopping websites or be re-purposed for use on other types of websites (e.g., travel and ticket booking websites).
- We contribute a novel descriptive taxonomy of dark pattern characteristics that aids researchers, regulators, and policy-makers to understand and compare the underlying influence and harmful effects of dark patterns.
- We demonstrate that many instances of dark patterns are enabled by third-party entities, which provide shopping websites with scripts and plugins to easily implement these patterns on their websites. We identify these third parties and describe how they enable these patterns. This list of third-parties can be used by existing tracker and ad-blocking extensions (e.g., Ghostery[1], Adblock Plus[2]) to limit their use.

2 RELATED WORK

2.1 Online Shopping and Influencing User Behavior

Starting with Hanson and Kysar, numerous scholars have examined how companies abuse users' cognitive limitations and biases for profit, a practice they call market manipulation [49]. For instance, studies have shown that users make different decisions from the same information based on how it is framed [78, 79], giving readily accessible information greater weight [77], and becoming susceptible to impulsively changing their decision the longer the reward from their decision is delayed [27]. Some argue that because users are not always capable of acting in their own best interests, some forms of *paternalism*—a term referring to the regulation or curation of the user's options—may be acceptable [76]. However, determining the kinds of curation that are acceptable is less straightforward, particularly without documenting the practices that already exist.

More recently, Calo has argued that market manipulation is exacerbated by digital marketplaces since they posses capabilities that increase the chance of user harm culminating in financial loss, loss of privacy, and the ability to make independent decisions [33]. For example unlike brick-and-mortar stores, digital marketplaces can capture and retain user behavior information, design and mediate user interaction, and proactively reach out to users. Other studies have suggested that certain elements in shopping websites can influence impulse buying behaviors [59, 84]. For instance studies have shown how perceived urgency, perceived scarcity, social influence (e.g., *social proof*—informing users of others' behaviour—and shopping with others [32, 60]) can lead to higher spending. More recently, Moser et al. conducted a study [63] to measure the prevalence of elements that encourage impulse buying. They identified 64 such elements—such as product reviews/ratings, discounts, and quick add-to cart buttons—by manually scraping 200 shopping websites.

2.2 Dark Patterns in User Interface Design

Coined by Brignull in 2010, dark patterns is a catch-all term for how user interface design can be used to adversely influence users and their decision-making abilities online. Brignull described dark patterns as "tricks used in websites and apps that make you buy or sign up for things that you

[1] https://ghostery.com
[2] https://adblockplus.com

didn't mean to", and created a website that contained examples of these patterns from shopping and travel websites. The website documented patterns such as *Bait and Switch* (the user sets out to do one thing, but a different, undesirable thing happens instead), and *Confirmshaming* (using shame tactics to steer the user into making a choice).

2.2.1 Dark Pattern Taxonomies. A growing number of studies have expanded on Brignull's original taxonomy more systematically. Conti and Sobiesk [37] were the first to create a taxonomy of malicious interface design techniques, which they defined as interfaces that manipulate, exploit, or attack users. While their taxonomy contains no examples and details on how the authors created the taxonomy are limited, it contains several categories that overlap with Brignull's dark patterns including *Confusion* (asking the user questions or providing information that they do not understand) and *Obfuscation* (hiding desired information and interface elements). More recently, Bösch et al. [30] presented a similar, alternative breakdown of privacy-specific dark patterns as *Dark Strategies*, creating new patterns: *Forced Registration* (requiring account registration to access some functionality) and *Hidden Legalese Stipulations* (hiding malicious information in lengthy terms and conditions). Finally, Gray et al. [47] presented a broader categorization of Brignull's taxonomy and collapse many patterns into categories such as *Nagging* (repeatedly making the same request to the user) and *Obstruction* (preventing the user from accessing functionality).

While these taxonomies have focused on the web, researchers have also begun to examine dark patterns in specific application domains. For instance, Lewis [56] analyzed design patterns in the context of web and mobile applications and games, and codified those patterns that have been successful in making apps *irresistible*, such as *Pay To Skip* (in-app purchases that skip levels of a game). In another example, Greenberg et al. [48] analyzed dark patterns and *antipatterns*—interface designs with *unintentional* side-effects on user behavior—that leverage users' spatial relationship with digital devices. They introduced patterns such as *Captive Audience* (taking advantage of users' need to be in a particular location or do a particular activity to insert an unrelated interaction) and *Attention Grabber* (visual effects that compete for users' attention).

2.2.2 Dark Patterns And Cognitive Biases. A growing body of work has drawn explicit connections between the concept of dark patterns and theories of psychological motivation that underlie the behavioral biases that these patterns exploit. Xiao and Benbasat [82] proposed a theoretical model for how users are affected by deceptive marketing practices in online shopping, including affective mechanisms (psychological or emotional motivations) and cognitive mechanisms (perceptions about a product). In another example, Bösch et al. [30] used Kahneman's Dual process theory [77] which describes how humans have two modes of thinking—"System 1" (unconscious, automatic, possibly less rational) and "System 2" (conscious, rational)—and noted how *Dark Strategies* exploit users' System 1 thinking to get them to make a decision desired by the designer. Lastly, Lewis [56] linked each of the dark patterns described in his book to *Reiss's Desires*, a popular theory of psychological motivators [69]. Finally, a recent study by the Norwegian Consumer Council (*Frobrukerrådet*) [45] examining how interface designs on Google, Facebook, and Windows 10 make it hard for users to exercise privacy-friendly options; the study highlighted the defaults and framing that such enable dark patterns.

2.3 Comparison to Prior Work

Our study differs from these studies in three ways. First, we develop automation techniques that enable identifying new dark patterns and measuring the prevalence of these patterns. These techniques can be used by researchers, journalists, and regulators to conduct repeated measurements of dark patterns across the web. Automated measurements of this nature have traditionally been complicated because of challenges in collecting user interface data at scale as well as because of reasoning and analyzing interface design data.

Second, unlike previous studies (Section 2.2.1) which have either used anecdotal data [30, 31] or crowdsourcing-like approaches [37, 47] to ground their taxonomies, we conduct a large crawl of online shopping websites to discover new instances of dark patterns on shopping websites—nearly an order of magnitude larger than previous work [47]. Large-scale measurement of this kind has proven useful in discovering, documenting, and mitigating privacy and security issues on the web, including third-party tracking [24, 39], the adoption of HTTPS [44], and vulnerabilities of using remote third-party JavaScript libraries [66].

Finally, unlike previous studies, we develop a taxonomy of dark pattern characteristics which defines and characterize the problematic nature of each dark pattern by classifying them based how they influence users' decision-making. We describe our definition and classification in the following section.

3 A TAXONOMY OF DARK PATTERN CHARACTERISTICS

Previous studies have have made considerable efforts in recognizing and creating dark pattern taxonomies. However, while these studies have informed us about the presence and types of dark patterns, we still lack a descriptive and shared understanding of the characteristics of dark patterns that make them *dark* in the first place.

We developed a taxonomy of dark pattern characteristics that allows researchers, policy-makers and journalists to have a descriptive, comprehensive, and comparative terminology for understanding the potential harm and impact of dark patterns on user decision-making. Our taxonomy is based upon the literature on online manipulation [33, 74, 81] and dark patterns highlighted in previous work [31, 47], and it consists of the following five dimensions, each of which poses a possible barrier to user decision-making:

- **Asymmetric**: Does the user interface design impose unequal weights or burdens on the available choices presented to the user in the interface[3]? For instance, a website may present a prominent button to accept cookies on the web but hide the opt-out button in another page.
- **Covert**: Is the effect of the user interface design choice hidden from users? A website may develop interface design to steer users into making specific purchases without their knowledge. Often, websites achieve this by exploiting users' cognitive biases, which are deviations from rational behavior justified by some "biased" line of reasoning [50]. In a concrete example, a website may leverage the Decoy Effect [51] cognitive bias, in which an additional choice—the decoy—is introduced to make certain other choices seem more appealing. Users may fail to recognize the decoy's presence is merely to influence their decision making, making its effect hidden from users.
- **Deceptive**: Does the user interface design induce false beliefs either through affirmative misstatements, misleading statements, or omissions? For example, a website may offer a discount to users that appears to be limited-time, but actually repeats when they visit the site again. Users may be aware that the website is trying to offer them a deal or sale; however, they may not realize that the influence is grounded in a false belief—in this case, because the discount is recurring. This false belief affects users decision-making i.e., they may act differently if they knew that this sale is repeated.
- **Hides Information**: Does the user interface obscure or delay the presentation of necessary information to the user? For example, a website may not disclose, hide, or delay the presentation of information about charges related to a product from users.

[3] We narrow the scope of asymmetry to only refer to explicit choices in the interface.

- **Restrictive**: Does the user interface restrict the set of choices available to users? For instance, a website may only allow users to sign up for an account with existing social media accounts such as Facebook or Google so they can gather more information about them.

In Section 5, we also draw an explicit connection between each dark pattern we discover and the cognitive biases they exploit. The biases we refer to in our findings are:

(1) Anchoring Effect [77]: The tendency for individuals to overly rely on an initial piece of information—the "anchor"—on future decisions.
(2) Bandwagon Effect [72]: The tendency for individuals to value something more because others seem to value it.
(3) Default Effect [53]: The tendency of individuals to stick with options that are assigned to them by default, due to inertia in the effort required to change the option.
(4) Framing Effect [78]: A phenomenon that individuals may reach different decisions from the same information depending on how it is presented or "framed".
(5) Scarcity Bias [62]: The tendency of individuals to place a higher value on things that are scarce.
(6) Sunk Cost Fallacy [28]: The tendency for individuals to continue an action if they have invested resource (e.g., time and money) into it, even if that action would make them worse off.

4 METHOD

Dark patterns may manifest in several different ways on shopping websites, and can rely heavily upon interface manipulation, such as changing the hierarchy of elements or prioritizing certain options over others using different colors. However, many instances of dark patterns share common traits such as the text they display (e.g., in the "Confirmshaming" dark pattern, which tries to shame the user into making a particular choice, many messages begin with *No thanks*), or the features they enable (e.g., subscriptions and memberships). We reasoned that if we could extract all such textual interface elements on each shopping website, we could group and organize them—using unsupervised clustering—for an expert analyst to sift through.

We aligned this data collection process closely with how an ordinary user would browse and make purchases on shopping websites: discover pages containing products on a website, add these products to the cart, and check out. We describe these steps, and the data we collected during each visit to a website below. Figure 1 illustrates an overview of our method.

We note that only analyzing textual information in this manner restricts the set of dark patterns we can discover, making our findings a lower bound on the dark patterns employed by shopping websites. We leave detecting other kinds of dark patterns—those that are enabled using style, color, and other non-textual features—to future work, and we discuss possible approaches in Section 7.

4.1 Creating a Corpus of Shopping Websites

We used the following criteria to evaluate existing lists of popular shopping websites and, eventually, construct our own: (1) the list must consist of shopping websites in English so that we would have the means to analyze the data collected from the websites, and (2) the list must be representative of the most popular shopping websites globally.

We retrieved a list of popular websites worldwide from Alexa using the Top Sites API [9]. Alexa is a web traffic analysis company that ranks and categorizes websites based on statistics it collects from users of its toolbar. We used the Top Sites list because it is more stable and is based on monthly traffic and not daily rank, which fluctuates often [70] [4]. The list contained 361,102 websites in total ordered by popularity rank.

We evaluated two website classification services to extract shopping websites from this list of the most popular websites: (1) Alexa Web Information Service [10], and (2) WebShrinker [22]. We evaluated the classification accuracy of these services using a random sample of 500 websites from our list of 361K websites, which we manually labeled as "shopping" or "not shopping". We considered a website to be a shopping website if it was offering a product for purchase. Of the 500 websites in our sample, we labeled 57 as shopping and 443 as not shopping. We then evaluated the performance of both classifiers against this "ground truth."

Table 3 in the Appendix summarizes the classifiers' results. Compared to Webshrinker, Alexa's classifications performed poorly on our sample of websites (classification accuracy: 89% vs. 94%), with a strikingly high false negative rate (93% vs. 18%). Although Webshrinker had a slightly higher false positive rate (0.2% vs. 0.4%), we used methods to determine and remove these false positives as we describe in Section 4.2.1.

We subsequently used Webshrinker to classify our list of 361K websites, obtaining a list of 46,569 shopping websites. To filter out non-English websites, we downloaded home pages of each site using Selenium [8] and ran language detection on texts extracted from the pages using the `polyglot` Python library [4]. Our final data set contained 19,455 English language shopping websites. We created this filtered list in August 2018.

4.2 Data Collection with a Website Crawl

We conducted all our crawls from a large American university using two off-the-shelf computers, both equipped with 16G of memory and quad-core CPUs. For each shopping website, we decided to start with its product pages, since these are where users make decisions about purchases and are also most likely to contain dark patterns. Therefore, the first step in our website crawl was to determine ways to automatically identify product URLs from shopping websites.

4.2.1 Discovering Product URLs on Shopping Websites. To effectively extract product URLs from shopping websites, we iteratively designed and built a Selenium-based web crawler that contained a classifier capable of distinguishing product URLs from non-product URLs.

We first built a naïve depth-first crawler that on visiting a website's home page, determined the various URLs on the page, selected one URL at random, and then repeated this process from the selected URL. Using this crawler, we assembled a data set of several thousand URLs from visiting a random sample of 100 websites from our data set of 19K shopping websites. We manually labeled a sample of these URLs either as "product" or "non-product" URLs, and created a balanced data set containing 714 labeled URLs in total.

We trained a Logistic Regression classifier on this data set of labeled URLs using the `SGDClassifier` class from scikit-learn [68]. We extracted several relevant features from the URLs, including the length of a URL, the length of its path, the number of forward slashes and hyphens in its path, and whether its path contained the words "product" or "category." We used 90% of the URLs for training and obtained an 83% average classification accuracy using five-fold cross validation.

We embedded this classifier into our original Selenium-based web crawler to help guide its crawl. As a result, rather than selecting and visiting URLs at random, the crawler first used the classifier to rank the URLs on a page by likelihood of being product URLs, and then visited the URL with the highest likelihood. The crawler declared a URL as product if its page contained an "Add to cart" or

[4] We did not use Alexa's list of Top/Shopping websites [21] because of two issues. First, its criteria of categorization are not fully disclosed. Second, most of the websites in the list had an average monthly rank > 500,000, which we did not consider to be representative of the most popular websites worldwide.

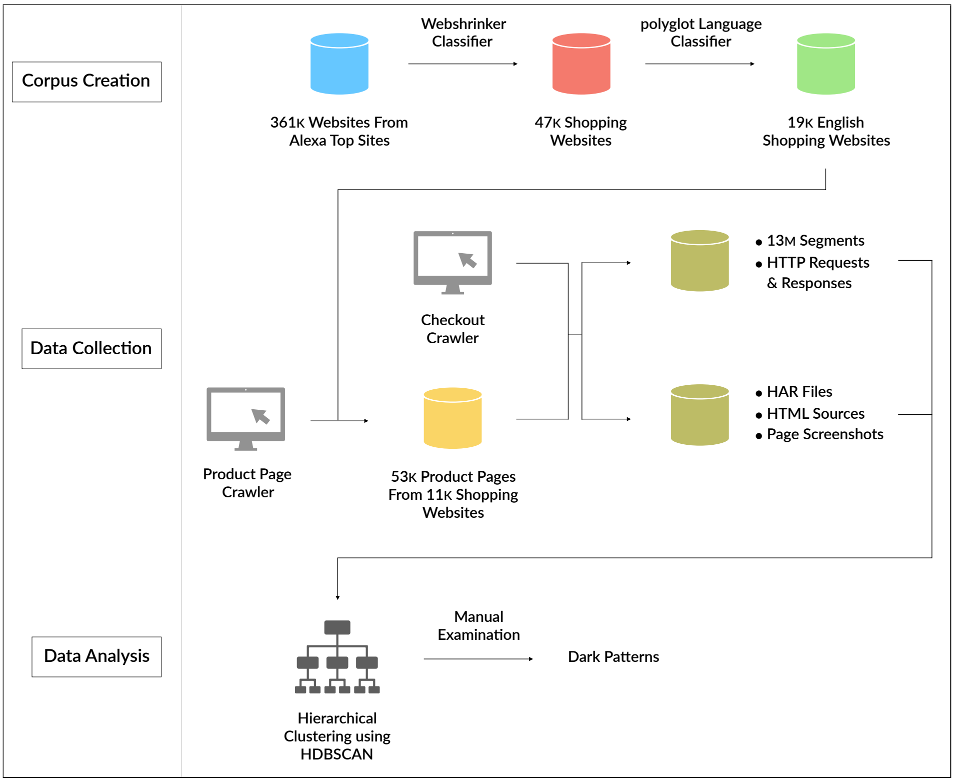


Fig. 1. Overview of the shopping website corpus creation, data collection using crawling, and data analysis using hierarchical clustering stages.

similar button. We detected this button by scoring visible HTML elements on a page by their size, color, and whether they matched certain regular expressions (e.g., “Add to (bag|cart|tote|. . .)”). This check also helped us weed out any false positives that may have resulted from the classification of shopping websites using Webshrinker (Section 4.1).

We tuned the crawler’s search process to keep its crawl tractable. The crawler returned to the home page after flagging a product URL. It did not visit a given URL more than two times to avoid exploring the same URLs, and it stopped after visiting 100 URLs or spending 15 minutes on a site. We determined these termination limits by running limited test crawls on random samples of shopping websites. Finally, we opted to extract no more than five product pages from each shopping website.

To evaluate our crawler’s performance, we randomly sampled 100 shopping websites from our corpus of 19K shopping websites and examined the product URLs the crawler returned for each of these websites. For 86 of those 100 websites, our crawler successfully extracted and returned legitimate product pages where they were present, and it returned no product pages where there were not any. For the remaining 14 websites, the crawler either timed out because the website was no longer reachable, the website included a step that the crawler could not handle (e.g., the website required selecting a country of origin), or the “Add to cart” button was incorrectly detected. We determined that the performance of the crawler was acceptable for the purpose of this study.

Therefore, we used the crawler on all of the 19K shopping websites, and in total we gathered 53,180 product pages from 11,286 shopping websites.

4.2.2 Simulating Product Purchase Flows. To simulate a user's typical shopping flow—which included selecting certain product options (e.g., size or color), adding the product to the cart, viewing the cart, and checking out—we designed and built an interactive "checkout crawler."

We based our checkout crawler on OpenWPM, a fully instrumented browser platform that is designed to conduct large-scale privacy and web-tracking measurement studies [39]. We extended OpenWPM in a number of ways to interact with the product pages we collected previously, including identifying various interface elements using scoring functions like the ones we described in Section 4.2. Each of these functions would output the most likely "Add to cart" buttons, "View cart" buttons, and "Checkout" buttons, which we would click in order. Because websites do not follow uniform HTML markup and design, our crawler needed to account for a variety of design alternatives and edge cases to simulate user interaction, such as dismissing popup dialogs.

We collected three types of data during this crawl for each product page. First, we saved the page source on visit. Second, we took screenshots each time the state of the page changed (e.g., clicking a button or selecting a product option). Third, we extended OpenWPM to store HTTP Archive (HAR) [13]) files for each crawled page. Unlike OpenWPM's existing HTTP instrumentation, HAR files are not limited to HTTP headers and contain full response contents that can be used for further analysis.

To evaluate our crawler's performance, we randomly sampled 100 product pages from the crawl in Section 4.2.1 and examined whether our crawler was able to simulate a user's shopping flow. In 66 of the 100 pages, our crawler reached the checkout page successfully. In 14 of the remaining 34, the crawler was able to add the product to cart but it was unable to proceed to the cart page; most often this was the result of complex product interaction (e.g., selecting the dimensions of a rug), which our crawler was not designed to perform. In the remaining 20 cases, either we produced Selenium exceptions, or failed to discover cart and checkout buttons.

We determined that the performance of our checkout crawler was acceptable for the purpose of this study. Therefore, we used the crawler on all of the 53K product pages. We divided the 53K product URLs into two equal-length lists to reduce the total crawling time. These crawls took approximately 90 hours to complete.

4.2.3 Capturing Meaningful Text Using Page Segmentation. To help discover dark patterns, the checkout crawler divided all the pages it visited into meaningful page segments, which can be thought of as "building blocks" of web pages representing meaningful smaller sections of a web page. These segments provided basic units for our data analysis and clustering.

We defined segments as *visible* HTML elements that contained no other block-level elements [6] and contained at least one text element—that is, elements of type `TEXT_NODE` [19]. However, since websites may use a virtually endless variety of markup and designs, we iteratively developed our segmentation algorithm, testing it on samples of shopping websites and accounting for possible edge cases. Algorithm 1 and Figure 11 in the Appendix detail the segmentation algorithm and illustrate its output for one web page, respectively.

To segment each web page, we waited for the page to load completely, also accounting for time needed for popup dialogs to appear. However, web pages may also display text from subsequent interactions, and with dynamically loaded content (e.g., a countdown timer). To capture possible segments from such updates to the web page during a crawl—no matter how minor or transient—we integrated the Mutation Summary [3] library into our checkout crawler. The Mutation Summary library combines DOM `MutationObserver` events [18] into compound event summaries that are easy to process. When the checkout crawler received a new Mutation Summary representing updates to the page, it segmented (Algorithm 1) this summary and stored the resulting segments.

For each segment, we stored its HTML Element type, its element text (via `innerText`), its dimensions and coordinates on the page, and its style including its text and background colors.

4.3 Data Analysis with Clustering

To discover dark patterns from the data set of segments, we employed hierarchical clustering. Our use of clustering was not to discover a set of latent constructs in the data but rather to organize the segments in a manner that would be conducive to scanning, making it easier for an expert analyst to sift through the clusters for possible dark patterns.

4.3.1 Data Preprocessing. Our crawls resulted in ~13 million segments across the 53K product URL pages. Many of these segments were duplicates, such as multiple "Add to Bag" segments across multiple websites. Since we only used text-based features for our analyses, we retained unique pieces of text across the websites in our data set (e.g., one segment containing the text "Add to Bag" across all the websites in our data set). We also replaced all numbers with a placeholder before performing this process to further reduce duplicates. This preprocessing reduced the ~13 million segments by 90% to ~1.3 million segments.

4.3.2 Feature Representations and Hierarchical Clustering. We created a variety of text representations and made several clustering passes through the data. Specifically, we used the Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) feature representations. In each case, we filtered all stop words (from Python NLTK [29]) and punctuation—except currency symbols, since these are indicative of product price—and only retained tokens that appeared in at least 100 segments. This resulted in a vocabulary of 10,133 tokens.

Given this large size of our vocabulary—and thus the dimensions of the segment-term matrix—we performed Principal Component Analysis (PCA) on both the BoW and TF-IDF matrices. We retained enough components from PCA to capture at least 95% of the variance in the data, resulting in 3 and 10 components, respectively.

To extract clusters from this data, we used the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm [34] implemented in the HDBSCAN Python library [14]. We chose HDBSCAN over other clustering algorithms since it is robust to noise in the data, and it allows us to vary the minimum size of the clusters (`min_cluster_size`). We made a total of eight passes at clustering: two inputs (BoW, TF-IDF) × two `min_cluster_size` values (5, 10) × two distance metrics (Manhattan distance, Euclidean distance). We picked sufficiently small values for the `min_cluster_size` parameter to keep the size of the noise cluster small and to avoid forcing segments into one cluster.

On examining the clustering output, we discovered that the clusters resulting from the TF-IDF input resulted in anywhere between 70%-75% of the segments being classified as noise. We believe this may have been because of the incorrect IDF scaling factor since the segments were not all drawn from a pool of independent observations—i.e., multiple segments originated from the same website; we subsequently ignored all clustering output resulting from the TF-IDF matrices. The clustering output across the BoW input did not vary too much. As expected, a `min_cluster_size` of 10 resulted in a larger noise cluster compared to a `min_cluster_size` of 5—but only marginally larger regardless of the distance metric. However, since the `min_cluster_size` of 10 produced significantly fewer clusters, we picked its output over the others. It contained 5,124 clusters.

4.3.3 Examining and Analyzing the Clusters. Once the clustering was complete, we made two passes through the data. The goal of pass one was to include clusters that contained any segments

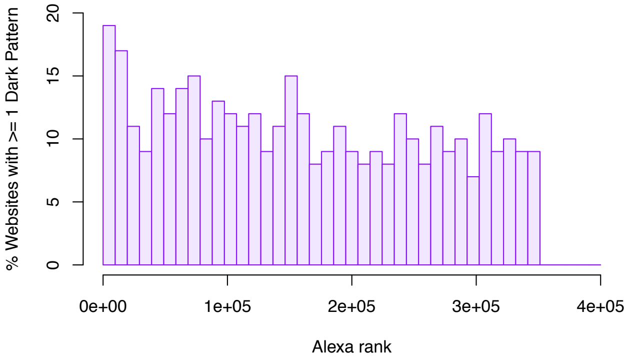


Fig. 2. Distribution of the dark patterns we discovered over the Alexa rank of the websites. Each bin indicates the percentage of shopping websites in that bin that contained at least one dark pattern.

that might manifest as dark patterns. In this pass, one researcher scanned the clusters and identified possible clusters of interest, recording all those clusters that represented specific types of user interfaces (e.g., login choices, cart totals), website characteristics (e.g., stock notifications), and product options (e.g., small/medium/large). This step filtered down the clusters from 5,124 to 1,768.

We then extracted all the websites that corresponded to these segments for further examination. The research team used the literature on dark patterns [31, 47] and impulse buying [63], and media coverage of high-pressure sales and marketing tactics (e.g., [15]) to create a shared understanding of possible dark patterns that could arise out of the data. Two researchers examined a sample of 200 of the 1,768 clusters, and recorded any dark patterns they encountered. The researchers also examined each website's set of screenshots and visited the websites to gain context and additional information surrounding the segments (e.g., discovering practices associated with the flagged pattern). To measure agreement between the researchers, we computed Cohen's kappa between the segments that were recorded—resulting in a score of 0.74. The team discussed and resolved all disagreements, and one researcher then examined the remaining clusters. The team then discussed the resulting dark patterns, and iteratively grouped them into types and broader categories.

4.4 Detecting Deceptive Dark Patterns

We further examined many of the dynamic dark patterns—those patterns that displayed transient values—for deceptive practices. To this end, we used our checkout crawler to "monitor" the websites containing dark patterns of interest once every four hours for a period of five days. We combined this data with several dark pattern-specific heuristics—which we describe in the following sections—to uncover instances of deceptive practices.

5 CATEGORIES OF DARK PATTERNS

Our analyses revealed 15 types of dark patterns representing 7 categories. Where applicable, we use the dark pattern labels proposed by Gray et al. [47] and Brignull [31] to describe these types and categories. In total, we discovered 1,841 instances of dark patterns from 1,267 (~11.2%) websites in our data set of 11K shopping websites. Given that our crawler explored the product page sections of websites and our analyses only took text-based user interfaces into account, this number represents a lower-bound estimate of the prevalence of dark patterns. Table 1 summarizes our findings and Figure 2 shows the distribution of the websites containing dark patterns over their Alexa ranks. The distribution suggests that dark patterns are more likely to appear on popular websites (Spearman's

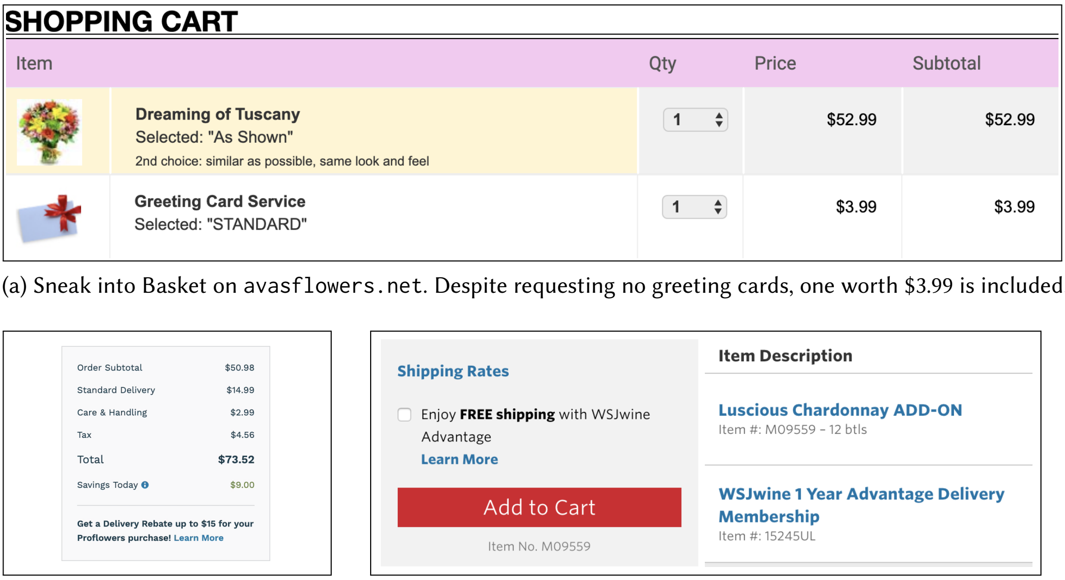
Table 1. Categories and types of dark patterns along with their description, prevalence, and definitions. Legend: ●= Always, ◐= Sometimes, ○= Never

Category	Type	Description	# Instances	# Websites	Asymmetric?	Covert?	Deceptive?	Hides Info.?	Restrictive?	Cognitive Biases
Sneaking	Sneak into Basket	Adding additional products to users' shopping carts without their consent	7	7	○	○	◐	●	○	Default Effect
	Hidden Costs	Revealing previously undisclosed charges to users right before they make a purchase	5	5	○	○	◐	●	○	Sunk Cost Fallacy
	Hidden Subscription	Charging users a recurring fee under the pretense of a one-time fee or a free trial	14	13	○	○	◐	●	○	None
Urgency	Countdown Timer	Indicating to users that a deal or discount will expire using a counting-down timer	393	361	○	◐	◐	○	○	Scarcity Bias
	Limited-time Message	Indicating to users that a deal or sale will expire will expire soon without specifying a deadline, thus creating uncertainty	88	84	○	◐	○	●	○	Scarcity Bias
Misdirection	Confirmshaming	Using language and emotion (shame) to steer users away from making a certain choice	169	164	●	○	○	○	○	Framing Effect
	Visual Interference	Using style and visual presentation to steer users to or away from certain choices	25	24	◐	●	◐	○	○	Anchoring & Framing Effect
	Trick Questions	Using confusing language to steer users into making certain choices	32	32	●	●	○	○	○	Default & Framing Effect
	Pressured Selling	Pre-selecting more expensive variations of a product, or pressuring the user to accept the more expensive variations of a product and related products	67	62	◐	◐	○	○	○	Anchoring & Default Effect, Scarcity Bias
Social Proof	Activity Message	Informing the user about the activity on the website (e.g., purchases, views, visits)	313	264	○	◐	◐	○	○	Bandwagon Effect
	Testimonials	Testimonials on a product page whose origin is unclear	12	12	○	○	◐	○	○	Bandwagon Effect
Scarcity	Low-stock Message	Indicating to users that limited quantities of a product are available, increasing its desirability	632	581	○	◐	◐	◐	○	Scarcity Bias
	High-demand Message	Indicating to users that a product is in high-demand and likely to sell out soon, increasing its desirability	47	43	○	◐	○	○	○	Scarcity Bias
Obstruction	Hard to Cancel	Making it easy for the user to sign up for a service but hard to cancel it	31	31	○	○	○	◐	●	None
Forced Action	Forced Enrollment	Coercing users to create accounts or share their information to complete their tasks	6	6	●	○	○	○	●	None

Rho = -0.62, $p < 0.0001$). In the following sections, we describes the various categories and types of dark patterns we discovered.

5.1 Sneaking

Coined by Gray et al. in their taxonomy [47], "Sneaking" refers to the category of dark patterns that attempt to misrepresent user actions, or disguise and delay information that if made available to users, they would likely object to. We observed three types of the Sneaking dark pattern: "Sneak into



SHOPPING CART

Item		Qty	Price	Subtotal
	Dreaming of Tuscany Selected: "As Shown" 2nd choice: similar as possible, same look and feel	1	$52.99	$52.99
	Greeting Card Service Selected: "STANDARD"	1	$3.99	$3.99

(a) Sneak into Basket on `avasflowers.net`. Despite requesting no greeting cards, one worth $3.99 is included.

Order Subtotal	$50.98
Standard Delivery	$14.99
Care & Handling	$2.99
Tax	$4.56
Total	**$73.52**
Savings Today	$9.00

Get a Delivery Rebate up to $15 for your Proflowers purchase! Learn More

Shipping Rates

□ Enjoy **FREE shipping** with WSJwine Advantage
Learn More

Item No. M09559

Item Description

Luscious Chardonnay ADD-ON
Item #: M09559 - 12 btls

WSJwine 1 Year Advantage Delivery Membership
Item #: 15245UL

(b) Hidden Costs on `proflowers.com`. The Care & Handling charge ($2.99) is disclosed on the last step.

(c) Hidden Subscription on `wsjwine.com`. Left: The website fails to disclose that the *Advantage* service is an annual subscription unless the user clicks on *Learn More*. Right: The *Advantage* service in cart.

Fig. 3. Three types of the Sneaking category of dark patterns.

Basket" [31], "Hidden Costs" [31], and "Hidden Subscription" (Brignull's "Forced Continuity" [31]) on 23 shopping websites. Figure 3 highlights instances of these three types.

5.1.1 *Sneak into Basket.*

The "Sneak into Basket" dark pattern adds additional products to users' shopping carts without their consent. Through its design, Sneak into Basket exploits the default effect cognitive bias in users, hoping that users will stick with the products it adds to cart, often promoting the added products as "bonuses" and "necessary". In one instance of Sneak into Basket as shown in Figure 3a, adding a bouquet of flowers to the shopping cart on `avasflowers.net` also adds a greeting card. In another instance on `laptopoutlet.co.uk`, adding an electronic product—such as a laptop—to the shopping cart also adds product insurance. Other websites such as `cellularoutfitter.com` added additional products (e.g., a USB charger) to the shopping cart using pre-selected checkboxes. While such checkboxes could be unselected by a vigilant user, the additional products would be added by default in the absence of any such intervention. In total, we found 7 instances of the Sneak into Basket dark pattern.

Based on our taxonomy of dark pattern characteristics, we classify Sneak into Basket as at least partially *deceptive* (it incorrectly represents the nature of the action of adding an item to the shopping cart) and *information hiding* (it deliberately disguises how the additional products were added to cart from users) in nature. However, it is not *covert*: users can visibly see and realize that additional products have been added to their shopping carts.

5.1.2 *Hidden Costs.*

The "Hidden Costs" dark pattern reveals new, additional, and often unusually high charges to users just before they are about to complete a purchase. Examples of such charges include "service fees" or "handling" costs. Often these charges are only revealed at the end of a checkout process, after the user has already filled out shipping/billing information, and consenting to terms of use. Through its design, the Hidden Costs dark pattern exploits the sunk cost fallacy cognitive bias: users are likely to feel invested in the process that they justify their effort by completing the purchase despite the additional charges. Figure 3b shows the Hidden Costs dark pattern on `proflowers.com`, where the "Care & Handling" charge of $2.99 is revealed immediately before confirming the order. In total, we found 5 instances of the Hidden Costs dark pattern.

Based on our taxonomy of dark pattern characteristics, we classify Hidden Costs as at least partially *deceptive* (it relies on minimizing and delaying information from users), and thus also *information hiding* in nature. Like Sneak into Basket, Hidden Costs is not *covert*: users can visibly see and realize that the website included additional charges.

5.1.3 Hidden Subscription. The "Hidden Subscription" dark pattern charges users a recurring fee under the pretense of a one-time fee or a free trial. Often, if at all, users become aware of the recurring fee once they are charged several days or months after their purchase. For example, we discovered that `wsjwine.com` offers users an *Advantage* service which appears to be a one-time payment of $89 but renews annually, as shown in Figure 3c. Further, Hidden Subscription often appears with the "Hard to Cancel" dark pattern—which we describe in Section 5.6—thereby making the recurring charges harder to cancel than signing up for them. In total, we found 14 instances of Hidden Subscription dark pattern.

Based on our taxonomy of dark pattern characteristics, we classify Hidden Subscription as at least partially *deceptive* (it misleads users about the nature of the initial offer) and *information hiding* (it withholds information about the recurring fees from users) in nature.

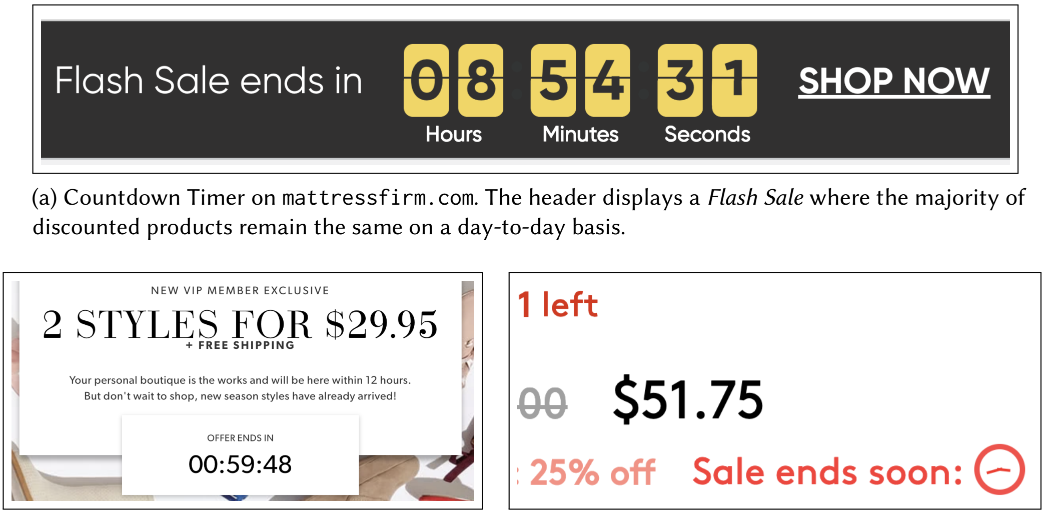
5.2 Urgency

"Urgency" refers to the category of dark patterns that impose a deadline on a sale or deal, thereby accelerating user decision-making and purchases [26, 36, 52]. Urgency dark patterns exploit the scarcity bias in users—making sales and deals more desirable than they would otherwise be, and signaling that inaction would result in losing out on potential gains from purchases. These dark patterns create a potent "fear of missing out" effect particularly when combined with the "Scarcity" (Section 5.5) and "Social Proof" (Section 5.4) dark patterns.

We observed two types of the Urgency dark pattern: "Countdown Timers" and "Limited-time Messages" on 437 shopping websites across their product, cart, and checkout pages. In product pages, these indicated deadlines about site-wide sales and coupons, sales on specific products, or shipping deadlines; in cart pages, they indicated deadlines about product reservation (e.g., "Your cart will expire in 10:00 minutes, please check out now") and coupons, urging users to complete their purchase. Figure 4 highlights instances of these two types.

5.2.1 Countdown Timers. The "Countdown Timer" dark pattern is a dynamic indicator of a deadline, counting down until the deadline expires. Figure 4a and Figure 4b show the Countdown Timer dark pattern on `mattressfirm.com` and `justfab.com` respectively, one indicating the deadline for a recurring *Flash Sale* and the other a *Member Exclusive*. In total, we found 393 instances of the Countdown Timer dark pattern.

Deceptive Countdown Timers. Using the visit-and-record method described in Section 4.4, we examined the countdown timers in our data set for deceptive practices. We stitched the screenshots of each countdown timer from the repeated visits of our crawler to a website into a video, and viewed the resulting videos to observe the behavior of the timers. We considered a countdown timer deceptive if (1) the timer reset after timeout with the same sale or offer still valid, or (2) the timer expired but the offer it claimed was expiring was still valid even following expiration.



Flash Sale ends in 08 54 31 SHOP NOW

Hours Minutes Seconds

(a) Countdown Timer on `mattressfirm.com`. The header displays a *Flash Sale* where the majority of discounted products remain the same on a day-to-day basis.

NEW VIP MEMBER EXCLUSIVE

2 STYLES FOR $29.95

+ FREE SHIPPING

Your personal boutique is the works and will be here within 12 hours.
But don't wait to shop, new season styles have already arrived!

OFFER ENDS IN

00:59:48

1 left

~~00~~ $51.75

25% off Sale ends soon:

(b) Countdown Timer on `justfab.com`. The offer is available even after the timer expires.

(c) Limited-time Message on chicwish.com. The sale claims to end soon without stating a deadline.

Fig. 4. Two types of the Urgency category of dark patterns.

In total, we discovered 157 instances of deceptive Countdown Timers on 140 shopping websites. One such example is shown in Figure 4b on `justfab.com`, where the advertised offer remains valid even after the countdown timer of 60 minutes expires.

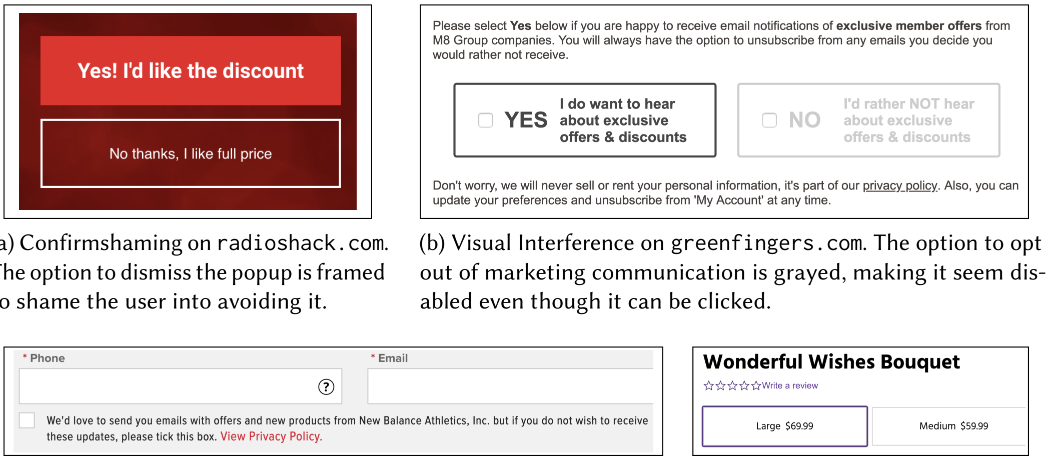
Based on our taxonomy of dark pattern characteristics, we classify Countdown Timers as partially *covert* (it creates a heightened sense of immediacy, unbeknownst to at least some users), and sometimes *deceptive* (it can mislead users into believing a sale or offer is expiring when in is reality it is not, creating false urgency) in nature.

5.2.2 Limited-time Messages. Unlike Countdown Timers, the "Limited-time Message" dark pattern is a static urgency message without an accompanying deadline. By not stating the deadline, websites withhold information from users, which results in uncertainty and increased urgency, further depriving users of the possibility of a delayed purchase. Figure 4c shows an instance of the Limited-time Message dark pattern on `chicwish.com`, where the advertised sale is stated to end *soon* without any mention of the end date. For every such instance we discovered, we verified that the shopping website made no disclosure about the accompanying deadline (e.g., in the fine print and the terms of sale pages). In total, we found 88 instances of the Limited-time Message dark pattern.

Based on our taxonomy of dark pattern characteristics, we classify Limited-time Messages as at least partially *covert* similar to Countdown Timers, and *information hiding* (unlike Countdown Timers they do not reveal the deadline in their offers) in nature.

5.3 Misdirection

The "Misdirection" category of dark patterns uses visuals, language, and emotion to steer users toward or away from making a particular choice. Misdirection functions by exploiting different affective mechanisms and cognitive biases in users without ever restricting the set of choices available to users. Our version of the Misdirection dark pattern is inspired by Brignull's original Misdirection dark pattern [31]. However, while Brignull considered Misdirection to occur exclusively using



Please select **Yes** below if you are happy to receive email notifications of **exclusive member offers** from M8 Group companies. You will always have the option to unsubscribe from any emails you decide you would rather not receive.

□ **YES** **I do want to hear about exclusive offers & discounts**

□ NO I'd rather NOT hear about exclusive offers & discounts

Don't worry, we will never sell or rent your personal information, it's part of our privacy policy. Also, you can update your preferences and unsubscribe from 'My Account' at any time.

□ We'd love to send you emails with offers and new products from New Balance Athletics, Inc. but if you do not wish to receive these updates, please tick this box. View Privacy Policy.

(a) Confirmshaming on `radioshack.com`. The option to dismiss the popup is framed to shame the user into avoiding it.

(b) Visual Interference on `greenfingers.com`. The option to opt out of marketing communication is grayed, making it seem disabled even though it can be clicked.

(c) Trick Questions on `newbalance.co.uk`. Opting out of marketing communication requires ticking the checkbox.

(d) Pressured Selling on `1800flowers.com`. The most expensive product is default.

Fig. 5. Four types of the Misdirection category of dark patterns.

stylistic and visual manipulation, we take a broader view of the term, also including Misdirection caused by language and emotional manipulation.

We observed four types of the Misdirection dark pattern: "Confirmshaming" [31], "Trick Questions" [31], "Visual Interference" [47], and "Pressured Selling" on 293 shopping websites. While the first three have been documented [31, 47], we discovered Pressured Selling from our analyses. Figure 5 highlights instances of these four types.

5.3.1 Confirmshaming. Coined by Brignull [31], the "Confirmshaming" dark pattern uses language and emotion to steer users away from making a certain choice. Confirmshaming appeared most often in popup dialogs that solicited users' email addresses in exchange for a discount, where the option to decline the offer—which the website did not want users to select—was framed as a shameful choice. Examples of such framing included "No thanks, I like full paying price", "No thanks, I hate saving money", "No thanks, I hate fun & games", or similar. Through its design, the Confirmshaming dark pattern exploits both the framing effect cognitive bias in users and shame, a powerful behavior change agent [57]. Figure 5a shows one instance of the Confirmshaming dark pattern on `radioshack.com`. In total, we found 156 such instances.

Based on our taxonomy of dark pattern characteristics, we classify Confirmshaming as *asymmetric* (the opt-out choice shames users into avoiding it) in nature. However, Confirmshaming is not *covert*: users can visibly see and realize that the design is attempting to influence their choice.

5.3.2 Visual Interference. The "Visual Interference" dark pattern uses style and visual presentation to steer users into making certain choices over others (Brignull's original description of Misdirection [31]). Although we excluded style information in our clustering analysis, we extracted these patterns as a consequence of examining the text they displayed. In some instances, websites used the Visual Interference dark pattern to make certain courses of action more prominent over others. For example, the subscription offering on `exposedskincare.com` is stylistically more prominent and emphasized than the non-subscription offering. In other instances, websites used visual effects on textual descriptions to inflate the discounts available for products. For example, websites such as dyson.co.uk and justfab.com offered free gifts to users, and then used these gifts to inflate the savings on users' purchases in the checkout page—even when the originally selected product was not on discount. In one instance on greenfingers.com, we discovered that the option to decline marketing communication is greyed out, creating an illusion that the option is unavailable or disabled even though it can be clicked, as shown in Figure 5b. In total, we found 25 instances of the Visual Interference dark pattern.

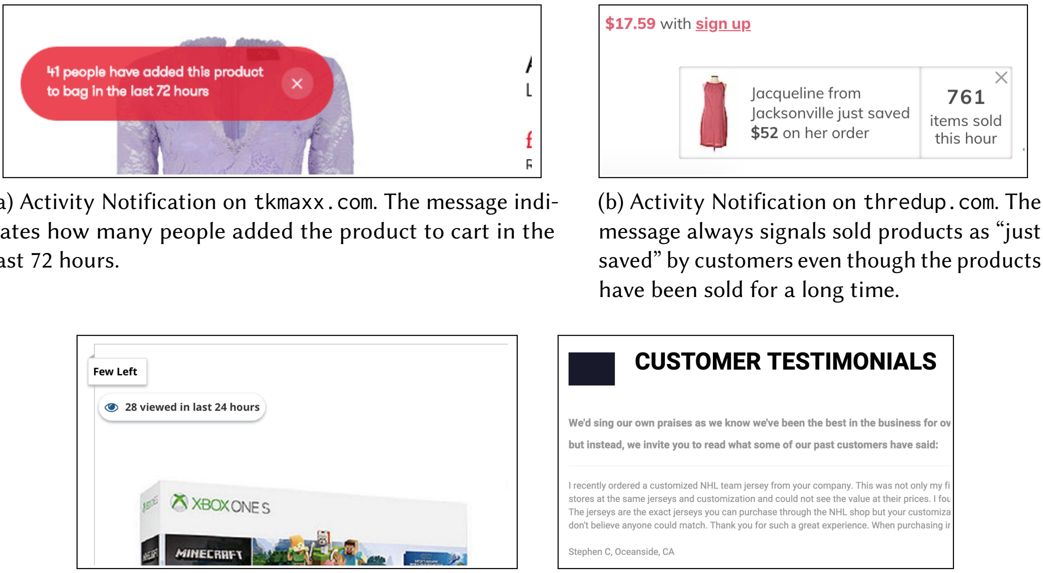
Based on our taxonomy of dark pattern characteristics, we classify Visual Interference as partially *asymmetric* (in some instances it creates unequal choices, steering users into one choice over the other), *covert* (users may not realize the effect the visual presentation has had on their choice), and sometimes *deceptive* (e.g., a website presents users with a "lucky draw" from a list of potential deals but the draw process is deterministic unknown to the user) in nature.

5.3.3 Trick Questions. Also originating from Brignull's taxonomy [31], the "Trick Questions" dark pattern uses confusing language to steer users into making certain choices. Like Confirmshaming, Trick Questions attempts to overcome users' propensity to opt out of marketing and promotional messages by subtly inverting the entire opt-out process. Most often, websites achieved this effect by introducing confusing double negatives (e.g., "Uncheck the box if you prefer not to receive email updates"), or by using negatives to alter expected courses of action such as unchecking a box to opt out (e.g., "If you do not wish to be contacted via email, please ensure that the box is not checked"). Through its design, Trick Questions exploits the default and framing effect cognitive biases in users, who become more susceptible to a choice they erroneously believe is aligned with their preferences. Figure 5c shows one instance of Trick Questions on newbalance.co.uk. In total, we found 32 such instances, occurring most often during the checkout process when collecting user information to complete purchases.

Based on our taxonomy of dark pattern characteristics, we classify Trick Questions as *asymmetric* (opting out is more burdensome than opting in) and *covert* (users fail to understand the effect of their choice as a consequence of the confusing language) in nature.

5.3.4 Pressured Selling. We discovered a new type of dark pattern that does appear in the taxonomies of Gray [47] and Brignull [31]. We term this the "Pressured Selling" dark pattern, and use it to refer to defaults or often high-pressure tactics that steer users into purchasing a more expensive version of a product (upselling) or into purchasing related products (cross-selling). The Pressured Selling dark pattern exploits a variety of different cognitive biases, such as the default effect, the anchoring effect, and the scarcity bias to drive user purchasing behavior. Figure 5d shows one such instance on 1800flowers.com, where the largest flower bouquet is selected by default. By selecting the most expensive option, the dark pattern makes it the point of comparison—an "anchor"—and thus increases the the probability of users avoiding the least expensive option [67]. In another instance, on fashionworld.co.uk, the website opened popup dialogs the user had to explicitly decline immediately after adding a product to cart. These dialogs urged users to buy more "Hot sellers", "Deals", and "Bundled" products. In total, we found 67 instances of the Pressured Selling dark pattern.

Based on our taxonomy of dark pattern characteristics, we classify Pressured Selling as partially *asymmetric* (it pushes users towards accepting more expensive product options) and at least partially *covert* (users fail to realize that they have purchased a more expensive product than they would have had they been defaulted with the least expensive product to begin with) in nature.



41 people have added this product to bag in the last 72 hours

$17.59 with sign up

Jacqueline from Jacksonville just saved $52 on her order

761 items sold this hour

Few Left

28 viewed in last 24 hours

CUSTOMER TESTIMONIALS

We'd sing our own praises as we know we've been the best in the business for ov

but instead, we invite you to read what some of our past customers have said:

I recently ordered a customized NHL team jersey from your company. This was not only my fi stores at the same jerseys and customization and could not see the value at their prices. I fou The jerseys are the exact jerseys you can purchase through the NHL shop but your customiza don't believe anyone could match. Thank you for such a great experience. When purchasing ir

Stephen C, Oceanside, CA

(a) Activity Notification on tkmaxx.com. The message indicates how many people added the product to cart in the last 72 hours.

(b) Activity Notification on thredup.com. The message always signals sold products as "just saved" by customers even though the products have been sold for a long time.

(c) Activity Notification on jcpenney.com. The message indicates the number of people who viewed the product in the 24 hours along with the quantity left in stock.

(d) Testimonials of Uncertain Origin on coolhockey.com. We found the same testimonials on ealerjerseys.com with different customer names.

Fig. 6. Two types of the Social Proof category of dark patterns.

5.4 Social Proof

According to the social proof principle, individuals determine the correct action and behavior for themselves in a given situation by examining the action and behavior of others [36]. The "Social Proof" dark pattern uses this influence to accelerate user decision-making and purchases, exploiting the bandwagon effect cognitive bias to its advantage. In fact studies have shown that individuals are more likely to impulse buy more when shopping with their peers, families and others [60].

We observed two types of the Social Proof dark pattern: "Activity Notifications" and "Testimonials of Uncertain Origin" on 275 websites across their product and cart pages. In all these instances, the Social Proof messages indicated other users' activities and experiences shopping for products and items. Figure 6 highlights instances of these two types.

5.4.1 Activity Notifications. The "Activity Notification" dark pattern is a transient, often recurring and attention grabbing message that appears on product pages indicating the activity of other users. We could group these into different categories: dynamic and periodic messages that indicated other users just bought a product (e.g., "Abigail from Michigan just bought a new stereo system"); static or dynamic text to indicate how many users have a specific item in their cart (e.g., "35 people added this item to cart"); and similar text to indicate how many users have viewed a product (e.g., "90 people have viewed this product"). Figure 6a, Figure 6b, and Figure 6c highlight three instances of Activity Notification on tkmaxx.com, thredup.com, and jcpenney.com respectively. In total, we found 313 such instances.

Deceptive Activity Notifications. We examined the Activity Notification messages in our data set for deceptive practices. To facilitate our analysis, we manually inspected the page source of each shopping website that displayed these notifications to verify their integrity. We ignored all those notifications that were generated server-side since we had limited insight into how and whether they were truly deceptive. We considered an instance of Activity Notification to be deceptive if the content it displayed—including any names, locations statistics, counts—was falsely generated or made misleading statements.

In total, we discovered 29 instances of deceptive Activity Notifications on 20 shopping websites. The majority of these websites generated their deceptive notifications in a random fashion (e.g., choosing a random number of users who are "currently viewing" a product) and others hard-coded generated the notifications, meaning they never changed. One notable case was `thredup.com` as shown in Figure 6b, where the website generated messages based on fictitious names and locations for an unvarying list of products that was always indicated to be "just sold".

Based on our taxonomy of dark pattern characteristics, we classify Activity Notifications as partially *covert* (in instances where the notifications are site-wide, users may fail to understand their purpose) and sometimes deceptive *the content of notifications can be deceptively generated or misleading* in nature.

5.4.2 Testimonials of Uncertain Origin. The "Testimonials of Uncertain Origin" dark pattern refers to the use of customer testimonials whose origin or how they were sourced and created is not clearly specified. For each instance of this dark pattern, we made two attempts to validate its origin. First, we inspected the website to check if it contained a form to submit them. Second, we performed exact searches of the testimonials on a search engine (`google.com`) to check if they appeared on other websites. Figure 6d shows one instance on `coolhockey.com`, where we found the same set of testimonials on `ealerjerseys.com` with different customer names attached to them. In total, we found 12 instances of this pattern.

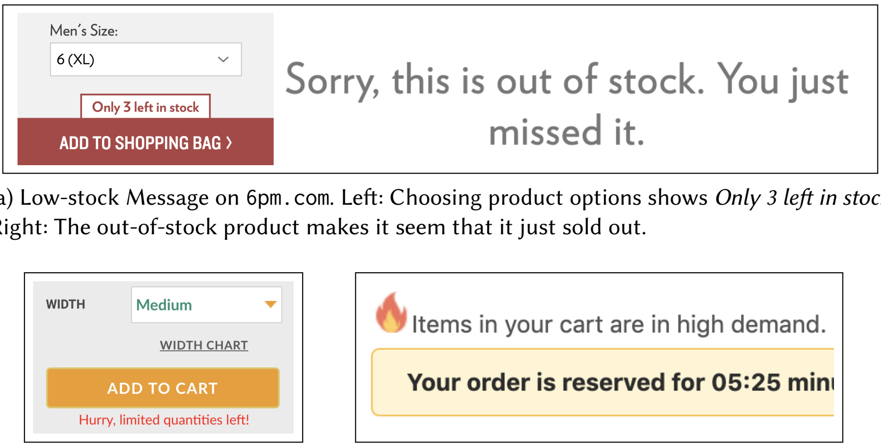
5.5 Scarcity

"Scarcity" refers to the category of dark patterns that signal the limited availability or high demand of a good, thus increasing its perceived value and desirability [54, 61]. We observed two types of the Scarcity dark pattern: "Low-stock Messages" and "High-demand Messages" on 609 shopping websites across their product and cart pages. In both pages, they indicated the limited availability of a product or that a product was in high demand and thus likely to become unavailable soon. Figure 7 highlights instances of these two types.

5.5.1 Low-stock Messages. The "Low-stock Message" dark pattern signals to users about limited quantities of a product. Figure 7a shows one such instance of this pattern on `6pm.com`, displaying the precise quantity in stock. In total, we found 632 instances of the Low-stock Message dark pattern. However, not all of these instances displayed stock quantities. 49 of these instances only indicated that stock was limited or low, without displaying the exact quantity, resulting in uncertainty and increased impulsive behavior. Figure 7b shows one such instance on `orthofeet.com`.

Deceptive Low-stock Messages. We examined all the Low-stock Message dark patterns for deceptive practices using the method described in Section 4.4. From the resulting data, we ignored those websites whose stock amounts remained the same between visits, reasoning that those are unlikely to be indicative of deceptive practices. We then manually examined the remaining sites and identified how the stock information was generated.

In total, we discovered 17 instances of deceptive Low-stock Messages on 17 shopping websites. On further examination, we observed that 16 of these sites decremented stock amounts in a recurring, deterministic pattern according to a schedule, and the one remaining site (`forwardrevive.com`) randomly generated stock values on page load. Exactly 8 of these sites used third-party JavaScript libraries to generate the stock values, such as Hurrify [17] and Booster [11]. Both of these are



Men's Size: 6 (XL)

Only 3 left in stock

Sorry, this is out of stock. You just missed it.

(a) Low-stock Message on `6pm.com`. Left: Choosing product options shows *Only 3 left in stock*. Right: The out-of-stock product makes it seem that it just sold out.

WIDTH Medium

Hurry, limited quantities left!

Items in your cart are in high demand.

Your order is reserved for 05:25 mini

(b) Low-stock on `orthofeet.com`. Appears for all products.

(c) High-demand Message on `fashionnova.com`. The message appears for all products in the cart.

Fig. 7. Two types of the Scarcity category of dark patterns.

popular plugins for Shopify—one of the largest e-Commerce companies—based websites. The remaining websites injected stock amounts through first-party JavaScript or HTML.

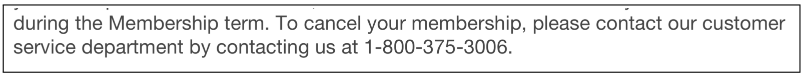
Besides the use—or non-use—of numeric data and deception, Low-stock Messages can be problematic in other ways. For example, we observed that several websites such as `6pm.com` and `orthofeet.com` displayed Low-stock indicators for nearly all their products—stating "Only X left" and "Hurry, limited quantities left!" respectively. The former, in particular, showed a "Sorry, this is out of stock. You just missed it" popup dialog for every product that was sold out, even if it had already been out of stock previously.

Based on our taxonomy of dark pattern characteristics, we classify Low-stock Messages as partially *covert* (it creates a heightened sense of impulse buying, unbeknownst to at least some users), sometimes *deceptive* (it can mislead users into believing a product is low on stock when in reality it is not, creating false scarcity), and partially *information hiding* (in some instances, it does not explicitly specify the stock quantities at hand in some instances) in nature.

5.5.2 High-demand Messages. The "High-demand Message" dark pattern signals to users that a product is in high demand, implying that it is likely to sell out soon. Figure 7c shows one such instance on `fashionnova.com` on the cart page, indicating that the products in cart are selling out fast. In total, we found 47 such instances of the High-demand dark pattern; 38 of these instances appeared consistently, regardless of the product displayed on the website, or regardless of the items in cart. Based on our taxonomy of dark pattern characteristics, we classify High-demand Messages as partially *covert* like Low-stock Messages.

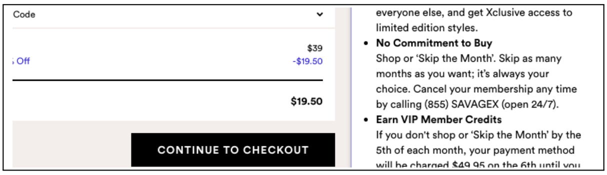
5.6 Obstruction

"Obstruction"—coined by Gray et al. [47]—refers to the category of dark patterns that make a certain action harder than it should be in order to dissuade users from taking that action. We observed one type of the Obstruction dark pattern: "Hard to Cancel"—a pattern similar to Brignull's *Roach*



during the Membership term. To cancel your membership, please contact our customer service department by contacting us at 1-800-375-3006.

(a) Hard to Cancel on sportsmanguide.com. The website only discloses in the terms PDF file that canceling the recurring service requires calling customer service.



everyone else, and get Xclusive access to limited edition styles.

- **No Commitment to Buy**
  Shop or 'Skip the Month'. Skip as many months as you want; it's always your choice. Cancel your membership any time by calling (855) SAVAGEX (open 24/7).
- **Earn VIP Member Credits**
  If you don't shop or 'Skip the Month' by the 5th of each month, your payment method will be charged $49.95 on the 6th until you

(b) Hard to Cancel on savagex.com. The website discloses upfront that the recurring service can only be canceled by calling customer care.

Fig. 8. The Hard to Cancel type from the Obstruction category of dark patterns.

Motel dark pattern [31]—on 31 websites, which made it easy for users to sign up for recurring subscriptions and memberships but hard for them to subsequently cancel and leave those.

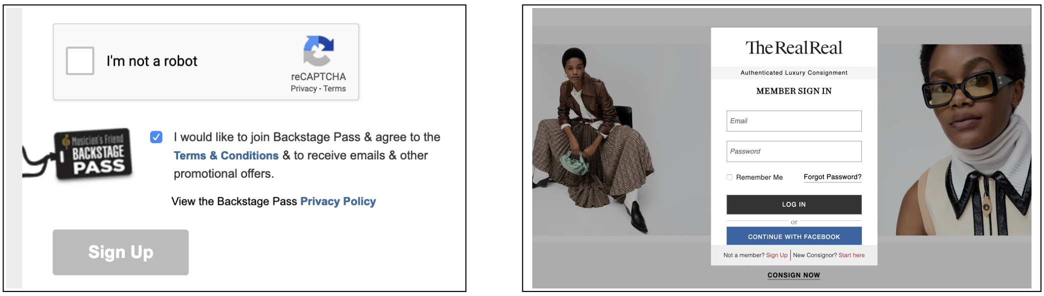
More often than not, shopping websites did not disclose upfront to users that canceling the subscription or membership could not be completed in the same manner they signed up for the memberships in the first place. For example, as shown in Figure 8a, sportsmansguide.com promotes a "buyer's club" discount membership price and makes it easy for users to sign up for this annual membership under the impression of "Cancel anytime." However, their terms of service reveal that the membership can only be cancelled by calling their customer service. In rare instances, as shown in Figure 8b, websites such as savagex.com disclosed upfront that cancellation required calling customer service.

The Hard to Cancel dark pattern is *restrictive* (it limits the choices users can exercise to cancel their services). In cases where websites do not disclose their cancellation policies upfront, Hard to Cancel also becomes *information hiding* (it fails to inform users about how cancellation is harder than signing up).

5.7 Forced Action

"Forced Action" refers to the category of dark patterns—originally proposed by Gray et al. [47]—that require users to take certain additional and tangential actions that in order to complete their tasks. We observed one type of the Forced Action dark pattern: "Forced Enrollment" on 6 websites, explicitly coercing users into signing up for marketing communication, or creating accounts to surrender their information in these instances. By using the Forced Enrollment dark pattern, online services and websites collected more information about their users than they might otherwise consent to—resulting from an all-or-nothing proposition.

On four websites, the Forced Enrollment dark pattern manifested as a checkbox in the user interface, requiring users to simultaneously consent to the terms of service *and* to receiving marketing emails as part of the consent process. Figure 9a shows one such instance on musiciansfriend.com. In another instance of the Forced Enrollment on therealreal.com—as shown in Figure 9b—the website displayed a popup dialog that prevented users from viewing product offerings on the website without creating an account—even if users decide against eventually making a purchase.



☑ I would like to join Backstage Pass & agree to the Terms & Conditions & to receive emails & other promotional offers.

View the Backstage Pass Privacy Policy

(a) Forced Enrollment on musiciansfriend.com. Agreeing to the terms of use also requires agreeing to receive emails and promotions.

(b) Forced Enrollment on therealreal.com. Browsing the website requires creating an account even if a purchase is not made.

Fig. 9. The Forced Enrollment type from the Forced Action category of dark patterns.

Based on our taxonomy of dark patterns, we classify Forced Enrollment as *asymmetric* (it requires competing the additional, tangential tasks, creating unequal choices) and *restrictive* (it mandates enrolling in marketing communication or creating accounts) in nature.

6 DARK PATTERNS AS A THIRD-PARTY SERVICE: A CASE STUDY OF SOCIAL PROOF ACTIVITY NOTIFICATIONS

In many instances, third-party entities—i.e., organizations and companies other than the shopping websites themselves—were often responsible for creating and presenting dark patterns on behalf of the shopping websites. We observed this frequently to be the case for one dark pattern in particular: Social Proof Activity Notifications (Section 5.4). In this section, we shed light on this ecosystem of third parties, using the list of websites that displayed Activity Notifications as our starting point.

6.1 Detecting Third-party Entities

In order to detect third-party entities, we had to uncover scripts that were served from third-party domains and were responsible for creating Social Proof Activity Messages. However, automatically attributing certain interface elements and page modifications to third-party scripts constitutes a more challenging task because modern browsers do not expose any means to attribute DOM changes (e.g. displaying a popup dialog) to particular scripts. Further, web pages may be modified by several different first and third-party scripts in the same visit, making attribution trickier.

To overcome this challenge, we employed a combination of automated and manual analyses. We used the following observation: when a third-party entity displays an Activity Notification on a shopping website, its content should be included in the HTTP response received from this third party's servers on that website. For example, if the notification states "Jane from Washington, DC just purchased this product", looking up the customer name and location—in this case "Jane" and "Washington, DC"—in the HAR file for that website should reveal the end point of the server that issued the notification. Thus for all notifications of this kind, we extracted the name and location pairs from the content, searched the HAR files for these pairs, and where successful, recorded the HTTP endpoints corresponding to the third-parties. We then manually verified these endpoints and determined the responsible entities by using the WHOIS database, visiting the script domains and using search engines to uncover the company identities and websites.

Table 2. List and prevalence of Social Proof Activity Notifications enabling third-party entities in our data set of 11K shopping websites and the home pages of Alexa top million websites [7]. Where available, we list additional dark patterns the third parties claim to offer. Nice/Bizzy, Woocommerce Notification, Boost, and Amasty are Shopify, Woocommerce, Wordpress and Magento plugins respectively.

Third-party Entity	Prevalence		Additional Dark Patterns
	# Shopping Websites	# Alexa Top Million	
Beeketing	406	4,151	Pressured Selling, Urgency, Scarcity
Dynamic Yield	114	416	Urgency
Yieldify	111	323	Urgency, Scarcity
Fomo	91	663	–
Fresh Relevance	86	208	Urgency
Insider	52	484	Scarcity, Urgency
Bizzy	33	213	–
ConvertCart	31	62	–
Taggstar	27	4	Scarcity, Urgency
Qubit	25	73	Pressured Selling, Scarcity, Urgency
Exponea	18	180	Urgency, Scarcity
Recently	14	66	–
Proof	11	508	–
Fera	11	132	Pressured Selling, Scarcity, Urgency
Nice	10	80	–
Woocommerce Notification	10	61	–
Bunting	5	17	Urgency, Scarcity
Credibly	4	67	–
Convertize	3	58	Scarcity, Urgency
LeanConvert	2	0	–
Boost	1	3	–
Amasty	1	0	Pressured Selling, Scarcity, Urgency

Where this analysis failed to return a HTTP endpoint from the HAR files, and for all other kinds of Social Proof Activity Notification (e.g., “This product was added to cart 10 times in the last day”), we manually visited the websites containing the message to determine the responsible third parties. We sped up this analysis using Google Chrome Developer Tool’s “DOM change breakpoints” feature [16], which helped us easily determine the responsible entities.

Having determined the third-party entities, we measured their prevalence across all the shopping websites in our data set. To do so, we searched the HTTP request data from checkout crawls for the third-party domains we identified. Finally, as a reference point, we also determined their prevalence on the web—beyond shopping websites—using the latest publicly available crawl data (November 2018) from the Princeton Web Census Project [7, 39]. This public project documents the prevalence of third-party scripts using periodic scans of home pages of Alexa top million sites and is available for external researchers to use.

6.2 The Ecosystem Of Third-party Entities

Table 2 summarizes our findings. We discovered a total of 22 third-party entities, embedded in 1,066 of the 11K shopping websites in our data set, and in 7,769 of the Alexa top million websites. We note that the prevalence figures from the Princeton Web Census Project data should be taken as a lower bound since their crawls are limited to home pages of websites. This difference in prevalence is particularly visible for certain third-party entities like Qubit and Taggstar, where their prevalence is higher in our data set compared to the Web Census data. By manually examining websites that contained these third parties, we discovered that many shopping websites only embedded them in their product—and not home—pages, presumably for functionality and performance reasons.

We learned that many third-party entities offered a variety of services for shopping websites, including plugins for popular e-commerce platforms such as Shopify[5] and Woocommerce[6]. To better understand the nature and capabilities of each third-party entity, we examined any publicly available marketing materials on their websites.

Broadly, we could classify the third-party entities into two groups. The first group exclusively provided Social Proof Activity Notifications integration as a service. The second group provided a wider array of marketing services that often enabled other types of dark patterns; most commonly these were Scarcity and Urgency dark patterns. We list all these additional dark pattern capabilities in the rightmost column of Table 2.

Many of the third-parties advertised practices that appeared to be—and sometimes unambiguously were—manipulative: "[p]lay upon [customers'] fear of missing out by showing shoppers which products are creating a buzz on your website" (Fresh Relevance), "[c]reate a sense of urgency to boost conversions and speed up sales cycles with Price Alert Web Push" (Insider), "[t]ake advantage of impulse purchases or encourage visitors over shipping thresholds" (Qubit). Further, Qubit also advertised Social Proof Activity Notifications that could be tailored to users' preferences and backgrounds.

In some instances, we found that third parties openly advertised the deceptive capabilities of their products. For example, Boost dedicated a web page—titled "Fake it till you make it"—to describing how it could help create fake orders [12]. Woocommerce Notification—a Woocommerce platform plugin—also advertised that it could create fake social proof messages: "[t]he plugin will create fake orders of the selected products" [23]. Interestingly, certain third parties (Fomo, Proof, and Boost) used Social Proof Activity Messages on their own websites to promote their products.

Finally, we also discovered that some of these deceptive practices resulted in e-commerce platforms taking action against third-party entities. For instance, Beeketing's—the most popular third party provider in our data set—"Sales Pop" Shopify plugin was temporarily removed from Shopify in an effort to crack down on deceptive practices [65, 73]. The plugin had allowed websites to create fake Activity Notifications by entering fabricated sales data.

In summary, we discovered that these third-parties could be easily integrated into websites and offered websites several "persuasive" capabilities, some of which were particularly deceptive.

7 DISCUSSION

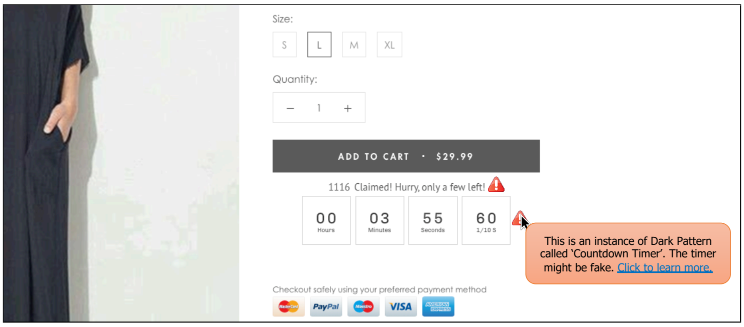
7.1 Dark Patterns and Implications For Consumers

Many dark patterns constitute manipulative and deceptive practices that past work has shown users are increasingly becoming aware of [35]. Further, when informed, users desire countermeasures to curb the effects of these practices [63].

Our current data set of dark patterns, comprising of screenshots and text segments, can be used to build countermeasures to help users make more informed decisions even in the presence of dark patterns. One such countermeasure could be a public facing website that scores shopping websites on how adversarial their design practices are. Our data set can also enable the development of browser extensions that automatically detect and flag dark patterns on e.g., shopping websites

[5] https://shopify.com

[6] https://woocommerce.com



This is an instance of Dark Pattern called ‘Countdown Timer’. The timer might be fake. Click to learn more.

Fig. 10. Mockup of a possible browser extension that can be developed using our data set. The extension flags instances of dark patterns with a red warning icon. By hovering over the icon, the user can learn more about the specific pattern.

(Figure 10). Such tools require further research, for example on methods to detect dark patterns on arbitrary websites. This could be done through a crowdsourced list of dark patterns with manual heuristics for detection (similar to how ad blockers work [25]), or through trained machine learning classifiers. Eventually, such tools could be integrated into browsers themselves. For example, in recent years, Firefox and Safari have shown interest in integrating tools that promote consumer privacy (e.g., features to block web tracking by default [64, 80]). However, finding the right incentives for companies to implement these solutions might be challenging in the context of dark patterns, since browsers might be weary of policing content on the web.

Additionally, future studies could leverage our taxonomy of dark pattern characteristics to better understand their effects on users, as well as to ascertain which dark patterns are considered most egregious by users (e.g., by means of users studies).

7.2 Implications for Consumer Protection Policy And Retailers

Our results demonstrate that a number of shopping websites use deceptive dark patterns, involving affirmative and false representations to consumers. We also found 22 different third-party entities that enable the creation of Social Proof dark patterns. Some of these entities promote blatantly deceptive practices and provide the infrastructure for retailers to use these practices to affect consumer behaviors for profit. These practices are unambiguously unlawful in the United States (under Section 5 of the Federal Trade Commission Act and similar state laws [43]), the European Union (under the Unfair Commercial Practices Directive and similar member state laws [40]), and numerous other jurisdictions.

We also find practices that are unlawful in a smaller set of jurisdictions. In the European Union, businesses are bound by an array of affirmative disclosure and independent consent requirements in the Consumer Rights Directive [41]. Websites that use the Sneaking dark patterns (Sneak into Basket, Hidden Subscription, and Hidden Costs) on European Union consumers are likely in violation of the Directive. Furthermore, user consent obtained through Trick Questions and Visual Interference dark patterns do not constitute freely given, informed and active consent as required by the General Data Protection Regulation (GDPR) [42]. In fact, the Norwegian Consumer Council filed a GDPR complaint against Google in 2018, arguing that Google used dark patterns to manipulate users into turning on the “Location History” feature on Android, and thus enabling constant location tracking [46].

In addition to demonstrating specific instances of unlawful business practices, we contribute a new approach for regulatory agencies and other consumer protection stakeholders (e.g., journalists and civil society groups) to detect dark patterns. The crawling and clustering methodology that we developed is readily generalizable, and it radically reduces the difficulty of discovering and measuring dark patterns at web scale. Furthermore, we already have a data set available of third-party entities which provide the infrastructure to enable certain deceptive dark patterns such as social proof messages across a large number of shopping websites. Regulators can also use this list as an initial starting point to inform policy and regulation around what kinds of practices should be allowable in online shopping.

7.3 Dark Patterns and Future Studies At Scale

We created an extensive infrastructure for Internet measurement that leverages automated techniques and machine learning to conduct measurements of dark patterns at scale. Researchers can extend this infrastructure to measure dark patterns that are not just text-based. For example, Urgency can be created by a blinking timer; similarly, Forced Actions that are beneficial to a company can be enforced by making the default option (e.g., subscribing to a paid service) visually more appealing and noticeable than its alternative (e.g., not subscribing). In our study, we do not take into account colors or other kinds of visually manipulative dark patterns. However, future research could extend our measurement infrastructure to also capture these type of visual manipulations; one starting point for doing so may be the design mining literature [55]. In addition, others can leverage our taxonomy of dark pattern characteristics to study and analyze dark patterns in other domains such as emails and mobile applications [38, 58].

7.4 Limitations

Our research has limitations. First, we only take into account text-based dark patterns and, therefore, leave out those that are inherently visual (e.g., a change of font size or color to emphasize one part of the text more than another from an otherwise seemingly harmless pattern). Second, during our crawls we experienced a small fraction of Selenium crashes, which did not allow us to either retrieve product pages or complete data collection on certain websites. Third, the crawler failed to completely simulate the product purchase flow on some websites Section 4. Fourth, we only crawled product pages and checkout pages, missing out on dark patterns present in other common pages such as the home page of websites, product search pages, and account creation pages.

8 CONCLUSION

In this paper, we studied the ~11K most popular shopping websites to identify their use of dark patterns. To build the data set, we created automated techniques to simulate user actions on shopping websites and collect both text and screenshots of these websites at scale. We defined and characterized dark patterns, describing their underpinnings and linking our definitions to the cognitive biases dark patterns leverage to affect user behavior. We found instances of at least one dark pattern on approximately 11.2% of the examined websites, where 183 of the websites displayed deceptive messages. Furthermore, we discovered that dark patterns are often enabled by third-party entities, of which we identify 22, two of which advertise practices that enable deceptive patterns. Based on these findings, we suggest that future work focuses on empirically evaluating the effects of dark patterns on user behavior, developing countermeasures against dark patterns so that users have a fair and transparent experience, and extending our work to conduct further measurements of this kind.

ACKNOWLEDGMENTS

We are grateful to Mihir Kshirsagar, Finn Myrstad, Vincent Toubiana, and Joe Calandrino for feedback on this draft.

REFERENCES

[1] [n. d.]. Affinion Group faces class action after paying out claims to AGs. https://www.washingtonexaminer.com/affinion-group-faces-class-action-after-paying-out-claims-to-ags. Accessed March 12, 2019.

[2] [n. d.]. Marketing Firm Agrees To $30 Million Settlement. https://www.wsj.com/articles/marketing-firm-agrees-to-30-million-settlement-1381441148. Accessed March 12, 2019.

[3] 2015. rafaelw/mutation-summary: A JavaScript library that makes observing changes to the DOM easy. https://github.com/rafaelw/mutation-summary Accessed March 17, 2019.

[4] 2018. aboSamoor/polyglot: Multilingual text (NLP) processing toolkit. https://github.com/aboSamoor/polyglot Accessed March 12, 2019.

[5] 2018. Facebook has been collecting call history and SMS data from Android devices. https://www.theverge.com/2018/3/25/17160944/facebook-call-history-sms-data-collection-android. Accessed April 2, 2019.

[6] 2018. MDN. https://developer.mozilla.org/en-US/docs/Web/HTML/Block-level_elements Accessed March 12, 2019.

[7] 2018. Princeton Web Census Data Release. https://webtransparency.cs.princeton.edu/webcensus/data-release/. Accessed April 2, 2019.

[8] 2018. Selenium. https://selenium-python.readthedocs.io Accessed March 12, 2019.

[9] 2019. Alexa Top Sites - API Reference. https://docs.aws.amazon.com/AlexaTopSites/latest/ApiReferenceArticle.html Accessed March 12, 2019.

[10] 2019. Alexa Web Information Service. https://docs.aws.amazon.com/AlexaWebInfoService/latest/index.html Accessed March 12, 2019.

[11] 2019. Booster. https://boostertheme.com Accessed March 12, 2019.

[12] 2019. Fake it till you make it - Social Proof. https://www.boostplugin.com/fake-boosts. Accessed April 4, 2019.

[13] 2019. .har - Wikipedia. https://en.wikipedia.org/wiki/.har Accessed March 12, 2019.

[14] 2019. The hdbscan Clustering Library. https://hdbscan.readthedocs.io/en/latest/index.html Accessed March 15, 2019.

[15] 2019. Hotel booking sites forced to end misleading sales tactics. https://www.theguardian.com/business/2019/feb/06/hotel-booking-sites-forced-to-end-misleading-sales-tactics Accessed March 12, 2019.

[16] 2019. How To Pause Your Code With Breakpoints In Chrome DevTools | Tools for Web Developers | Google Developers. https://developers.google.com/web/tools/chrome-devtools/javascript/breakpoints#dom. Accessed April 2, 2019.

[17] 2019. Hurrify - Countdown Timer. https://apps.shopify.com/hurrify-countdown-timer Accessed March 12, 2019.

[18] 2019. MDN. https://developer.mozilla.org/en-US/docs/Web/API/MutationObserver Accessed March 12, 2019.

[19] 2019. Node.nodeType - Web APIs | MDN. https://developer.mozilla.org/en-US/docs/Web/API/Node/nodeType Accessed March 12, 2019.

[20] 2019. SENATORS INTRODUCE BIPARTISAN LEGISLATION TO BAN MANIPULATIVE âĂŸDARK PATTERNSâĂŹ. https://www.fischer.senate.gov/public/index.cfm/2019/4/senators-introduce-bipartisan-legislation-to-ban-manipulative-dark-patterns. Accessed April 4, 2019.

[21] 2019. The top 500 sites on the web. https://www.alexa.com/topsites/category/Top/Shopping Accessed March 12, 2019.

[22] 2019. Webshrinker. https://www.webshrinker.com Accessed March 12, 2019.

[23] 2019. WooCommerce Notification | Boost Your Sales - Recent Sales Popup - Live Feed Sales - Upsells - WordPress plugin| WordPress.org. https://wordpress.org/plugins/woo-notification/. Accessed April 4, 2019.

[24] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. ACM, New York, NY, USA, 674–689. https://doi.org/10.1145/2660267.2660347

[25] AdBlock. 2019. AdBlock. https://adblockplus.org. Accessed April 4, 2019.

[26] Praveen Aggarwal and Rajiv Vaidyanathan. 2015. Use It Or Lose It: Time-Limited Promotions And Purchase Behavior. In *Proceedings of the 2002 Academy of Marketing Science (AMS) Annual Conference*, Harlan E. Spotts (Ed.). Springer International Publishing, Cham, 2–2.

[27] George Ainslie. 1975. Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control, Vol. 82. American Psychological Association, 463–496. https://doi.org/10.1037/h0076860

[28] Hal R Arkes and Peter Ayton. 1999. The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological bulletin* 125, 5 (1999), 591.

[29] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python* (1st ed.). O'Reilly Media, Inc.

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

[31] Harry Brignull. 2018. Dark Patterns. https://darkpatterns.org/. Accessed March 12, 2019.

[32] Will Browne and Mike Swarbrick Jones. 2017. What works in e-commerce - a meta-analysis of 6700 online experiments. *Qubit Digital Ltd* (2017).

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

[34] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. 2013. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*, Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 160–172.

[35] Shruthi Sai Chivukula, Chris Watkins, Lucca McKay, and Colin M. Gray. 2019. &#34;Nothing Comes Before Profit&#34;: Asshole Design In the Wild. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article LBW1314, 6 pages. https://doi.org/10.1145/3290607.3312863

[36] Robert B Cialdini. 2009. *Influence: Science and practice*. Vol. 4. Pearson education Boston.

[37] Gregory Conti and Edward Sobiesk. 2010. Malicious interface design: exploiting the user. In *Proceedings of the 19th international conference on World wide web*. ACM, 271–280.

[38] Biplab Deka, Zifeng Huang, Chad Franzen, Joshua Hibschman, Daniel Afergan, Yang Li, Jeffrey Nichols, and Ranjitha Kumar. 2017. Rico: A Mobile App Dataset for Building Data-Driven Design Applications. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 845–854. https://doi.org/10.1145/3126594.3126651

[39] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-million-site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, New York, NY, USA, 1388–1401. https://doi.org/10.1145/2976749.2978313

[40] European Parliament and Council of European Union. 2011. Directive 2011/83/EU. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0083. Accessed March 12, 2019.

[41] European Parliament and Council of European Union. 2011. Directive 2011/83/EU of the European Parliament and of the Council. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0083. Accessed March 12, 2019.

[42] 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 March 12, 2019.

[43] Federal Trade Commission. 1914. 15 U.S. Code §45. Unfair methods of competition unlawful; prevention by Commission. https://www.law.cornell.edu/uscode/text/15/45. Accessed March 12, 2019.

[44] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. 2017. Measuring HTTPS Adoption on the Web. In *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, Vancouver, BC, 1323–1338. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/felt

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

[46] Frobrukerrådet. 2018. New study: Google manipulates users into constant tracking. https://www.forbrukerradet.no/side/google-manipulates-users-into-constant-tracking. Accessed March 12, 2019.

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

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

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

[50] Martie G Haselton, Daniel Nettle, and Damian R Murray. 2015. The evolution of cognitive bias. *The handbook of evolutionary psychology* (2015), 1–20.

[51] Joel Huber, John W. Payne, and Christopher Puto. 1982. Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research* 9, 1 (1982), 90–98. http://www.jstor.org/stable/2488940

[52] J. Jeffrey Inman and Leigh McAlister. 1994. Do Coupon Expiration Dates Affect Consumer Behavior? *Journal of Marketing Research* 31, 3 (1994), 423–428. http://www.jstor.org/stable/3152229

[53] Eric J. Johnson, Steven Bellman, and Gerald L. Lohse. 2002. Defaults, Framing and Privacy: Why Opting In-Opting Out1. *Marketing Letters* 13, 1 (01 Feb 2002), 5–15. https://doi.org/10.1023/A:1015044207315

[54] Jae Min Jung and James Kellaris. 2004. Cross-national differences in proneness to scarcity effects: The moderating roles of familiarity, uncertainty avoidance, and need for cognitive closure. *Psychology and Marketing* 21 (09 2004), 739 – 753. https://doi.org/10.1002/mar.20027

[55] Ranjitha Kumar, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R. Klemmer, and Jerry O. Talton. 2013. Webzeitgeist: Design Mining the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3083–3092. https://doi.org/10.1145/2470654.2466420

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

[57] Brian Lickel, Kostadin Kushlev, Victoria Savalei, Shashi Matta, and Toni Schmader. 2014. Shame and the motivation to change the self. *Emotion* 14, 6 (2014), 1049.

[58] Thomas F. Liu, Mark Craft, Jason Situ, Ersin Yumer, Radomir Mech, and Ranjitha Kumar. 2018. Learning Design Semantics for Mobile Apps. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 569–579. https://doi.org/10.1145/3242587.3242650

[59] Sheng Luo, Bin Gu, Xingbiao Wang, and Zhaoquan Zhou. 2018. Online Compulsive Buying Behavior: The Mediating Role of Self-control and Negative Emotions. In *Proceedings of the 2018 International Conference on Internet and e-Business*. ACM, 65–69.

[60] Xueming Luo. 2005. How Does Shopping With Others Influence Impulsive Purchasing? *Journal of Consumer Psychology* 15 (12 2005), 288–294. https://doi.org/10.1207/s15327663jcp1504_3

[61] Michael Lynn. 1991. Scarcity effects on value: A quantitative review of the commodity theory literature. *Psychology & Marketing* 8, 1 (1991), 43–57.

[62] Luigi Mittone and Lucia Savadori. 2009. The scarcity bias. *Applied Psychology* 58, 3 (2009), 453–468.

[63] Carol Moser, Sarita Y Schoenebeck, and Paul Resnick. 2019. Impulse Buying: Design Practices and Consumer Needs. (2019).

[64] Mozilla. 2019. What Happened To Tracking Protection? https://support.mozilla.org/en-US/kb/tracking-protection-pbm. Accessed April 4, 2019.

[65] Dang Van Nhan. 2018. Beeketing eCommerce Success Community Public Group | Facebook. https://www.facebook.com/groups/beeketing.ecommerce.community/permalink/1691157507670306/. Accessed April 2, 2019.

[66] Nick Nikiforakis, Luca Invernizzi, Alexandros Kapravelos, Steven Van Acker, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2012. You Are What You Include: Large-scale Evaluation of Remote Javascript Inclusions. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS '12)*. ACM, New York, NY, USA, 736–747. https://doi.org/10.1145/2382196.2382274

[67] C. Whan Park, Sung Youl Jun, and Deborah J. MacInnis. 2000. Choosing What I Want versus Rejecting What I Do Not Want: An Application of Decision Framing to Product Option Choice Decisions. *Journal of Marketing Research* 37, 2 (2000), 187–202. http://www.jstor.org/stable/1558499

[68] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[69] Steven Reiss. 2004. Multifaceted Nature of Intrinsic Motivation: The Theory of 16 Basic Desires. *Review of General Psychology* 8, 3 (2004), 179–193. https://doi.org/10.1037/1089-2680.8.3.179 arXiv:https://doi.org/10.1037/1089-2680.8.3.179

[70] Quirin Scheitle, Oliver Hohlfeld, Julien Gamba, Jonas Jelten, Torsten Zimmermann, Stephen D. Strowes, and Narseo Vallina-Rodriguez. 2018. A Long Way to the Top: Significance, Structure, and Stability of Internet Top Lists. In *Proceedings of the Internet Measurement Conference 2018 (IMC '18)*. ACM, New York, NY, USA, 478–493. https://doi.org/10.1145/3278532.3278574

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

[72] Muzafer Sherif. 1936. The psychology of social norms. (1936).

[73] Ben Shuffer. 2018. Shopify Cracking Down On Fake Scarcity? https://medium.com/@benshaffer_83355/shopify-cracking-down-on-fake-scarcity-e1509b11cb75/. Accessed April 2, 2019.

[74] Daniel Susser, Beate Roessler, and Helen F. Nissembaum. December 23, 2018. Online Manipulation: Hidden Influences in a Digital World. (December 23, 2018). https://doi.org/10.2139/ssrn.3306006

[75] Pew Research Center Internet & Technology. 2016. Online Shopping and E-Commerce. http://www.pewinternet.org/2016/12/19/online-shopping-and-e-commerce/ Accessed March 12, 2019.

[76] Richard H. Thaler and Cass R. Sunstein. 2003. Libertarian Paternalism. *American Economic Review* 93, 2 (May 2003), 175–179. https://doi.org/10.1257/000282803321947001

[77] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 4157 (1974), 1124–1131.

[78] Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211, 4481 (1981), 453–458. https://doi.org/10.1126/science.7455683 arXiv:http://science.sciencemag.org/content/211/4481/453.full.pdf

[79] Amos Tversky and Daniel Kahneman. 1989. Rational Choice and the Framing of Decisions. In *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*, Birsen Karpak and Stanley Zionts (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 81–126.

[80] WebKit. 2019. Intelligent Tracking Protection. https://webkit.org/blog/7675/intelligent-tracking-prevention. Accessed April 4, 2019.

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

[82] Bo Xiao and Izak Benbasat. 2011. Product-Related Deception in E-Commerce: A Theoretical Perspective. *MIS Quarterly* 35, 1 (2011), 169–195. http://www.jstor.org/stable/23043494

[83] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark patterns in the design of games. In *Foundations of Digital Games 2013*.

[84] Fuzheng Zhang, Nicholas Jing Yuan, Kai Zheng, Defu Lian, Xing Xie, and Yong Rui. 2015. Mining consumer impulsivity from offline and online behavior. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1281–1292.

A APPENDIX

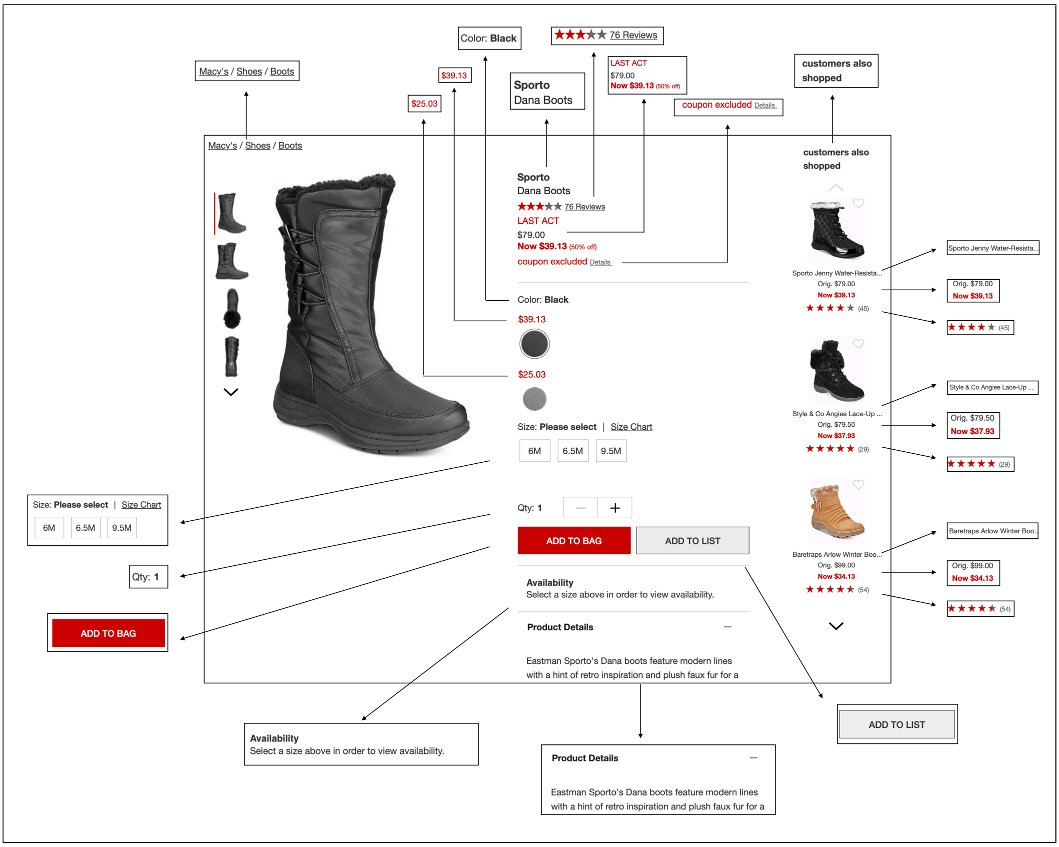
Table 3. Confusion Matrices From Our Evaluation of Alexa's and Webshrinker's Website Classifiers.

		Alexa Prediction		Webshrinker Prediction	
		Not Shopping	Shopping	Not Shopping	Shopping
Truth	Not Shopping	442	1	423	20
	Shopping	53	4	10	47

Algorithm 1 Page Segmentation

```
ignoredElements ← [‘script’, ‘style’, ‘noscript’, ‘br’, ‘hr’]
blockElements ← [‘div’, ‘section’, ‘article’, ‘aside’, ‘nav’, ‘header’, ‘footer’, ‘main’, ‘form’, ‘fieldset’, ‘table’]

function SEGMENTS(element)                                          ▷ Returns a list of segments
    if not element then
        return empty list
    end if
    tag ← element.tagName
    if tag in ignoredElements or element not visible or element not bigger than 1 pixel then
        return empty list
    end if
    if tag in blockElements then
        if element does not contain visible blockElements then
            if all of element’s children in ignoredElements then
                return empty list
            else
                if element occupies more than 30% of the page then
                    return list of segments(child) for each child in element’s children
                else
                    return [element]
                end if
            end if
        else if element contains text nodes then
            return [element]
        else
            return list of segments(child) for each child in element’s children
        end if
    else
        if element has at least one child with tag in blockElements then
            return list of segments(child) for each child in element’s children
        else
            if element occupies more than 30% of the page then
                return list of segments(child) for each child in element’s children
            else
                return [element]
            end if
        end if
    end if
end function
```



Availability
Select a size above in order to view availability.

Product Details

Eastman Sporto's Dana boots feature modern lines with a hint of retro inspiration and plush faux fur for a

Fig. 11. An illustration of the page segmentation algorithm. The page is segmented into smaller meaningful "building blocks" or segments. Only segments containing text are recorded.