Towards an Understanding of Dark Pattern Privacy Harms

JOHANNA GUNAWAN, Northeastern University, USA
DAVID CHOFFNES, Northeastern University, USA
WOODROW HARTZOG, Northeastern University, USA
CHRISTO WILSON, Northeastern University, USA

Current efforts to regulate dark patterns suffer from a lack of defined, legally cognizable harms, which makes bills like the DETOUR Act and CPRA largely ineffective in prohibiting the use of malicious interfaces. In this position paper, we discuss the challenges of articulating these harms, then outline a research agenda for empirically measuring the labor costs, or effort burdens, that users may incur while trying to avoid or overcome dark patterns. We posit that effort measurements provide a path towards defining dark pattern harms, with the goal of addressing dark patterns from a privacy-focused policy perspective.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; • **Social and professional topics → Computing / technology policy**; • **Information systems →** *Web interfaces.*

Additional Key Words and Phrases: UX design, dark patterns, privacy, policy

1 INTRODUCTION

Recent scholarship has allowed us to better describe and categorize dark patterns, [2, 3, 9], discover new ones [14], find where they hide in certain environments [8], and more. We can learn more about dark patterns and their mechanics through other disciplines' perspectives, such as behavioral psychology and legal scholarship [13, 20]. Dark patterns have begun to receive real regulatory attention through the Deceptive Experiences to Online Users Reduction (DETOUR) Act [21] and the California Privacy Rights Act of 2020 (CPRA) [1]. These regulatory attempts, though sorely needed, are currently unconvincing. Blanket statements to prohibit dark patterns are better than none, but these policies are doomed to fail until they provide stricter definitions of when a dark pattern becomes 'too dark.'

In this position paper, we explore the concept of darkness through a policy lens, focusing especially on privacy and legal definitions of harm. We then propose future work to empirically explore one potential measure of darkness: the effort a user must exert in order to avoid or overcome a dark pattern's persuasiveness. For the purpose of this position paper, we define dark patterns as interface designs that lead users towards outcomes that benefit the platform over the user, or that steer users away from what they are intending to do.

2 UNDERSTANDING HARMS: HOW DARK IS 'DARK?'

At the core of darkness considerations lies the fundamental assumption that a dark pattern transgresses against another party for the benefit of whoever controls an interface's design. All four of Mathur et. al.'s normative approaches take this format – one party (individual welfare, collective welfare, regulatory objectives, or autonomy) is disadvantaged, while those who employ dark patterns benefit [15]. These are skewed power dynamics in a landscape where tech platforms have unprecedented levels of control users' digital lives – not just their online experiences.

CHI'21, May 8–13, 2021, Online Virtual Conference

When it comes to privacy, articulating transgressions in the law is difficult. As a legal concept, privacy struggles to provide a clear-cut definition of harm [6, 7] – and yet definitions are necessary to improve privacy regulation and meaningfully enforce it. Courts and lawmakers demand concrete harms before enforcing privacy laws, but the ostensible harm from dark patterns do not always rise to relevant legal thresholds. Similarly, harms resulting from manipulation [12, 19] are difficult to clearly identify and articulate as significant, discrete, and concrete. It is an ongoing challenge for courts to tell when attempts to persuade people have gone too far.

Recent scholarship provides new ways of understanding privacy harms. Danielle Citron and Daniel Solove offer a privacy harms roadmap [5] that stresses why some types of privacy harms should be considered cognizable, and their typology maps neatly to many of the consequences outlined by several dark patterns taxonomies [2, 3, 9, 13–15]. This considerable overlap provides a promising option for legislating dark patterns – and it reflects the privacy and data protection goals of both the DETOUR Act and the CPRA. Another proposal for defining privacy harms [4] reflects the same transgressional, adversarial concepts in our current understanding of dark patterns: Ryan Calo suggests a "subjective" privacy harm, arising from the "individual... perception of unwanted observation," and an "objective" harm, which involves the "forced and unanticipated use of information about [an individual against that individual]" [4].

Even as definitions of privacy harms improve, dark pattern harms remain challenging to distinguish. Dark pattern harms have yet to be explicitly addressed by regulation. The CPRA's two references to dark patterns only mention that business should not use them, and that "agreement obtained through use of dark patterns does not constitute consent" [1]. The DETOUR Act does not mention dark patterns explicitly, but frames dark patterns as interface designs that "[obscure, subvert, or impair user autonomy, decision-making, or choice to obtain consent or user data]" [21]. While recent scholarship shows that dark patterns create a slew of other costs, like financial costs or cognitive burdens [13–15], privacy-specific approaches to dark patterns are all that are currently available in legislation.

3 FUTURE WORK: MEASUREMENTS OF EFFORT

The potential consequences of dark patterns for the end user are not purely financial, temporal, or privacy-eroding, but potentially 'all of the above,' as with preselected subscription checkboxes for marketing emails. These potential harms must be meaningfully understood before they can be legislated, and we need not only a common descriptive language but tools or benchmarks by which to determine dark patterns' potential impact. Enforcement requires legally cognizable harms. Several teams have already begun to study the impact of dark patterns, putting dark patterns research down a path that will lead to articulated harms. Mathur et al. highlight an initial set of dark pattern outcomes based on three different normative lenses [15]. Some observed the impact of dark patterns on users' consent or privacy choices [16, 22]; others measured the usability of privacy choice paths [10].

Many open questions still stand regarding dark pattern impact. Our team is interested in exploring what 'legally cognizable' dark pattern privacy harms might be, and we are especially interested in understanding where thresholds for this criteria might lie. Our interdisciplinary team's experience spans several privacy research domains, such as web, mobile, IoT, advertising, law, and misinformation. Our interest in dark patterns is driven by our privacy research focus. For example, our team members' previous research [17] investigated privacy behaviors that are reminiscent of the Bad Defaults pattern and the Violate dark pattern strategy [2]; other team members' work has also focused especially on regulating design for improved privacy [11].

We aim to conduct preliminary investigations into where legally cognizable harm thresholds might lie for dark patterns, primarily from an individual welfare perspective [15]. Specifically, we

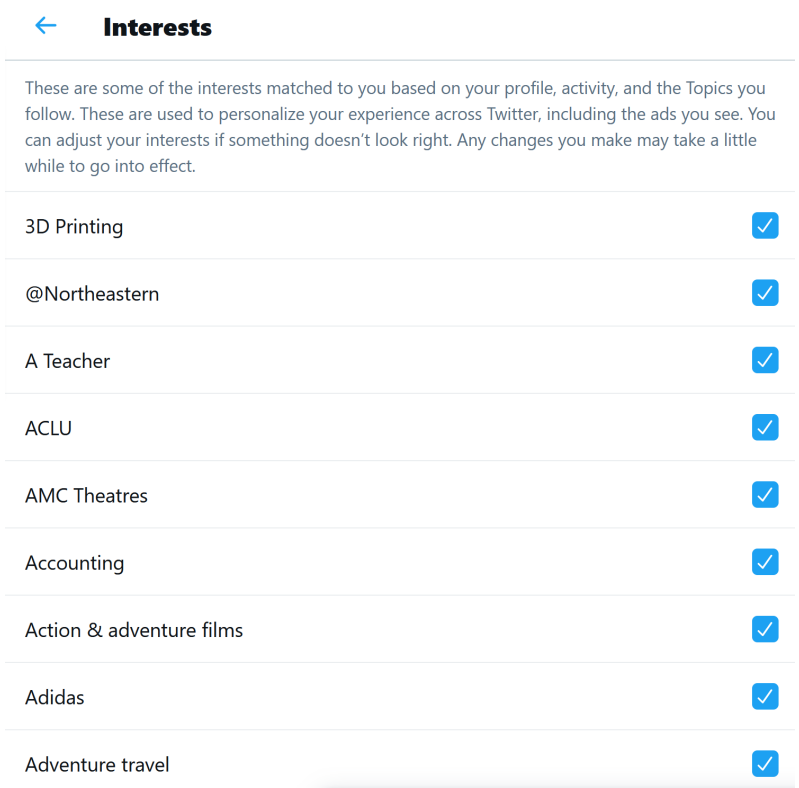

Fig. 1. Example of preselected Bad Defaults en masse, taken from the first author's personal Twitter under *Settings>Privacy and safety>Ads preferences>Interests*. The image only shows a subset of 505 interests which must be individually toggled if the author wished to depersonalize their Twitter ad experience. These dark patterns be could measured by the 505 toggles, compounded patterns of Preselection and Bad Defaults, the number of nested settings pages required to access this setting, and other metrics.

will prioritize patterns that impact users' decision making regarding their privacy rights [20]. Our proposed corpus of patterns to investigate includes privacy dark patterns [2] as well as general dark patterns that may still erode user privacy like many Nagging, Obstruction, and Interface Interference patterns [9]. Some dark patterns are more difficult to measure against privacy risk (like Sneak into Basket patterns [9]) or the effort to avoid them (like Testimonials or Low-Stock Messages [14]) – we believe measurements for these patterns is of high importance, but omit these as out of scope for the work proposed in this position paper.

We hypothesize that dark pattern maliciousness is partially dependent on the level of effort a user must exert to extricate themselves from the pattern's grasp. Like Liguri and Strahilevitz' experiment on the financial cost of dark patterns [13], our proposed work aims to capture the labor costs a user might accrue due to a dark pattern's presence. This effort could be examined via metrics such as the number of clicks, popups, or pages required to access a privacy setting or account deletion option; the time to complete a privacy-forward task in the presence of dark patterns; or levels of navigation required to achieve the task. We envision this investigation through user studies, as in Mathur et al.'s suggestion of measuring the cognitive burden imposed on users [15], as well as through manual and automatic experiments to calculate total values for the above metrics.

4 WHAT WE HOPE TO DISCOVER

The expected outcomes of this work are to: (1) investigate which types of dark patterns are amenable to effort-based measurements and (2) evaluate dark patterns against measures of effort to avoid or overcome a pattern, with a longer-term goal of assisting designers and lawmakers in formalizing dark pattern standards. These measurements may also supplement our understanding of dark pattern characteristics like asymmetry, deceptiveness, and restrictiveness [14]. For example, this data might reveal *how* restrictive 'Hard to Cancel' or 'Forced Enrollment' [14] patterns are, or how deeply Sneaking [9] patterns hide information.

4.1 Compound Effects of Dark Patterns

We know that dark patterns may be combined in the same interface element [8, 9, 14] – for example, when preselected boxes are provided with confusing language – but we don't currently know whether there is a meaningful increase or multiplier of darkness when several patterns are simultaneously present in one interface element, or if certain dark patterns contribute more than others to the interface's 'darkness' when appearing concurrently. Furthermore, we currently do not know the impact of multiple dark patterns on a user's experience when the patterns are presented in succession during a task flow, in several locations on a given page or screen, or in other environments. Conversely, we don't know whether compound dark patterns are easier to circumvent or address on the user side, if they might have the ability to 'kill two birds with one stone,' or one click. Our hope is that labor cost measurements can help begin to answer some of these questions.

4.2 Aggregate and Longitudinal Effects of Dark Patterns

Scope also matters when considering a dark pattern's impact, and labor cost measurements may provide a way to understand cumulative dark pattern impact. One preselected mailing list checkbox may be simple enough to un-check during a registration flow, but how burdensome are preselected cookie management notices for a user opening several browser sites in the span of a few minutes? In a vacuum, one dark pattern may seem rather innocuous, but dark patterns may be more problematic when viewed at scale. Quantitative user studies on the impact of dark patterns on user privacy might help us better understand when a pattern becomes too dark in a greater context – which may help avoid dark patterns being dismissed as de minimis harms.

4.3 Revealing Other Unknowns

It is also possible that collecting effort measurements might reveal that certain privacy controls are not provided to users at all, thus exposing potential violations of privacy law – or such experiments might reveal that users are well-equipped to evade some patterns, but not others. We are additionally interested in whether effort analysis might detect other dark patterns in the process of examining a different one (for example, finding Obstruction patterns while trying to look at Interface Interference samples [9]) or if effort measurements might reveal noncompliance to existing regulation, which may answer questions under a regulatory objectives lens [15].

5 CONCLUSION

While our inspiration and focus for this proposed work comes largely from our team's privacy interests and research backgrounds, we believe that labor cost measurements are useful for dark patterns research more broadly. For example, we believe effort measurements can be taken to investigate users' digital well-being under the digital welfare lens [15], especially for dark patterns

in games [23]. Measurements for this purpose may help create dark patterns provisions in regulation against addictive technologies [18].

This work will not fully close the gap between proposed legislation and interface-level privacy protection, but we believe that measurements like these will contribute to eventual definitions of legally cognizable dark pattern harms. Our proposed work is one small part in the effort to combat dark patterns, and we hope to participate in robust discussions with the greater research community at the "What Can CHI Do About Dark Patterns" workshop.

REFERENCES

[1] 2020. California Privacy Rights Act of 2020 (CCPA).

[2] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proc. of PETS* 2016, 4 (2016), 237–254. https://content.sciendo.com/view/journals/popets/2016/4/article-p237.xml

[3] Harry Brignull. 2010. Dark Patterns. https://www.darkpatterns.org/.

[4] Ryan Calo. 2011. The Boundaries of Privacy Harm. *Indiana Law Journal* 86 (2011). Issue 3. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1641487

[5] Danielle Keats Citron and Daniel J. Solove. 2021. Privacy Harms. (2021). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3782222

[6] Ignacio N. Cofone. 2019. Privacy law needs privacy harm. *The Hill* (2019). https://thehill.com/opinion/cybersecurity/459427-privacy-law-needs-privacy-harm

[7] Ignacio N. Cofone and Adriana Robertson. 2018. *Hastings Law Journal* 69 (2018).

[8] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proc. of CHI.*

[9] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proc. of CHI.*

[10] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2019. An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites. In *Proc. of the Workshop on Usable Security.*

[11] Woodrow Hartzog. 2018. *Privacy's Blueprint: The Battle to Control the Design of New Technologies.* Harvard University Press.

[12] Ido Kilovaty. 2019. Legally Cognizable Manipulation. *Berkeley Technology Law Journal* 34 (2019). Issue 2.

[13] Jamie Luguri and Lior Strahilevitz. 2019. Shining a Light on Dark Patterns. *U of Chicago, Public Law Working Paper No. 719; University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 879* (2019). https://ssrn.com/abstract=3431205

[14] Arunesh Mathur, Gunes Acar, Michael Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (2019).

[15] Arunesh Mathur, Jonathan Mayer, and Kihir Kshirsagar. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. https://arxiv.org/pdf/2101.04843.pdf

[16] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. In *Proc. of CHI.*

[17] Elleen Pan, Jingjing Ren, Martina Lindorfer, Christo Wilson, and David R. Choffnes. 2018. Panoptispy: Characterizing Audio and Video Exfiltration from Android Applications. (2018).

[18] Sen. Hawley Introduces Legislation to Curb Social Media Addiction 2019. Sen. Hawley Introduces Legislation to Curb Social Media Addiction. https://www.hawley.senate.gov/sen-hawley-introduces-legislation-curb-social-media-addiction.

[19] Daniel Susser, Beate Roessler, and Helen Nissenbaum. 2019. Online Manipulation: Hidden Influences in a Digital World.

[20] Ari Ezra Waldman. 2020. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology* 31 (Feb. 2020), 105–109.

[21] Mark R. Warner. 2019. Deceptive Experiences To Online Users Reduction (DETOUR) Act. https://www.congress.gov/bill/116th-congress/senate-bill/1084/text

[22] Fiona Westin and Sonia Chiasson. 2019. Opt out of Privacy or "Go Home": Understanding Reluctant Privacy Behaviours through the FoMO-Centric Design Paradigm. In *Proceedings of the New Security Paradigms Workshop* (San Carlos, Costa Rica) *(NSPW '19).* Association for Computing Machinery, New York, NY, USA, 11 pages. https://doi.org/10.1145/3368860.3368865

[23] Jose P. Zagal, Staffan Bjork, and Chris Lewis. 2013. Dark Patterns in the Design of Games. In *Proc. of Foundations of Digital Games Conference (FDG '13).*