

Deceptive Interfaces Are Perceived as Unattractive: Correlational Evidence from User Evaluations of Dark Patterns

2025-10-16

Abstract

Dark patterns are manipulative interface designs that exploit cognitive biases to steer user behavior. While research has documented their prevalence and categorization, limited empirical work examines how users perceive these manipulations when provided vocabulary to articulate their experiences, and whether deceptive patterns relate to aesthetic judgments. This study presents correlational evidence from 126 social media users who evaluated 15 dark patterns across six strategic categories using an adapted User Experience Questionnaire supplemented with autonomy-focused items (pressuring, addictive, covert, deceptive). Two key findings emerged: First, a strong correlation ($r = 0.93$) between perceived deception and attractiveness demonstrates that users judge deceptive interfaces as unattractive (annoying, unfriendly); suggesting manipulation detection operates through aesthetic judgment. Second, within-category correlations reveal that certain dark pattern strategies constitute coherent perceptual constructs, with Interface Interference patterns ($r = 0.77$) and Social Engineering patterns ($r = 0.72$) showing high internal consistency. These findings suggest users possess tacit awareness of manipulation that becomes articulable when provided appropriate vocabulary, challenging assumptions that manipulation detection requires expert analysis rather than user judgment. Results extend psychological understanding of user experience beyond traditional pragmatic and hedonic dimensions to include autonomy violations as a distinct experiential quality manifesting through aesthetic reactions.

Hauke Sandhaus¹ (ORCID: 0000-0002-4169-0197), Doris Maria Rhomberg² (ORCID: 0009-0003-7542-3438)

¹Digital Life Initiative, Cornell University

²Faculty of Informatics, TU Wien

Keywords: dark patterns, user experience, deception, manipulation, interface design, aesthetics

Corresponding Author

Hauke Sandhaus
Digital Life Initiative
Cornell Tech
2 W Loop Rd
New York, NY 10044
USA
hgs52@cornell.edu

Acknowledgments

This research was conducted during Doris Maria Rhomberg's visiting researcher period at the Digital Life Initiative at Cornell Tech.

Conflict of Interest

The authors declare no conflicts of interest.

Publication Ethics

Informed consent was obtained from all participants included in the study. All procedures were performed in accordance with the ethical standards of the institution.

Authorship

Hauke Sandhaus (ORCID: 0000-0002-4169-0197): conceptualization, methodology, writing – review & editing; Doris Maria Rhomberg (ORCID: 0009-0003-7542-3438): data curation, formal analysis, investigation, writing – original draft. All authors approved the final version of the article.

Open Science

Raw data, analysis scripts, and supplementary materials are available through the Open Science Framework at https://osf.io/nw2tj/?view_only=9f62bbab141841808118ca6339802bc4.

Introduction

Digital interfaces shape billions of daily decisions, yet not all design choices serve user interests. Dark patterns (the term encompasses manipulative and deceptive interface design practices that exploit cognitive biases to steer users toward behaviors benefiting service providers over user interests) have proliferated across digital platforms (Gray et al. 2018; Mathur et al. 2019). Recent regulatory attention, including explicit prohibition in the EU Digital Services Act (European Commission 2022) and enforcement actions (Federal Trade Commission 2025), underscores their societal significance.

Despite growing documentation of dark pattern taxonomies (Gray et al. 2023; Mathur et al. 2021) and their behavioral effects (Luguri and Strahilevitz 2021), fundamental questions about user perception remain contested. How do users perceive interfaces designed to manipulate them? Do manipulative designs produce coherent psychological experiences?

And critically, how do perceptions of deception relate to other dimensions of user experience?

Existing research presents a puzzle: dark patterns demonstrably influence behavior (Luguri and Strahilevitz 2021), yet standard user experience metrics often fail to identify them as problematic (Bongard-Blanchy et al. 2021). This disconnect raises a crucial question: are users genuinely unaware of manipulation, or do traditional UX frameworks simply fail to capture manipulation awareness?

We propose that users possess tacit awareness of manipulation that becomes articulable when provided appropriate vocabulary. This awareness may manifest not through conscious deliberation but through immediate aesthetic reactions, a form of manipulation detection through aesthetic judgment. Finding deceptive interfaces unattractive (annoying, unfriendly) may serve as a rapid, affective signal of autonomy violation. Such aesthetic responses would align with research showing moral violations trigger visceral reactions (Haidt 2001) and that aesthetic judgments serve as rapid trust assessments (Lindgaard et al. 2006).

Additionally, if dark pattern taxonomies capture meaningful psychological distinctions, users should perceive patterns within the same strategic category as perceptually similar. Demonstrating within-category consistency would validate that manipulation strategies constitute stable psychological constructs, not merely researcher-imposed categorizations.

Present Study

This exploratory study examines how end users perceive dark patterns in social media when provided vocabulary (specific adjective pairs) to articulate their experiences. We investigated relationships between perceived manipulation and user experience dimensions, seeking to understand how deception perceptions relate to aesthetic judgments and whether users detect strategic coherence across different pattern implementations. Two findings emerged as particularly robust from this exploratory analysis:

1. **Deception-Attractiveness Relationship:** Do users judge deceptive interfaces as unattractive? Such a relationship would suggest manipulation awareness manifests through aesthetic reactions.
2. **Strategic Category Coherence:** Do patterns sharing manipulation strategies produce consistent user experiences? Within-category consistency would validate that manipulation strategies constitute stable psychological constructs rather than merely researcher-imposed categorizations.

We addressed these questions through a survey study in which 126 social media users evaluated 15 dark patterns spanning six strategic categories. Participants used an adapted User Experience Questionnaire (Laugwitz et al. 2008) supplemented with four autonomy-

focused adjective pairs (pressuring/suggesting, addictive/non-addictive, covert/revealed, deceptive/benevolent). This design enabled direct comparison between traditional UX dimensions and autonomy-violation constructs while examining relationships among evaluation dimensions.

Our findings suggest that when provided appropriate adjective pairs for evaluation, users detect manipulation through aesthetic judgment: correlations between deception and attractiveness ratings (annoying, unfriendly) indicate users perceive deceptive interfaces as unattractive. Additionally, certain manipulation strategies (Interface Interference, Social Engineering) constitute coherent perceptual constructs identifiable across specific implementations. These results suggest evaluation frameworks should explicitly incorporate autonomy-violation dimensions alongside traditional pragmatic and hedonic UX qualities.

Related Work

Dark Patterns: Definition and Taxonomies

Dark patterns are manipulative interface designs that exploit cognitive biases and heuristics (Kahneman 2011; Thaler and Sunstein 2009) to steer users toward behaviors that benefit service providers over user interests. Current taxonomies distinguish broad manipulation strategies (Gray et al. 2023; Mathur et al. 2021; Brignull et al. 2023), including nagging (repeated prompts), obstruction (making actions difficult), sneaking (hiding information), interface interference (manipulating visual hierarchy), forced action (requiring unwanted actions), and social engineering (exploiting social dynamics).

Each strategic category encompasses specific implementation patterns. For example, interface interference includes false hierarchy (styling preferred options prominently), trick wording (misleading labels), and emotional manipulation (guilt-inducing language). Social engineering includes gamification, social pressure, and engagement mechanics like infinite scroll. These taxonomies, while widely adopted, remain based primarily on designer intent and mechanism analysis rather than empirical validation of user perception.

User Awareness of Manipulation

Research presents conflicting evidence regarding whether users consciously detect manipulative design. Some studies suggest users fail to recognize dark patterns during interaction (Sluijpers 2022), with interface styling significantly altering decisions without awareness (Bielova et al. 2024; Schneider et al. 2018).

This apparent contradiction may reflect measurement limitations. Standard user experience frameworks assess pragmatic qualities (efficiency, perspicuity) and hedonic qualities (stimulation, novelty) (Hassenzahl and Tractinsky 2006; Laugwitz et al. 2008) but lack constructs explicitly capturing manipulation or autonomy violation. Studies

documenting behavioral effects without corresponding UX deficits (Bongard-Blanchy et al. 2021) may thus reflect not user insensitivity but rather the inadequacy of traditional UX metrics for detecting manipulation awareness.

Aesthetic Responses to Moral Violations

Psychological research demonstrates interconnections between aesthetic and moral judgments. Moral violations activate disgust responses similar to those triggered by physical contaminants (Haidt 2001), with shared neural substrates suggesting aesthetic reactions signal ethical concerns. In interface evaluation, users employ aesthetic cues when assessing trustworthiness and credibility (Lindgaard et al. 2006), with negative aesthetic reactions emerging rapidly—within 50 milliseconds of exposure.

This aesthetic-moral coupling suggests manipulation awareness might manifest through immediate aesthetic responses rather than deliberative reasoning. If users detect deception at a tacit level, this awareness could surface as negative aesthetic judgments (finding interfaces unattractive, annoying, unfriendly) before conscious articulation of the manipulation mechanism.

Taxonomic Validity

Dark pattern taxonomies organize patterns by strategic intent and manipulation mechanism (Gray et al. 2023; Mathur et al. 2021), but empirical validation of these categorizations remains limited. A valid taxonomy should correspond to meaningful psychological distinctions: patterns within the same strategic category should produce consistent user experiences, while patterns from different categories should be perceived as distinct.

Prior work has documented individual pattern effects (Luguri and Strahilevitz 2021) and cross-pattern prevalence (Mathur et al. 2019), but whether users perceive strategic coherence—recognizing patterns as variations of a shared manipulation approach—remains unexplored. Demonstrating within-category perceptual consistency would validate that manipulation strategies constitute stable psychological constructs detectable across specific implementations.

Method

Participants

A total of 143 participants completed the online survey between mid-January and mid-February 2024. After data processing, 126 participants who identified as social media users comprised the final analytical sample (53 female, 73 male; age range 19-80 years, $M = 32.4$, $SD = 12.7$). Three-quarters (75%) resided in Austria, 9% in the United States, 5% in Switzerland or Germany, and 11% elsewhere or did not report location. The majority (87%)

considered themselves at least somewhat technically savvy (56% "Yes", 31% "A bit", 13% "No").

Materials

Dark Pattern Taxonomy and Mockups

We developed a taxonomy of 15 social media-specific dark patterns organized into six strategic categories. This taxonomy synthesized existing classifications (EDPB 2023; Mildner et al. 2023; Monge Roffarello and De Russis 2022), organized considering the strategic framework established by Gray et al. (2023). The six strategic categories were: Nagging (1 pattern), Obstruction (2 patterns), Sneaking (2 patterns), Interface Interference (3 patterns), Forced Action (1 pattern), and Social Engineering (6 patterns). Table 1 presents the complete taxonomy.

Table 1. Dark Pattern Taxonomy: Six Strategic Categories and 15 Specific Patterns

Strategy	Pattern	Description
Nagging	Nagging	Repeated interruptions or prompts that persist until user complies
Obstruction	Overcomplicated Process	Unnecessarily complex procedures that discourage desired user actions
Obstruction	Hindering Account Deletion	Deliberate barriers making it difficult to close or delete accounts
Sneaking	Sneaky Bad Defaults	Pre-selected options favoring service provider over user interests
Sneaking	Expectation Result Mismatch	Interface elements that produce unexpected outcomes when clicked
Interface Interference	False Hierarchy	Visual design emphasizing undesired options over user-preferred choices
Interface Interference	Trick Wording	Ambiguous or misleading language obscuring true action consequences
Interface Interference	Toying with Emotion	Emotional manipulation through guilt, fear, or social pressure
Forced Action	Forced Access	Requiring account creation or data disclosure to access basic features
Social Engineering	Gamification	Game-like elements (points, badges) encouraging prolonged engagement
Social Engineering	Social Pressure	Exploiting social norms and peer influence to drive user behavior
Social Engineering	Social Connector	Leveraging social relationships to increase platform engagement

Strategy	Pattern	Description
Social Engineering	Content Customization	Personalized feeds optimized for engagement over user wellbeing
Social Engineering	Endlessness	Infinite scroll or autoplay removing natural stopping points
Social Engineering	Pull To Refresh	Gesture-based content updates exploiting checking behaviors

Note. Patterns synthesized from social media dark pattern classifications by EDPB (2023), Mildner et al. (2023), and Monge Roffarello and De Russis (2022), organized according to strategic framework by Gray et al. (2023).

For each pattern, we created realistic interface mockups designed to appear as authentic social media interfaces. Mockups incorporated visual elements typical of major platforms (color schemes, typography, layout patterns) while implementing the specific dark pattern mechanism. Twelve patterns were presented as static images; three patterns (Nagging, Endlessness, Pull To Refresh) required video mockups to demonstrate their temporal dynamics. All materials are publicly available through the Open Science Framework (Anonymous 2025). Further methodological detail is available in Rhomberg (2024).

Evaluation Instrument

We adapted the User Experience Questionnaire-Short (UEQ-S) (Schrepp et al. 2017), which employs semantic differential scales using opposing adjective pairs rated on 7-point scales (-3 to +3). The UEQ-S assesses four dimensions: Perspicuity (e.g., complicated/easy), Efficiency (e.g., inefficient/efficient), Dependability (e.g., unpredictable/predictable), and Stimulation (e.g., boring/exciting). We supplemented the UEQ-S with Attractiveness items from the full UEQ (e.g., annoying/enjoyable, unfriendly/friendly) to enable assessment of overall aesthetic appeal.

Critically, we introduced four autonomy-focused items designed to capture experiences absent from standard UX frameworks:

- **Pressuring / Suggesting:** assessing perceived coercion
- **Addictive / Non-addictive:** capturing compulsive engagement
- **Covert / Revealed:** measuring perceived transparency
- **Deceptive / Benevolent:** directly assessing manipulation perception

These four items were developed after creating pre-survey assumptions about how each dark pattern would be perceived. The author identified expected characteristics for each pattern (e.g., "annoying" for Nagging, "addictive" for Gamification), reviewed these expectations against the Microsoft Desirability Toolkit (Benedek and Miner 2006) to identify

suitable adjective pairs already present in the UEQ, and then created new autonomy-focused adjective pairs for characteristics not captured by existing UEQ dimensions.

The final instrument comprised 14 items: 10 from UEQ-S/UEQ organized into 5 scales (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation) and 4 autonomy-focused standalone items. See Rhomberg (2024) for additional instrument development detail.

Procedure

Participants accessed the survey through a web link distributed via social media, university mailing lists, and snowball sampling. After providing informed consent and answering demographic questions, participants were randomly assigned to evaluate 5 of the 15 dark patterns. This between-subjects randomization ensured each participant evaluated patterns from multiple strategic categories while limiting survey length.

For each assigned pattern, participants:

1. Viewed the dark pattern mockup (image or video)
2. Rated the interface using all 14 evaluation items
3. Optionally provided open-ended comments

Item presentation order was randomized within each pattern evaluation to control for order effects. The entire survey required approximately 15-20 minutes to complete. Participants were debriefed about dark patterns and provided resources for managing social media usage upon completion.

This between-subjects design yielded approximately 42 independent evaluations per pattern (126 participants \times 5 patterns each / 15 total patterns), providing adequate statistical power for pattern-level analyses and detection of medium-to-large correlational effects.

Analytical Strategy

Scale scores were computed by averaging constituent items. For autonomy-focused items assessed via single adjective pairs, individual item scores served as scale proxies. Following UEQ interpretation guidelines (Schrepp et al. 2017), we classified scores as positive (≥ 1), neutral (-1 to 1), or negative (≤ -1). This classification scheme provides context for interpreting overall pattern evaluations across all dimensions.

To address our research questions, we conducted two primary analyses:

Research Question 1 (Deception-Attractiveness Correlation): We computed Pearson correlation coefficients between Deception and Attractiveness scores across all 15 dark patterns. Pattern-level means (aggregating across participants) served as the unit of

analysis, as we were interested in whether patterns perceived as more deceptive were also perceived as less attractive.

Research Question 2 (Within-Category Consistency): For each strategic category containing multiple patterns, we calculated correlation matrices including all evaluation items. We then computed average correlation coefficients using Fisher z-transformation to appropriately aggregate correlations. This yielded a single within-category consistency coefficient representing the average correlation among patterns sharing strategic intent.

Categories with only one pattern (Nagging, Forced Action) could not be assessed for within-category consistency.

All analyses were conducted in R version 4.3.1. Raw data, analysis scripts, and supplementary materials are available through the Open Science Framework (Anonymous 2025).

Results

Descriptive Overview

Table 2 presents mean evaluation scores across all 15 dark patterns for both traditional UX dimensions and autonomy-focused items. Standard UEQ-S metrics (Perspicuity, Efficiency, Dependability, Stimulation) showed predominantly neutral ratings: 13 of 15 patterns received neutral overall UEQ-S scores (between -1 and +1), with only Nagging ($M = -1.06$) clearly negative. Excluding Attractiveness (which we added from the full UEQ specifically to assess aesthetic appeal), the traditional UEQ-S would have identified only 4 problematic patterns: 3 on Dependability and 1 on Stimulation. This limited detection contrasts sharply with autonomy-focused items, which revealed substantially more negative perceptions.

Table 2. Descriptive Statistics: Aggregated Scores Across All Dark Patterns

Scale	Min	Max	Mean	SD	Negative	Positive
<i>Traditional UX Dimensions</i>						
Attractiveness	-2.15	0.58	-0.93	0.74	6	0
Dependability	-1.13	0.36	-0.32	0.42	3	0
Efficiency	-0.82	0.85	0.10	0.48	0	0
Perspicuity	-0.98	1.45	0.26	0.63	0	3
Stimulation	-1.56	0.40	-0.37	0.52	1	0
UEQ-S Overall	-1.06	0.59	-0.17	0.41	1	0
<i>Autonomy-Focused Items</i>						
Pressure	-2.26	-0.04	-1.08	0.67	8	0
Addiction	-1.47	1.26	0.11	0.79	2	0

Scale	Min	Max	Mean	SD	Negative	Positive
Covert	-0.44	1.49	0.39	0.50	0	0
Deception	-2.09	0.07	-1.03	0.63	6	0

Note. Negative = number of patterns with $M \leq -1$; Positive = number with $M \geq 1$.

Values aggregated across $n = 15$ dark patterns; each pattern evaluated by ~ 42 participants.

Autonomy-focused items revealed more consistently negative perceptions. Pressure emerged as the most negative dimension ($M = -1.08$, $SD = 0.67$), with 8 of 15 patterns rated as pressuring (scores ≤ -1). Deception showed similarly strong negative perceptions ($M = -1.03$, $SD = 0.63$), with 6 patterns perceived as deceptive. Attractiveness showed comparable negative ratings ($M = -0.93$, $SD = 0.74$), with the same 6 patterns rated as both unattractive and deceptive.

Research Question 1: Deception and Attractiveness Correlation

Figure 1 displays the relationship between perceived deception and attractiveness across all 15 dark patterns. A strong negative correlation emerged ($r = 0.93$, $p < .001$, 95% CI [0.82, 0.97]), indicating that patterns perceived as more deceptive were consistently judged as less attractive (more annoying, more unfriendly).

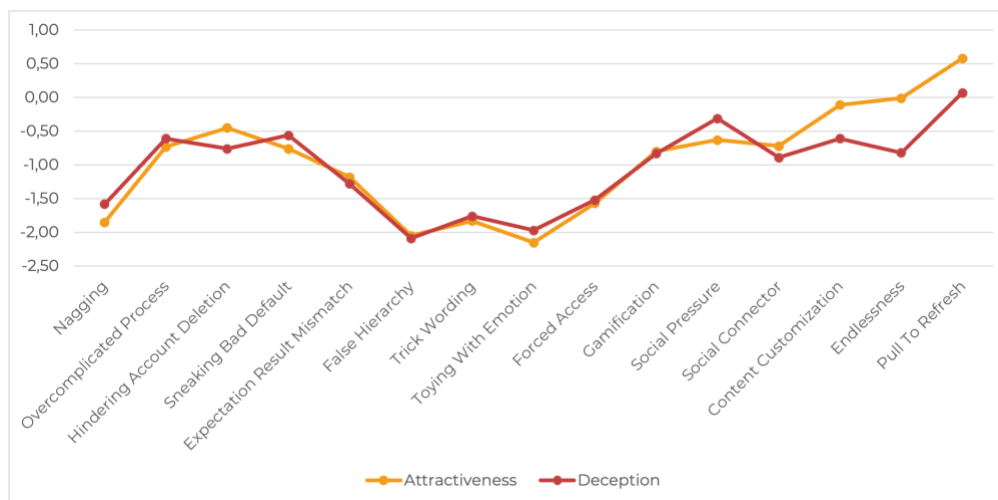


Figure 1. Correlation between perceived deception and attractiveness ratings across 15 dark patterns. The points shows mean scores across patterns for perceived deception (red) and attractiveness (yellow). The strong negative correlation ($r = 0.93$) demonstrates that users judge deceptive interfaces as unattractive (axis flipped for visualization).

The shared variance between deception and attractiveness ($r^2 = 0.86$) indicates that 86% of variability in attractiveness judgments can be explained by deception perceptions.

Pattern-Specific Observations. The three patterns with strongest deception perceptions (False Hierarchy: $M = -2.09$; Toying with Emotion: $M = -1.97$; and Trick Wording: $M =$

–1.76) all belong to the Interface Interference category. These same three patterns received the most negative attractiveness ratings (False Hierarchy: $M = -2.15$; Toying with Emotion: $M = -2.03$; Trick Wording: $M = -1.78$).

Pull To Refresh represented an exception, receiving slightly positive ratings on both dimensions (Deception: $M = 0.07$; Attractiveness: $M = 0.58$) despite high addiction ratings ($M = -1.00$).

Specificity Analysis. To determine whether this correlation was specific to deception or reflected general negativity, we examined relationships between attractiveness and other autonomy-focused items. Attractiveness showed moderate correlation with Pressure ($r = 0.68$) and weak correlation with Covert ($r = 0.31$) but negligible correlation with Addiction ($r = -0.12$).

Similarly, deception correlated moderately with Dependability ($r = 0.58$) and Stimulation ($r = 0.52$) but weakly with Efficiency ($r = 0.21$) and Perspicuity ($r = 0.15$). The deception-attractiveness relationship was notably stronger than these alternative pairings.

Research Question 2: Within-Category Consistency

Table 3 presents average within-category correlation coefficients for strategic categories containing multiple patterns. Two categories demonstrated substantial internal consistency: Interface Interference ($r = 0.77$) and Social Engineering ($r = 0.72$). These high correlations indicate that patterns sharing manipulation strategies produced remarkably similar user experiences across all evaluation dimensions.

Table 3. Within-Category Correlation Coefficients

Strategic Category	<i>n</i> Patterns	Average <i>r</i>	Interpretation
Nagging	1	—	Single pattern
Obstruction	2	-0.23	Weak consistency
Sneaking	2	0.29	Weak consistency
Interface Interference	3	0.77	Strong consistency
Forced Action	1	—	Single pattern
Social Engineering	6	0.72	Strong consistency

Note. Correlation coefficients computed using Fisher z-transformation aggregation across all pairwise pattern correlations within each category.

Interface Interference Consistency. The three Interface Interference patterns (False Hierarchy, Trick Wording, and Toying with Emotion) showed highly correlated evaluations ($r = 0.77$). Figure 2 illustrates this consistency: all three patterns were consistently rated as pressuring, deceptive, obstructive, annoying, and unfriendly.

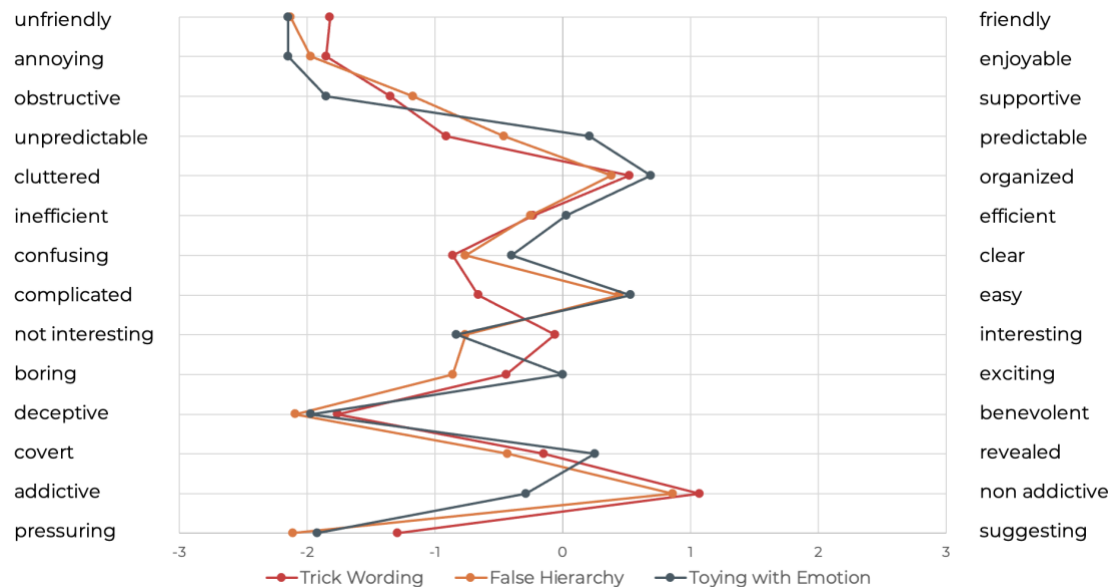


Figure 2. Evaluation profiles for the three Interface Interference dark patterns demonstrate high within-category consistency ($r = 0.77$). Despite differing surface implementations (visual hierarchy manipulation, misleading wording, emotional appeals), users perceive these patterns similarly across dimensions.

Figure 6 shows the three Interface Interference pattern mockups: False Hierarchy (manipulating button prominence and positioning), Trick Wording (using misleading labels and double negatives), and Toying with Emotion (employing guilt-inducing language). Despite these substantial implementation differences, users perceived all three patterns consistently across evaluation dimensions, indicating recognition of their shared strategic approach of cognitively manipulating user attention and emotions.

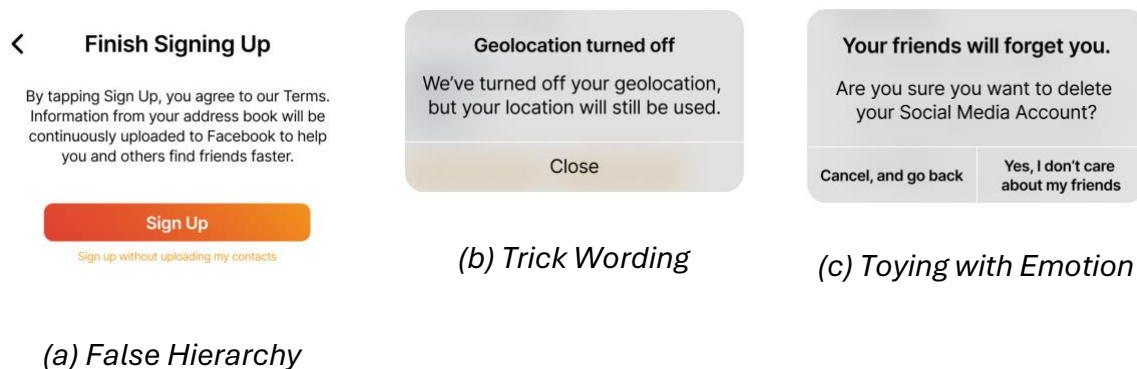


Figure 3. Interface Interference pattern mockups showing three distinct implementations of cognitive manipulation: (a) manipulating visual hierarchy to bias choice, (b) using misleading wording to confuse users, and (c) employing emotional language to pressure decisions.

Social Engineering Consistency. The six Social Engineering patterns (Gamification, Social Pressure, Social Connector, Content Customization, Endlessness, and Pull To Refresh)

showed strong consistency ($r = 0.72$). These patterns leverage social proof, fear of missing out (FOMO), and engagement hooks to influence behavior. Unlike Interface Interference's uniformly negative profile, Social Engineering patterns showed a mixed profile: users rated them as easy, clear, and efficient (positive UX dimensions) while simultaneously rating them as pressuring (negative autonomy dimension).

Obstruction Category. The Obstruction category showed negative correlation ($r = -0.23$). These two patterns (Overcomplicated Process and Hindering Account Deletion) both obstruct user goals but through different mechanisms: Overcomplicated Process adds unnecessary steps and confusing language; Hindering Account Deletion imposes waiting periods and confirmation screens.

Sneaking Category. The two Sneaking patterns showed weak consistency ($r = 0.29$). Sneaky Bad Defaults and Expectation Result Mismatch both involve hiding information or presenting misleading expectations, but users did not evaluate them as consistently as Interface Interference or Social Engineering patterns.

Exploratory Findings

Autonomy-Focused Item Patterns. Beyond the two main research questions, the autonomy-focused items revealed interpretable patterns. Pressure emerged as the most universal negative dimension, with 8 of 15 patterns rated as pressuring. Deception showed more selective application, concentrated in Interface Interference patterns. Addiction applied specifically to Social Engineering patterns designed for prolonged engagement (Endlessness, Pull To Refresh, Gamification).

Covert ratings centered near neutral ($M = 0.39$), suggesting most patterns operated neither particularly hidden nor fully transparent. This may reflect the mockup presentation format: participants evaluated patterns individually rather than encountering them embedded within functional interfaces where covert characteristics might be more salient.

Cross-Cutting Findings. Forced Access (requiring contact upload to proceed) received the worst overall evaluation ($M = -0.93$) and most negative ratings on Covert ($M = -0.44$), Deception ($M = -1.52$), and Pressure ($M = -1.59$). This pattern violated multiple autonomy dimensions simultaneously.

Conversely, Pull To Refresh received positive evaluations across most dimensions (overall attractiveness: $M = 0.58$) despite high addiction ratings ($M = -1.00$).

Discussion

This study provides correlational evidence that users detect manipulation through aesthetic judgment: interfaces perceived as deceptive are consistently judged as unattractive (annoying, unfriendly). Additionally, certain dark pattern strategies (Interface Interference and Social Engineering) constitute coherent perceptual constructs, while others (Obstruction, Sneaking) do not. These findings extend psychological understanding

of user experience beyond traditional pragmatic and hedonic dimensions to include autonomy violations as distinct experiential qualities.

Manipulation Detection Through Aesthetic Judgment

The strong correlation between perceived deception and attractiveness ($r = 0.93$) demonstrates that manipulation awareness manifests through immediate aesthetic reactions rather than conscious deliberation. Users judge deceptive interfaces as annoying and unfriendly (the two adjective pairs comprising the UEQ Attractiveness scale (Laugwitz et al. 2008), designed to capture immediate, affective reactions to interface appeal). These negative aesthetic responses may serve as rapid, affective signals of autonomy violation.

This finding aligns with moral psychology research showing intertwined aesthetic and moral judgments (Haidt 2001), extending this relationship to digital interface design. Users don't merely judge deceptive interfaces as instrumentally problematic; they experience visceral aesthetic aversion. This suggests manipulation detection operates partially below conscious awareness, triggering rapid aesthetic responses similar to disgust reactions to moral violations (Chapman and Anderson 2013).

The specificity of this relationship supports the aesthetic judgment interpretation. Deception correlated more strongly with attractiveness than other autonomy-focused dimensions (pressure, addiction), suggesting particular sensitivity to truthfulness in interface communication. This aligns with research on trust calibration (McKnight et al. 2011), where deception detection serves as critical input to trust formation. Aesthetic reactions may function as rapid trust assessments; "unattractive" serving as a heuristic signal for "untrustworthy."

Strategic Coherence: When Users Detect Manipulation Intent

Within-category correlations revealed that certain (but not all) manipulation strategies constitute coherent perceptual constructs. Interface Interference ($r = 0.77$) and Social Engineering ($r = 0.72$) showed strong consistency, while Obstruction ($r = -0.23$) and Sneaking ($r = 0.29$) did not. This selective pattern provides important insights into which manipulation strategies users detect as unified approaches versus perceiving as distinct implementations.

Interface Interference as Cognitive Manipulation. The three Interface Interference patterns (False Hierarchy, Trick Wording, and Toying with Emotion) produced highly consistent user experiences despite substantial implementation differences. False Hierarchy manipulates button prominence; Trick Wording uses misleading labels; Toying with Emotion employs guilt-inducing language. Yet users recognized these as variations of a shared manipulation approach: exploiting visual hierarchy and emotional framing to bias information processing.

This coherence indicates users detect the underlying strategic intent (cognitive manipulation at perceptual or emotional levels) rather than merely responding to surface features. The uniformly negative profile across all evaluation dimensions (pressuring, deceptive, annoying, unfriendly) suggests users perceive these patterns as particularly egregious autonomy violations.

Social Engineering's Mixed Profile. Social Engineering patterns showed comparable coherence despite greater implementation diversity (gamification, social pressure, infinite scroll, pull-to-refresh). These patterns share strategic reliance on social proof, fear of missing out (FOMO), and engagement mechanics.

Critically, Social Engineering patterns exhibited a mixed evaluation profile absent in Interface Interference: users rated them positively on traditional UX dimensions (easy, clear, efficient) while simultaneously rating them negatively on autonomy dimensions (pressuring). This duality illuminates why such patterns persist; they deliver genuine usability benefits (infinite scroll eliminates pagination friction) while manipulating engagement through addictive mechanics. Users detect both aspects: they appreciate the functionality while recognizing the manipulation.

Failed Coherence as Informative. The negative correlation for Obstruction and weak correlation for Sneaking provide valuable information. These categorizations, while logical from designer intent perspectives, do not correspond to coherent user experiences. Users perceived Overcomplicated Process and Hindering Account Deletion as distinct frustrations rather than instances of unified manipulation strategy, despite both obstructing user goals.

This selective coherence validates our analytical approach—positive correlations reflect genuine perceptual consistency rather than methodological artifacts. It also suggests that not all designer-intent taxonomies align with user psychology. Future taxonomy development should incorporate empirical validation of perceptual coherence alongside mechanism analysis.

Tacit Awareness Enabled by Vocabulary Provision

These findings challenge the assumption that users are "blind" to dark patterns (Sluijpers 2022). Our results demonstrate users clearly perceive manipulation when provided appropriate vocabulary; they rate patterns as deceptive, pressuring, and unattractive. However, this awareness doesn't necessarily translate to behavioral resistance, highlighting the distinction between perceptual awareness and behavioral protection.

Prior work documenting behavioral effects without corresponding UX deficits (Bongard-Blanchy et al. 2021) may reflect measurement limitations rather than user insensitivity. Standard UX frameworks emphasize dimensions on which many dark patterns perform well: Social Engineering patterns averaged positive ratings on Perspicuity and Efficiency. Traditional metrics thus systematically miss the autonomy-violation dimension that becomes evident when users have vocabulary to articulate it.

Our autonomy-focused items revealed what standard UX measures obscured. The strong deception-attractiveness correlation emerged only because we provided evaluative vocabulary (deceptive/benevolent, annoying/enjoyable, unfriendly/friendly) enabling users to articulate tacit awareness. Without such vocabulary, this awareness might manifest only as vague unease insufficient to overcome behavioral manipulation mechanisms.

This vocabulary-provision framework reconciles apparent contradictions in the literature. Users possess tacit awareness of manipulation operating through affective channels (aesthetic reactions) but may lack conceptual vocabulary to articulate this awareness without appropriate prompting. Providing autonomy-focused evaluation dimensions transforms tacit affective responses into articulable judgments, making manipulation awareness measurable and communicable.

Practical Implications

For UX Evaluation. Our findings demonstrate that traditional UX frameworks systematically fail to capture manipulation dimensions detectable through aesthetic judgment. The UEQ-S (a well-validated instrument (Schrepp et al. 2017)) identified only 1 of 15 dark patterns as problematic overall, despite users clearly detecting manipulation when provided autonomy-focused vocabulary.

We recommend augmenting standard UX evaluations with autonomy-focused constructs. Minimally, evaluations should assess: (1) perceived coercion (pressuring/suggesting), (2) transparency (covert/revealed), and (3) deception (deceptive/benevolent). These additions would enable practitioners to identify manipulation detectable through aesthetic judgment (designs users perceive as annoying and unfriendly despite functional adequacy) that traditional metrics miss entirely.

For Regulation. The strong deception-attractiveness correlation provides an objective, user-perception-based metric for regulatory assessment. Designs producing dissociations between functionality and attractiveness (rated as efficient yet unattractive, annoying, unfriendly) warrant scrutiny as potentially manipulative. This approach grounds regulation in demonstrated user perception rather than expert judgment alone.

Strategic coherence findings inform regulatory scope. Rather than prohibiting individual patterns, regulations might target coherent manipulation strategies (Interface Interference, Social Engineering). Our evidence that users perceive these categories as unified approaches suggests strategy-based prohibitions would be psychologically meaningful to end users, not merely regulatory abstractions.

For Design Education. The manipulation detection through aesthetic judgment finding suggests design education should explicitly connect aesthetics and ethics. "Good design" encompasses not only usability and visual appeal but also respect for user autonomy. Teaching designers that manipulation manifests as aesthetic unpleasantness (users find deceptive interfaces annoying and unfriendly) may help internalize ethical considerations as intrinsic to aesthetic quality rather than external constraints.

Limitations and Future Directions

Several limitations warrant consideration. First, our study employed mockups rather than functional interfaces. While mockups enabled controlled comparison across patterns, they may not fully capture manipulation effects emerging through prolonged interaction (e.g., addiction to infinite scroll) or how covert characteristics manifest in functional contexts. The neutral Covert ratings ($M = 0.39$) may reflect this limitation: participants evaluated patterns individually rather than encountering them embedded within interfaces where hidden mechanics might be more salient.

Second, our sample skewed toward technically savvy users (87% self-reported technical savvy), potentially overestimating manipulation awareness in general populations. Research on vulnerable populations—children, elderly adults, individuals with lower technical literacy—may reveal different patterns. However, if technically savvy users detect manipulation through aesthetic judgment, less-savvy users likely experience similar affective responses, perhaps with reduced ability to articulate them even when provided vocabulary.

Third, we examined social media dark patterns specifically. Strategic coherence may differ in other domains (e-commerce, health apps, financial services) where different manipulation strategies dominate. Whether manipulation detection through aesthetic judgment generalizes across domains or reflects social media-specific dynamics remains an empirical question requiring cross-domain research.

Fourth, our correlational design cannot establish causality. While we interpret the deception-attractiveness correlation as indicating aesthetic reactions to perceived manipulation, reverse causation remains possible: perhaps unattractive designs trigger suspicion. Experimental manipulations systematically varying deception while controlling aesthetics would test directional hypotheses about whether manipulation drives aesthetic responses or vice versa.

Pull To Refresh and Normalization. Pull To Refresh represented an intriguing exception, receiving slightly positive ratings on both deception ($M = 0.07$) and attractiveness ($M = 0.58$) despite high addiction ratings ($M = -1.00$). This ubiquitous pattern (now standard across mobile interfaces) may have achieved normalization: users no longer detect its addictive mechanics as deceptive, despite recognizing addiction. Longitudinal research examining whether novel manipulations trigger stronger aesthetic aversion that diminishes over time would clarify normalization processes and inform proactive regulation before patterns become entrenched.

Individual Differences. Future research should explore individual differences in manipulation detection through aesthetic judgment. Do certain personality traits (need for cognition, trust propensity, aesthetic sensitivity) predict stronger deception-unattractiveness associations? Can training enhance manipulation detection or vocabulary for articulation? Understanding individual variation could inform targeted interventions for vulnerable populations.

Behavioral Validation. Finally, behavioral validation remains critical. Our findings demonstrate perceptual awareness through aesthetic judgment but don't establish whether this awareness predicts behavioral resistance. Research examining relationships between manipulation perception, aesthetic reactions, and actual behavior (click-through rates, choice patterns, engagement duration) would clarify whether aesthetic-based detection provides meaningful protection or merely reflects post-hoc rationalization of manipulated behavior.

Theoretical Contributions

This work extends psychological theory in several ways. First, we demonstrate that manipulation awareness constitutes a distinct dimension of user experience, irreducible to traditional pragmatic and hedonic qualities. Users can perceive interfaces as efficient yet deceptive, enjoyable yet pressuring—combinations impossible under frameworks collapsing evaluation into single "user experience" constructs.

Second, we provide evidence that aesthetic reactions signal moral judgment in digital contexts. The aesthetic-moral coupling documented in moral psychology (Haidt 2001) extends to interface evaluation, suggesting humans employ aesthetic responses as rapid ethical assessments across diverse domains.

Third, we validate user perception of manipulation strategies as coherent psychological constructs. Strategic coherence demonstrates that users possess tacit knowledge about manipulation intent and mechanics that organizes their experiences beyond surface features. This finding supports treating manipulation strategies as genuine psychological phenomena rather than researcher conveniences.

Conclusion

Dark patterns exploit cognitive biases to influence behavior, but our findings demonstrate users detect this manipulation through aesthetic judgment when provided vocabulary to articulate their awareness. Deceptive interfaces are perceived as unattractive—annoying and unfriendly—suggesting manipulation awareness operates through rapid affective responses rather than conscious deliberation alone.

Moreover, users detect certain manipulation strategies (Interface Interference, Social Engineering) as coherent psychological constructs, recognizing patterns sharing manipulative intent despite implementation differences. This strategic awareness emerges selectively: not all designer-intent categorizations correspond to coherent user experiences, highlighting the importance of empirical validation in taxonomy development.

These findings have immediate practical implications. Traditional UX frameworks systematically fail to capture manipulation detectable through aesthetic judgment. Augmenting evaluation with autonomy-focused vocabulary enables users to articulate tacit awareness, revealing negative perceptions invisible to standard metrics. For regulators, user-perceived unattractiveness despite functional adequacy provides

objective evidence of potential manipulation. For researchers, our results challenge "user blindness" assumptions and demonstrate that manipulation awareness can manifest through affective channels when appropriate evaluative vocabulary is provided.

Ultimately, this work contributes to a more complete psychological account of user experience—one recognizing that aesthetic judgment serves not only as evaluation of visual appeal but also as detection mechanism for autonomy violations. Users perceive deceptive interfaces as unattractive through a form of aesthetic-moral coupling, and this perception becomes articulable when provided appropriate vocabulary, even as behavioral systems may remain susceptible to manipulation mechanics.

References

Anonymous. 2025. *Data and Materials for: Ethics-Focused User Metrics: How User Experience Metrics Can Trigger Ethical Design Sensibility*.

https://osf.io/nw2tj/?view_only=9f62bbab141841808118ca6339802bc4.

Benedek, Joey, and Trish Miner. 2006. *Product Reaction Cards*.

Bielova, Nataliia, Laura Litvine, Anyisia Nguyen, Mariam Chammat, Vincent Toubiana, and Estelle Harry. 2024. "The Effect of Design Patterns on (Present and Future) Cookie Consent Decisions." *USENIX Security Symposium*. USENIX Association. *Accepted for Publication*.

Bongard-Blanchy, Kerstin, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "'I Am Definitely Manipulated, Even When i Am Aware of It. It's Ridiculous!' - Dark Patterns from the End-User Perspective." *Designing Interactive Systems Conference 2021* (New York, NY, USA), June.

<https://doi.org/10.1145/3461778.3462086>.

Brignull, H, M Leiser, C Santos, and K Doshi. 2023. *Deceptive patterns – user interfaces designed to trick you*. <https://www.deceptive.design/>.

Chapman, Hanah A, and Adam K Anderson. 2013. "The Evolution of Disgust." *Psychological Science* 24 (11): 2321–29.

EDPB. 2023. *Guidelines 03/2022 on Deceptive Design Patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them. Version 2.0* European Data Protection Board. https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032022-deceptive-design-patterns-social-media_en.

European Commission. 2022. *Digital Services Act*. <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>.

Federal Trade Commission. 2025. *FTC Settlement on Dark Patterns*.

Gray, Colin M, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. 2018. "The Dark (Patterns) Side of UX Design." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14.

Gray, Colin, Cristiana Santos, and Nataliia Bielova. 2023. *Towards a Preliminary Ontology of Dark Patterns Knowledge*. <https://doi.org/10.1145/3544549.3585676>.

Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (4): 814–34. <https://doi.org/10.1037/0033-295x.108.4.814>.

Hassenzahl, Marc, and Noam Tractinsky. 2006. "User Experience-a Research Agenda." *Behaviour & Information Technology* 25: 91–97.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Farrar, Straus; Giroux.

Laugwitz, Bettina, Theo Held, and Martin Schrepp. 2008. "Construction and Evaluation of a User Experience Questionnaire." In *HCI and Usability for Education and Work*, edited by Andreas Holzinger. Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-540-89350-9_6.

Lindgaard, Gitte, Gary Fernandes, Cathy Dudek, and Judith Brown. 2006. "Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression!" *Behaviour & Information Technology* 25 (2): 115–26.

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

Mathur, Arunesh, Gunes Acar, Michael J Friedman, et al. 2019. "Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites." *Proceedings of the ACM on Human-Computer Interaction* 3: 1–32.

Mathur, Arunesh, Mihir Kshirsagar, and Jonathan Mayer. 2021. "What Makes a Dark Pattern... Dark?: Design Attributes, Normative Considerations, and Measurement Methods." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan), May, 1–18. <https://doi.org/10.1145/3411764.3445610>.

McKnight, D Harrison, Michelle Carter, Jason Bennett Thatcher, and Paul F Clay. 2011. "Trust in a Specific Technology: An Investigation of Its Components and Measures." *ACM Transactions on Management Information Systems (TMIS)* 2 (2): 1–25.

Mildner, Thomas, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. "About Engaging and Governing Strategies: A Thematic Analysis of Dark Patterns in Social Networking Services." *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA), CHI '23, April, 1–15. <https://doi.org/10.1145/3544548.3580695>.

Monge Roffarello, Alberto, and Luigi De Russis. 2022. "Towards Understanding the Dark Patterns That Steal Our Attention." *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA), CHI EA '22, April, 1–7. <https://doi.org/10.1145/3491101.3519829>.

Rhomberg, Doris. 2024. "Quantifying Ethical UX: User Perceptions of Dark Patterns on Social Media." Master's Thesis, TU Wien.

Schneider, Christoph, Markus Weinmann, and Jan vom Brocke. 2018. "Digital Nudging: Guiding Online User Choices Through Interface Design." *Communications of the ACM* 61 (7): 67–73. <https://doi.org/10.1145/3213765>.

Schrepp, Martin, Andreas Hinderks, and Jörg Thomaschewski. 2017. "Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)." *International Journal of Interactive Multimedia and Artificial Intelligence* 4 (January): 103. <https://doi.org/10.9781/ijimai.2017.09.001>.

Sluijpers, M. 2022. *Investigating Dark Pattern Blindness*. <https://studenttheses.uu.nl/handle/20.500.12932/41972>.

Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Penguin Books.