

Exploatering genom design

En undersökning av Dark Patterns förekomst på mobilapplikationer för sociala medieplattformer

Mathias Erdin

Linda Schönfeldt

Institutionen för data-
och systemvetenskap

Examensarbete 15 hp

Data- och systemvetenskap

Vårterminen 2024

Handledare: Lena Norberg

English title: Exploitation by Design - An Investigation of the
Prevalence of Dark Patterns in Mobile Applications for Social Media
Platforms



Stockholms
universitet

Sammanfattning

Dark Patterns är designmönster som är ämnade att manipulera användare och de används av företag inom flertalet områden. Tidigare forskning har huvudsakligen fokuserat på e-handel, men på senare år har andra domäner utforskats, däribland sociala medieplattformar. Denna studie ämnar undersöka hur *Dark Patterns* förekommer på mobilapplikationerna för de nio största sociala medieplattformarna i Sverige. Studien har tre frågeställningar: Vilka typer av *Dark Patterns* förekommer på nio av de mest populära SNS-mobilapplikationerna i Sverige? Med vilka övertygande argument försöker designen påverka användaren med *Dark Patterns*? Vad har denna påverkan för effekt på användares autonomi?

Studien utförs via kartläggning och använder sig av dokumenstudie för att dokumentera och analysera insamlade data med hjälp av tematisk analys. Till hjälp för detta används en tidigare definierad ontologi gällande *Dark Patterns*. Författarna finner 34 unika *Dark Patterns*, varav 13 formuleras av dem. Det finns flera olika sätt som dessa *Dark Patterns* försöker övertyga och manipulera användare. Exempel på övertygande argument som används är att användaren avråds från vissa handlingsalternativ medan andra favoriseras av designen. Användarens autonomi undergrävs sedermera när deras beslut guidas och försvåras; deras uppmärksamhet fångas och håller dem kvar på plattformen; eller när deras psykologiska sårbarheter på annat sätt exploateras för systemets vinning.

Nyckelord

Dark Patterns, *Social Networking Services*, Sociala medieplattformar, Övertygande design, Autonomi

Abstract

Dark Patterns are design patterns intended to manipulate users and are used by companies in various fields. Previous research has primarily focused on e-commerce, but in recent years other domains have been explored, including social media platforms. This study aims to investigate how Dark Patterns appear in the mobile applications of the nine largest social media platforms in Sweden. The study has three research questions: What types of Dark Patterns appear in nine of the most popular SNS mobile applications in Sweden? With what persuasive arguments does the design attempt to influence the user with Dark Patterns? What effect does this influence have on users' autonomy?

The study is conducted through a survey and uses document analysis to document and analyze the collected data with the help of thematic analysis. For this purpose, a previously defined ontology regarding Dark Patterns is used. The authors find 34 unique Dark Patterns, 13 of which they formulate themselves. There are several ways in which these Dark Patterns attempt to persuade and manipulate users. Examples of persuasive arguments used include discouraging users from certain actions while others are favored by the design. The user's autonomy is subsequently undermined when their decisions are guided and made more difficult; their attention is captured and kept on the platform; or when their psychological vulnerabilities are otherwise exploited for the system's gain.

Keywords

Dark Patterns, Social Networking Services, Social media, Persuasive design, Autonomy

Synopsis

<i>Bakgrund</i>	<i>Dark Patterns</i> utgör skadliga designmönster som skapats för att lura användare. Sedan introduktionen av begreppet har olika domäner utforskats.
<i>Problem</i>	<i>Dark Patterns</i> på sociala medieplattformar har börjat uppmärksammas de senaste åren men ännu saknas forskning på vilka som förekommer där. Om all design innehåller övertygande argument vilken påverkan har <i>Dark Patterns</i> på användare av mobilapplikationer för sociala medier?
<i>Frågeställningar</i>	Vilka typer av <i>Dark Patterns</i> förekommer på de mest populära SNS-mobilapplikationerna i Sverige? Med vilka övertygande argument försöker designen påverka användaren med <i>Dark Patterns</i> ? Vad har denna påverkan för effekt på användares autonomi?
<i>Metod</i>	Studien använder sig av kartläggning som forskningsstrategi tillsammans med dokumentstudie för att samla in data från sociala medieapplikationer i form av mobilapplikationer. Detta görs med hjälp av skärminspelning för att kunna undersöka förekomsten av <i>Dark Patterns</i> . Analysmetoden som tillämpas är tematisk analys.
<i>Resultat</i>	Under analysen identifierades 34 unika <i>Dark Patterns</i> , varav 13 formulerades av författarna. Som del av denna process identifierades de övertygande argument som används av dem för att manipulera användare och därav effekten de har på deras autonomi.
<i>Diskussion</i>	De <i>Dark Patterns</i> som identifierats på sociala medier tjänar till att exploatera användare på många olika sätt. De inskränker deras autonomi genom att guida deras handlingar. Möjligheten finns för dessa företag att i stället anamma design som förebygger förekomsten av <i>Dark Patterns</i> .

Tack

Vi vill rikta ett stort tack till vår handledare Lena Norberg som har varit till stor hjälp under den här processen. Vi vill även tacka familj och vänner för all stöttning.

Innehåll

1	Introduktion	1
1.1	Bakgrund	1
1.2	Problem	2
1.3	Frågeställningar	2
2	Utökad bakgrund	3
2.1	Etiska problem gällande teknologi och design	3
2.2	Beroendeframkallande design av SNS	3
2.2.1	Bhargava och Velasquez tre argument	4
2.3	Effekter av beroendeframkallande design	5
2.4	Forskning om digitalt välmående	5
2.4.1	Designverktyg för digitalt välmående	5
2.5	Definitionen av Dark Patterns	6
2.5.1	Vald definition av Dark Patterns	6
2.6	Taxonomier av Dark Patterns	6
2.6.1	Vald ontologi av Dark Patterns	7
3	Metod	9
3.1	Metodval	9
3.1.1	Forskningsstrategi	9
3.1.2	Datainsamlingsmetod	9
3.1.3	Analysmetod	10
3.1.4	Forskningsetiska aspekter	11
3.2	Metodtillämpning	11
3.2.1	Urval	11
3.2.2	Genomförande	11
3.2.3	Protokoll	12
3.2.4	Dokumentstudie	12
3.2.5	Tematisk analys	13
3.2.6	Forskningsetiska aspekter	14
4	Resultat och analys	15
4.1	Förekomsten av Dark Patterns på nio av de mest populära SNS- mobilapplikationerna i Sverige	15
4.1.1	Sneaking	15
4.1.2	Obstruction	16
4.1.3	Interface Interference	18
4.1.4	Forced Action	20
4.1.5	Social Engineering	23

4.1.6	Odefinierad hög- eller mesonivå	24
4.2	Analys	24
5	Diskussion och slutsatser	26
5.1	Diskussion.....	26
5.2	Slutsatser.....	27
5.3	Begränsningar	28
5.4	Framtida forskning	28
5.5	Resultatetiska konsekvenser.....	28
5.6	Tekniska verktyg.....	29
	Källförteckning.....	30
	Bilaga A – Vald ontologi	33
	Bilaga B – Funna Dark Patterns.....	35

Figurer

Figur 1 Visualisering av de olika typerna av funna Dark Patterns	15
Figur 2 Visualisering av högnivån Sneaking med samtliga funna meso- och lågnivåer	16
Figur 3 Exempel på lågnivån Disguised Ads på TikTok	16
Figur 4 Visualisering av högnivån Obstruction med samtliga funna meso- och lågnivåer	17
Figur 5 Visualisering av högnivån Interface Interference med samtliga funna meso- och lågnivåer	18
Figur 6 Exempel på lågnivån Cuteness på Snapchat	19
Figur 7 Visualisering av högnivån Forced Action med samtliga funna meso- och lågnivåer	21
Figur 8 Exempel på mesonivån Unnecessary Onboarding Steps på X	22
Figur 9 Visualisering av högnivån Social Engineering med samtliga funna meso- och lågnivåer	23

Tabeller

Tabell 1 Protokoll med uppgifter att utföra under datainsamlingen

12

Förkortningar

DP - Dark Patterns

SNS - Social Networking Services

1 Introduktion

Så många som 85 procent av alla internetanvändare i Sverige använder sig dagligen av sociala medieplattformar, även kallade Social Networking Services (SNS). För barn och ungdomar (8–19 år) ligger samma siffra på 92 procent (Andersson et al., 2023). Dagens smarta enheter och SNS kan lätt bli distraherande element i människors liv. Distraktioner som i längden kan störa dagliga aktiviteter som studier och arbete samt leda till sämre produktivitet (Mark et al., 2015, refererad i Monge Roffarello et al., 2023) och ökad stress (Mark et al., 2014, refererad i Monge Roffarello et al., 2023). Det är ingen hemlighet att tech-företag utnyttjar psykologiska sårbarheter för att öka användares besöksfrekvens och vistelsetid (Monge Roffarello et al., 2023) men hur går detta till?

1.1 Bakgrund

Begreppet *Dark Patterns* (DP) myntades initialt av Harry Brignull år 2010 (Gray et al., 2023b). Den ursprungliga definitionen löd “a user interface that has been carefully crafted to trick users into doing things, such as buying insurance with their purchase or signing up for recurring bills” (Brignull, 2023, s. 5). Definitionen har främst berört e-handelsplattformar (Mildner et al., 2023a), men har vidgats och innefattar nu designmönster som, oavsett företagets intentioner, undergräver och försämrar användares autonomi, beslutsförmåga och valfrihet (CPRA, 2022). SNS förmåga att fånga användares uppmärksamhet och hålla dem kvar är således ingen slump utan medvetna designbeslut. Något som är särskilt förekommande i SNS-mobilapplikationer (Monge Roffarello & De Russis, 2023).

Utöver forskningen om DP finns ett växande forskningsområde om digitalt välmående (eng. *digital wellbeing*). Vissa användningsområden och vissa användare av SNS har visat sig vara mer mottagliga för negativa konsekvenser när det kommer till psykisk hälsa och välmående. Anledningen till detta kan vara att användare har svårt att bedöma hur ofta och länge de använder SNS. Denna diskrepans mellan självupplevd och faktisk tid på SNS kan tyda på bristande självkontroll och självbestämmande (Mildner et al., 2023b).

I den samtida uppmärksamhets-ekonomin (eng. *attention economy*) tävlar tech-bolag med sina digitala produkter om användares begränsade uppmärksamhet (Kane, 2019). De använder bland annat notifikationer och automatisk uppspelning av video för att fånga och hålla användares uppmärksamhet på SNS och lura dem till att ge upp sin upplevda handlingsförmåga (eng. *agency*) (Lukoff et al., 2021). Ständig åtkomst till SNS via mobila gränssnitt kan leda till *zone states* – ett fenomen av normativ dissociation där användaren omedvetet absorberar aktiviteter utan inneboende värde för dem (Baughan et al., 2022).

Om design har en inneboende *övertygande* natur, kan design av teknologi ses handla om att övertyga användare att göra något. Människor är vana att tänka på detta sätt om design när det handlar om reklam och hur sådan påverkar hur vi handlar. Det är dock inte lika klart hur design påverkar oss på andra sätt när det kommer till hur vi tänker och lever (Redström, 2006). Den amerikanska designteorikern Buchanan uttrycker det:” Design is an art of thought directed to practical action through the persuasiveness of objects and, therefore, design involves the vivid expression of competing ideas about social life” (1989, s. 94).

En genomgång av DPs historia, nutid och framtid av Narayanan et al. (2020) talar för att det finns tre olika tillgångar att hämta hos användare: pengar, data och uppmärksamhet. Då tidigare forskning om DP främst fokuserat på domäner som skadar användares ekonomi eller integritet (Monge Roffarello et al., 2023) behövs mer forskning på SNS och vilka DP som används där.

1.2 Problem

Under nutida förhållanden, där användares psykologiska sårbarheter utnyttjas av SNS-företag för att förleda och avleda deras uppmärksamhet och intentioner, är problemet som ligger till grund för studien att DPs förekommer i SNS-mobilapplikationer som dagligen används av många människor och att DPs har visat sig påverka människor när de använder SNS samt att människor ofta är omedvetna om att de manipuleras av att dessa DPs är designade för att påverka människors beteenden.

1.3 Frågeställningar

Vilka typer av Dark Patterns förekommer på de mest populära SNS-mobilapplikationerna i Sverige?

Med vilka övertygande argument försöker designen påverka användaren med Dark Patterns?

Vad har denna påverkan för effekt på användares autonomi?

2 Utökad bakgrund

Här presenteras utökad bakgrund för etiska problem gällande design, särskilt vad gäller beroendeframkallande design och dess effekter. Vidare nämns forskningsområdet digitalt välmående och hur design kan användas för att stärka användares autonomi och välmående. Slutligen diskuteras begreppet *Dark Patterns* (DP), existerande taxonomier, samt den definition och ontologi som studien valt att utgå ifrån.

2.1 Etiska problem gällande teknologi och design

Om design av teknologi ses kunna påverka människor till handling innehåller den således *övertygande argument* angående hur och var sådana artefakter ska användas. Detta innebär följaktligen att design är normativ eftersom den inte bara erbjuder utan *förespråkar* hur användning ska ske (Redström, 2006). Som Dunne (2005, s. 21) uttrycker det:

This enslavement is not, strictly speaking, to machines, nor to the people who build and own them, but to the conceptual models, values, and systems of thought the machines embody. User-friendliness helps to naturalise electronic objects and the values they embody. For example, while using electronic objects the use is constrained by the simple generalized model of a user these objects are designed around: the more time we spend using them the more time we spend as a caricature.

Innebär detta att *all* design är av övertygande natur? Om så är fallet skulle effekten snabbt bli urvattnad med dagens utbud. All påverkan av tankar och agerande behöver således inte vara övertygande. Vissa hävdar att intention från designern krävs men oavsett intention kan design ses som övertygande. Ta regn-paraplyet som exempel – även om designern inte kan påverka regnet finns en inneboende intention om hur användaren ska agera, och potentiellt tänka, när det regnar. Det är dock inte säkert att användare godtar dessa inneboende intentioner. I stället pågår en *dialog* mellan designer och användare där designern föreslår något via sin design och användaren gör valet att godta, modifiera eller neka förslaget. Att definiera övertygande design tycks därmed vara en svår men fundamental uppgift (Redström, 2006).

2.2 Beroendeframkallande design av SNS

De flesta företag livnär sig på användare av sin produkt eller tjänst – men det finns en annan typ av företag, så kallad *attention-economy business*. I stället för att livnära sig direkt på sina användare, kommer deras inkomst från företag (ofta reklamföretag) som köper deras användares uppmärksamhet. Den *annons- och uppmärksamhetsbaserade affärsmodellen* är inget nytt fenomen utan har existerat länge för TV, tidningar, radio (Wu, 2016), men sociala medie-företag är de mest inflytelserika och framgångsrika av sådana företag (PwC, 2018).

Bhargava och Velasquez (2021) menar att det finns tre återkommande element som SNS-företag använder. Det första *intermittent variable rewards* är belöningar som ges på ett oregelbundet sätt för att maximera användarens engagemang. Ett exempel som lyfts är hur SNS kan använda laddningsskärmar och liknande fördröjningar för att generera belöningar. Det andra elementet är hur SNS anammar design som spelar på människors behov av social validering och gemenskap. Exempelvis Facebooks gilla-knapp med vilken användare kan markera att de gillar ett inlägg av någon

annan. Det tredje elementet är design som eroderar naturliga stoppsignaler, av vilket det primära exemplet är *infinite scrolls*. Innan dessa fanns betydde slutet av en webbsida att användaren behövde göra ett aktivt val att ladda nästa sida, lämna plattformen eller liknande. Införandet av *infinite scrolls* innebär att sådana beslut aldrig behöver tas då innehållet aldrig tar slut. Användaren ges därmed inga pauser att reflektera över sitt användande. Slutsatsen som Bhargava och Velasquez (2021) drar är att SNS-företag har designat sina tjänster på ett sätt som gör dem beroendeframkallande. Uppsatsen kommer härmed benämna sådan typ av design beroendeframkallande design.

Bhargava och Velasquez (2021) beskriver också tydligt hur denna design bidrar till internetberoende med dess utnyttjande av psykologiska sårbarheter. Den ständiga stimulans som användare utsätts för gör det svårare för dem att begränsa sitt användande. Detta gäller särskilt individer som är sårbara för de psykologiska dimensioner som beroendeframkallande design spelar på. Graden till vilken de kan dras in i att använda internet och SNS kan inte liknas vid något annat än ett beroende (Bhargava & Velasquez, 2021).

2.2.1 Bhargava och Velasquez tre argument

Om designen av SNS är beroendeframkallande leder detta till moraliska problem. Nedan lyfts tre argument för detta formulerade av Bhargava och Velasquez (2021).

The Harm Argument: Det första argumentet handlar om hur SNS skadar användare. Det finns en viktig poäng i flertalet av de studier som gjorts på internetberoende – faktumet att skadorna är beroende av tiden användaren spenderar på teknologin. Den tid som användaren spenderar på SNS tas från tillgänglig tid att lägga på andra områden av livet, vilket ofta slutar med att skola, arbete, hälsa och socialt liv påverkas negativt. Skadan av SNS-beroende begränsas inte bara till aktiv användning då berörda användares uppmärksamhet tenderar att delas mellan SNS och andra aktiviteter under dagen. Som Ward et al. (2017) nämner tar det tid för berörda användare att kunna fokusera på annat efter att de kollat SNS då uppmärksamheten behöver återgå till full kapacitet.

The Adding Insult to Injury Argument: SNS-företag använder sig av adaptiva algoritmer. Detta betyder att plattformarna mäter engagemanget och individanpassar innehållet efter användarnas preferenser för att göra innehållet mer beroendeframkallande och öka deras vistelsetid (Lanier, 2018; Rader & Gray, 2015, refererad i Bhargava & Velasquez, 2021). Bhargava och Velasquez (2021) påpekar hur användardata länge påverkat innehåll, dock aldrig med den precision som adaptiva algoritmer tillåter. SNS skadar användare, och de använder sina användare för att kunna göra detta.

The Exploitation Argument: Att exploatera innebär enligt Wood (2005) att nyttja en svaghet eller sårbarhet hos en person eller föremål för egen vinning. Allt utnyttjande är dock inte moraliskt förkastligt – för detta krävs bristande respekt gentemot objektet som utnyttjas (Arnold, 2010; Wood, 1995, refererade i Bhargava & Velasquez, 2021). Att göra användare beroende av SNS är moraliskt förkastligt av tre anledningar: användare exploateras genom att (1) användarnas sårbarheter utnyttjas (2) för SNS-företagens vinning (3) på ett sätt som inte är respektfullt mot användarna (Bhargava & Velasquez, 2021).

Det finns två sårbarheter som SNS-företag kan exploatera hos SNS-beroende användare. Den första bygger på användarnas starka behov av tjänsterna och hur SNS-företagen utnyttjar detta för att tjäna pengar (Bhargava & Velasquez, 2021). Den andra handlar om internets konstanta närvaro i dagens samhälle och hur tidigare SNS-beroende användare kan falla tillbaka i beroendet på grund av påminnelser i sin omgivning (Lu et al., 2002; Niaura et al., 1988, refererad i Bhargava & Velasquez, 2021). Att se andra använda sociala medier på sina mobiltelefoner kan vara ett exempel på sådan påminnelse.

2.3 Effekter av beroendeframkallande design

När beroendeframkallande design leder till internetberoende kan det påverka människor på många olika sätt. Även om American Psychiatric Association (APA) ännu inte räknar överanvändning av internet som ett beroende (2013) menar Bhargava och Velasquez (2021) att detta är vida accepterat, däribland SNS-beroende. Forskning har bland annat funnit samband mellan internetberoende och minskad förmåga att tänka klart, resonera och engagera sig i tankar som kräver koncentration (Judd, 2014; Karpinski et al., 2013, refererad i Bhargava & Velasquez, 2021). Internetberoende tycks även påverka användares fysiska hälsa gällande fysisk aktivitet och sömn (Kojima et al., 2019).

Baughan et al. (2022) lyfter hur beroendeframkallande design leder till normativ dissociation, ett fenomen som omfattar många olika mentala tillstånd men generellt innefattar att personens uppmärksamhet om sig själv eller världen runt omkring begränsas. Under detta tillstånd har människor en minskad känsla av viljekraft och möjlighet att reflektera, vilket leder till att de kan interagera med SNS utan mål och därmed spendera mer tid på dessa än de ämnat göra.

Då sociala medier består av medvetna designbeslut för att fånga användares uppmärksamhet (Monge Roffarello & De Russis, 2023), kan medvetna designbeslut även tas för att motverka detta (Lukoff et al., 2021). Ett forskningsområde som berör detta är forskning om digitalt välmående.

2.4 Forskning om digitalt välmående

När det kommer till forskning om digital teknologi och mänskligt välmående har begreppet digitalt välmående alltmer hamnat i fokus. Forskningsområdet saknar en enhetlig definition men strävar efter att motverka de negativa effekterna av digital teknologi och ge användare verktyg för att kunna kontrollera deras användning av denna (Vanden Abeele, 2020). Exempel på tillämpningar av konceptet för att minska tiden som människor spenderar på SNS inkluderar försök att begränsa skärmtiden eller ändra estetiska aspekter för att göra dem mindre engagerande (Zhao et al., 2024). Digitalt välmående handlar således främst om att förebygga människors överanvändning och beroende av digital teknologi, inklusive SNS.

2.4.1 Designverktyg för digitalt välmående

När det kommer till designers förmåga att stärka digitalt välmående finner Baughan et al. (2022) att det finns flera olika redskap för att motverka normativ dissociation. I sin studie använder de en rad interna och externa interventioner i ett klientprogram för Twitter och finner att markeringar för redan sedda inlägg och påminnelser om hur mycket tid en användare spenderat är effektiva verktyg för att förebygga det tillståndet. För att stärka digitalt välmående kan således funktioner som spårar tidsbruk, uppmuntrar till avbrott, eller olika former av automatiska interventioner vara användbara (Monge Roffarello & De Russis, 2019).

Nära relaterat till dessa verktyg är konceptet *digital nudging*. Det är design menad att styra användares beslut utan att inskränka på deras valfrihet och består av interventioner – exempelvis notifikationer som informerar en individ om deras kaloriintag (Mirsch et al., 2017). När det kommer till digitala applikationer kan en *nudge* vara information om att en vara de kollar på är under stor efterfrågan, vilket kan trigga användare till att handla. *Nudges* som verktyg kan således ha överlappning med DP, då de går att använda för att beskära användares autonomi, men det är också möjligt att använda dem för att minska individers användande av SNS och därigenom stärka deras digitala välmående. Ett exempel av detta är Purohit et al. (2020) studie där *digital nudging* var utgångspunkten för att designa ett Chrome-tillägg menat att minska tiden användare spenderar på Facebook. Tillägget inkluderade

flertalet *digital nudges*, varav en dolde personaliserade rekommendationer. Om användaren ville se sådana fick de själva klicka på dem. Då personliga rekommendationer är menade att hålla kvar användare genom att visa innehåll som kan vara intressant för dem kan *nudges* minska tiden på SNS genom att uppmäna till medveten handling.

2.5 Definitionen av Dark Patterns

Mathur et al. (2021) undersöker 19 definitioner av begreppet DP och sammanfattar hur forskningen om DP saknar enad definition. Det är således en definition som är öppen för tolkning (Gray et al., 2018; Luguri & Strahilevitz, 2019). Beroende på kontext kan DP dessutom vara önskvärda. Exempel på detta är *autoplay*-funktionen, när videos automatiskt fortsätter spela, som kan anses skadlig när den stjälar användares uppmärksamhet, men som kan bidra till en fördjupad upplevelse om den används för att spela musikvideos i bakgrunden (Monge Roffarello et al., 2023).

Ovan nämnda studie av Mathur et al. (2021) sammanfattar undersökta definitioner till fyra aspekter:

1. **Egenskaper hos DP:** Vissa definitioner innehåller ord som "luriga" och "vilsledande" medan andra beskriver dem som "styrande", "tvingande" och "bedrägliga".
2. **Påverkan på användare:** Elva av de nitton definitionerna innehåller unika sätt som DP anses påverka användare. Några beskriver hur användares intentioner undergrävs, andra hur deras preferenser undergrävs. Vissa definitioner menar att användare "luras" medan andra talar om hur användares autonomi minskas. Flera av definitionerna anser att DP påverkar på flera sätt.
3. **Gränssnittsdesignerns roll:** Vissa definitioner menar att designers använder domänspecifik kunskap om mänskligt beteende för att skapa DP. Andra menar att designerns medvetet använder DP för att nå ett mål.
4. **För- och nackdelar med gränssnittsdesign:** Vissa påtalar hur syftet med DP är att gynna onlinetjänster. Andra understryker hur DP kan skada användare.

2.5.1 Vald definition av Dark Patterns

Den här studien har valt att definiera DP utifrån de fyra aspekter som Mathur et al. (2021) lyfter fram. För denna studie är ett DP således designmönster som medvetet utformats för att lura, styra eller manipulera användare och därmed undergräva deras autonomi, vilket leder till att de fattar beslut som de annars inte skulle ha gjort. Detta kan resultera i olika former av skada för användarna, samtidigt som det gynnar de onlinetjänster som använder dessa designmönster.

2.6 Taxonomier av Dark Patterns

Brignulls försök att definiera de DP han hade identifierat kan ses som den första taxonomin gällande detta fenomen. Majoriteten av de han identifierade verkar inom e-handel, med ett mindre antal applicerbara i SNS. Denna taxonomi var således inte allomfattande, men målet när han definierade dessa DP var att skapa uppmärksamhet om dem som fenomen (Brignull, 2023).

Efter det initiala arbetet av Brignull har det skapats flera taxonomier gällande DP. Vissa applicerbara på specifika områden, andra mer abstrakta. Ett exempel på en specifik taxonomi är Zagal et al. (2013) som är menad att illustrera DP inom spel. Taxonomin inkluderar DP som *Pay to Skip* där en spelare kan betala för att hoppa över delar av ett onlinespel och därigenom spara tid, eller *Grinding* där aktiviteter i ett spel medvetet görs tidsintensiva för att hålla kvar spelare längre.

En annan mer specifik taxonomi är av Monge Roffarello et al. (2023). Denna taxonomi är ett försök att brygga kunskapen mellan DP och digitalt välmående och Monge Roffarello et al. (2023) formulerar en rad DP som är menade att stjåla och skada användares uppmärksamhet, vilka de benämner *Attention-Capture Damaging Patterns* (ACDP). En stor del av dessa kan förekomma i SNS, som *Disguised Ads and Recommendations*, vilket är reklam och rekommendationer menade att posera som normalt innehåll. Andra förekommer enbart i andra kontexter som *Time Fog*, vilket är när videostreamingplattformar försöker minska användarens medvetenhet om tiden genom att dölja mobiltelefonens klocka. Vissa som *Neverending Autoplay*, en kontinuerlig uppspelning av videos, kan förekomma i både SNS och streamingplattformar. Specifika taxonomier kan således omfatta specifika områden som spel, men även mer abstrakta sådana som fokuserar på uppmärksamhet.

När det kommer till försök att skapa en abstrakt taxonomi har Mathur et al. (2019) utifrån tidigare forskning om manipulation online och DP-taxonomier försökt att fånga in DP genom fem olika dimensioner. Dessa är *Asymmetric*, *Covert*, *Deceptive*, *Hides Information* och *Restrictive*. *Deceptive* betyder att effekten av gränssnittets design inger användaren falska övertygelser, antingen via direkta eller missledande påståenden, eller genom att underlätta information. Exempelvis kan användaren missledas att tro att en rabatt är tidsbegränsad, på grund av en nedräkning, och känna tidspress på grund av detta. En nedräkning som egentligen startas om varenda gång sidan laddas in. Taxonomin kom senare att utökas med ännu en dimension *Disparate Treatment*, vilken involverar att vissa grupper av användare behandlas bättre eller sämre än andra (Mathur et al., 2021).

2.6.1 Vald ontologi av Dark Patterns

För den här studien valdes ontologin som definierats av Gray et al. (2024). Den kopplar ihop tidigare taxonomier och består av tre nivåer; hög, meso och låg. Den höga nivån beskriver abstrakt kunskap och är kontextagnostisk, vilket innebär att de kategorier som ingår kan appliceras oberoende om det rör sig om mobil, dator, e-handel eller SNS. Mesonivån är nivån som brygger hög- och lågnivån och är innehållsagnostisk, vilket innebär att de kan tolkas på ett sätt som är lämpligt för den kontext de återfinns i. Kategorierna i den låga nivån är mest specifika i hur deras funktion och utseende beskrivs. De kan troligtvis identifieras via algoritmiskt, manuellt eller annat tekniskt tillvägagångssätt. Precis som Mathur et al. (2019) har Gray et al. (2024) konstruerat dessa nivåer utifrån tidigare forskning. Exempelvis är lågnivån *Disguised Ads* taget från minst sex tidigare källor, varav hälften uttrycker detta DP direkt och de resterande tre indirekt, vilket innebär att de använt sig av DP som Gray et al. (2024) tolkat som motsvarande till *Disguised Ads*.

För att illustrera hur de tre nivåerna av kategorier kan höra ihop med varandra ges nedanstående exempel. Högnivåns *Obstruction* som minskar användarens autonomi genom att hindra dem att göra vad de vill kan kopplas till mesonivån *Adding Steps*. Denna karaktäriseras av att extra steg läggs till för att förhindra användaren att göra som de vill. På lågnivån kan detta sedan kopplas till *Privacy Maze*, ett DP som innebär att integritetsinställningar är svåra att hitta med syftet att hindra användaren från att nå dem (Gray et al., 2024).

När det gäller de olika nivåerna kommer denna studie att utgå från samtliga för att identifiera DP. Både DP från låg- och mesonivån kommer att tillämpas om de matchar de som identifieras på SNS. Om nya DP identifieras kommer de att formuleras utefter hög- eller mesonivåerna i Gray et al. (2024) nuvarande ontologi. Om inget applicerbart finns på den låga nivån behöver det dock inte betyda att något nytt formuleras om det fångas in tillfredsställande av mesonivån. Även om inget DP matchar på mesonivån är högnivån tillräckligt abstrakt för att de DP som finns på den ska kunna användas för att formulera nya på mesonivån.

Då DP bland SNS inte är ett område där det finns en definitiv taxonomi att utgå från togs beslutet att för denna uppsats utgå från en sådan med en abstrakt nivå. Detta då en abstrakt nivå gör en taxonomi helt funktionell när det kommer till att identifiera DP, oavsett vilket område det appliceras på. Anledningen till att den tidigare nämnda abstrakta taxonomin av Mathur et al. (2019) bortsågs från är på grund av att en tidigare tematisk analys av Mildner et al. (2023b) om DP i SNS hade identifierat dimensionen *Disparate Treatment* som något de inte kunde applicera när det kom till SNS. Taxonomin av Monge Roffarello et al. (2023) inkluderades som komplement för dess applicerbarhet på SNS.

3 Metod

Inledningsvis redovisar och argumenterar detta kapitel för valen av forskningsstrategi, datainsamlings- och analysmetoder. Därefter beskrivs hur samtliga metoder tillämpades. Slutligen diskuteras forskningsetiska överväganden och aspekter.

3.1 Metodval

Nedan diskuteras valen av forskningsstrategi, datainsamlingsmetod, analysmetod samt forskningsetiska överväganden.

3.1.1 Forskningsstrategi

Då denna uppsats ämnar undersöka förekomsten av *Dark Patterns* (DP) på sociala medieplattformar (SNS) valdes kartläggning som forskningsstrategi. Kartläggning involverar sökning av nödvändig information från relevanta platser, vilket i denna studie kan appliceras på sociala medier. Kartläggning involverar också undersökningen av detta över en begränsad ögonblicksbild, vilket är passande då det räcker för studiens ändamål att det går att få fram vilka DP som förekommer på vilka SNS under en viss period. Forskningsstrategin tillåter även studien att omfattande belysa fenomenet, eftersom det inte bara är en eller två SNS som uppsatsen ämnar undersöka (Denscombe, 2014). Vanligtvis involverar kartläggning insamlande av information från personer via enkäter, intervjuer etcetera, men kan också involvera analys av dokument (Denscombe, 2014).

En alternativ forskningsstrategi som övervägdes var fallstudie, då det likt kartläggning skulle ge möjlighet att kunna urskilja de DP som manifesterar sig hos de SNS som ska undersökas. Fallstudier är dock mer lämpade för att undersöka ett specifikt fall eller instans av ett fenomen (Denscombe, 2014), medan denna uppsats ämnar undersöka flera. Ett annat mål med fallstudier är att utforska det aktuella fallet djupgående. Exempelvis för att förstå de relationer och processer som äger rum för att kunna förklara varför saker är som de är (Denscombe, 2014). Målet med denna studie är dock att undersöka förekomsten av DP, inte de processer som skulle kunna förklara varför de ser ut som de gör, eller varför de används.

3.1.2 Datainsamlingsmetod

Som datainsamlingsmetod valdes dokumentstudie för att undersöka de nio valda SNS-mobilapplikationerna. Dokument kan ta formen av textuella data, men kan också innehålla bilder eller generell audiovisuell information, vilket inkluderar sociala medieströmmar (Johannesson & Perjons, 2012). Den här studien behandlar således SNS som onlinedokument som dokumenteras med video och ljud för att sedan kunna analyseras för att identifiera förekommande DP. Fördelen med att använda dokumentstudie som datainsamlingsmetod är att datan kommer från en originalkälla. Det finns således mindre risk för misstolkningar eller manipulering från tredje part (Denscombe, 2014).

När det kommer till att bedöma kvaliteten av dokument som datakälla ger Scott (1990) fyra kriterier:

1. **Äkthet**, är dokumentet genuint och av obestridligt ursprung?
2. **Kredibilitet**, är dokumentet fritt från fel och förvrängning?

3. **Representativitet**, är dokumentet typiskt för dess sort, och ifall inte, är utsträckningen av dess otypiskhet känd?

4. **Mening**, är dokumentet tydligt och begripligt?

Ifall dessa kriterier appliceras på de SNS som utgör datasetet utifrån uppsatsens forskningsfråga, har de hög kvalitet som datakällor. Deras äkthet är obestridbar, då de utgör direkta källor av allmänt känt ursprung. Deras kredibilitet utifrån att DP ska finnas är också stark, då det inte finns någon möjlighet för DP att vara felaktiga eller förvrängda. De kan endast finnas eller inte finnas. När det kommer till representativitet så är de SNS som är aktuella för denna studie inte helt typiska, då SNS kan skilja sig en del från varandra, men utsträckningen av deras otypiskhet gentemot varandra är känd. Sista kriteriet kan däremot vara svårt att direkt applicera i denna kontext då de SNS som ska undersökas inte nödvändigtvis är tydliga och begripliga, men detta hämmar dem inte som källor att upptäcka DP hos. SNS bör således utgöra en datakälla av bra kvalitet för denna uppsats.

Studiens frågeställningar försvårar en alternativ datainsamlingsmetod. Enkäter och intervjuer samlar in information om vad deltagarna tycker, med observation samlas information in genom observation av händelser eller fenomen (Denscombe, 2014). Att försöka identifiera vilka DP som finns på SNS med hjälp av de två första metoderna skulle vara en tidskrävande och omständlig process. Deltagarna skulle behöva kunskap om DP och ontologin av Gray et al. (2024). Det skulle även finnas en risk att deltagarna skulle ha olika tolkningar om vad som utgör ett DP, eller var i ontologin ett DP passar in. Den tredje metoden, observation, framstår som ett mer komplicerat alternativ till dokumentstudie, eftersom författarna skulle behöva observera andra interagera med SNS.

Om frågeställningarna ämnade besvara huruvida människor upplever att de utsätts för DP på SNS, samt deras attityder till dessa, skulle enkäter kunna användas för datainsamling. De består vanligtvis av tre delar: en bakgrund, instruktioner och frågor (Johannesson & Perjons, 2012). Bakgrunden skulle kunna förklara vad DP är, instruktionerna och frågorna skulle kunna variera beroende på om enkäten använder sig av stängda eller öppna frågor – alternativt skulle en blandning kunna användas.

3.1.3 Analysmetod

Som analysmetod valdes tematisk analys då den tillåter identifiering, analysering och rapportering av de mönster som finns i datasetet, i detta fall DP (Braun & Clarke, 2006). Ett tema kan förstås som något som representerar mönster eller meningar i datasetet utifrån den aktuella forskningsfrågan. Det tillåter forskaren att fånga upp såväl förekomsten av mönster som dess frekvens. Något centralt under analysens gång är kodning som innebär att forskaren identifierar mönster eller meningar som är intressanta för studien. Koderna arrangeras sedan i teman och underteman för att fånga in vidden av det som uttrycks i materialet. Utförandet av tematisk analys görs i sex olika faser och gemensamt för samtliga, förutom den sista, är att kodningen är konstant närvarande. Kodning och omkodning pågår således under hela processen tills dess att resultaten ska sammanfattas (Braun & Clarke, 2006).

Tematisk analys kan utföras induktivt utan något teoretiskt ramverk. De teman som identifieras blir därmed bundna till det insamlade datasetet. Alternativt kan analysen drivas deduktivt med tidigare teorier och koncept (Braun & Clarke, 2006). För denna studie har den teoretiska varianten av tematisk analys valts.

En alternativ analysmetod som övervägdes var innehållsanalys. Likt tematisk analys involverar den identifiering av mönster, analysering av dessa och skapande av kategorier (Denscombe, 2014). Innehållsanalys förutsätter dock att det som ska mätas passar in i förutbestämda och enhetliga kategorier. Detta överensstämmer inte med den utgångspunkt som denna studie har i ontologin av Gray et al. (2024). Tematisk analys valdes således för dess flexibilitet att fånga in icke enhetliga DP.

3.1.4 Forskningsetiska aspekter

I enlighet med god forskningssed (Vetenskapsrådet, 2017) har denna studie ämnat upprätthålla hög kvalitet genom god metodtillämpning i syftet att uppnå tillförlitliga resultat. Nedan följer andra forskningsetiska överväganden som berörs.

Valet av dokumentstudie underlättar åtkomst till studiens insamlade materialet då detta finns offentligt publicerat på SNS-plattformarna. Onlinedokument kan dock medföra svårigheter gällande integritetsfrågor då innehållet inte laddats upp i forskningssyfte. Det kan även förekomma upphovsrättsskyddat material (Denscombe, 2014). Vid studiens utforskande och dokumenterande av valda SNS kan individer och upphovsrättsskyddat material förekomma i dessa. Även om intresset ej ligger i att dokumentera sådant kan namn, bilder, musik och skriftligt innehåll komma att medverka i inspelningarna. Vad gäller individer bör informerat samtycke inhämtas av samtliga som utgör en del av forskning (Vetenskapsrådet, 2017). Den stora volymen av olika individer som fångas upp när valda SNS med flöden spelas in gör detta praktiskt omöjligt. För att skydda individers intressen kommer åtgärder i form av maskering appliceras på samtliga element som kan användas för att identifiera individerna. Detsamma gäller upphovsrättsskyddat material (Denscombe, 2014).

3.2 Metodtillämpning

Nedan diskuteras hur studiens urval, datainsamling och analys tillämpades.

3.2.1 Urval

Urvalet av SNS motiveras av popularitet hos det svenska folket. Med 'popularitet' menas att valda SNS använts av minst tio procent av det svenska folket någon gång under tolv månader (Andersson et al., 2023). Urvalet består därmed av nio plattformar: Facebook, Youtube, Instagram, Snapchat, LinkedIn, TikTok, X, Pinterest och Reddit.

Då studien avser att undersöka mobilapplikationer av SNS har Flashback uteslutits på grund av brist på sådan. Valet av mobilapplikationer motiveras utifrån att de utgör ett potentiellt större hot mot användares digitala välmående. Detta eftersom de uppmuntrar till passiv och överdriven användning (Monge Roffarello & De Russis, 2023). Valet av mobilapplikationer motiveras även av den höga förekomsten av DP i sådana. Di Geronimo (2020) undersökte 240 populära Android-applikationer och fann att DP förekom i 95 procent av dem.

3.2.2 Genomförande

Studien genomförs genom att undersöka de SNS som valts för studien på två mobiltelefoner med Androids operativsystem med samtliga mobilapplikationer installerade. Undersökningen genomförs individuellt av författarna efter ett protokoll (Tabell 1) med färdigformulerade uppgifter. Detta görs för samtliga SNS och säkerställer att alla applikationer undersöks på samma sätt, vilket säkerställer studiens reproducerbarhet och reliabiliteten av dess resultat.

Undersökningen dokumenteras via ljud- och skärminspelning som lagras på externa hårddiskar. Det insamlade datasetet analyseras sedan med båda författare närvarande. Analysen görs enligt riktlinjerna för tematisk analys. Under analysen tas skärmdumpar och videoklipp för att dokumentera de DP som hittas. Skärmdumparna och videoklippen inspekteras efter personliga uppgifter, dessa maskeras sedan. Detsamma görs för upphovsrättsskyddat material. Dessa sparas också på externa hårddiskar.

3.2.3 Protokoll

Datainsamlingen utgår från ett protokoll som består av tio uppgifter (Tabell 1). Fem uppgifter inspirerades av Di Geronimo et al. (2020) * som undersökte populära applikationer på Google Play Store. Fem inspirerades av Mildner et al. (2023a) ** som undersökte förekomsten av DP i SNS. Till vissa används material (bilder och videoklipp) som tillhandahålls i förhand.

Tabell 1 Protokoll med uppgifter att utföra under datainsamlingen

1. **	Sätt på skärminspelning på den mobila enheten som ska användas.
2. *	Öppna applikationen och skapa ett konto, logga sedan in och ut.
3. *	Stäng applikationen och öppna den igen.
4. **	Skapa någon typ av innehåll, lägg upp det och ta sedan bort det.
5. **	Följ och avfölj andra konton.
6. *	Gå in på personliga inställningar. <i>I Hur mycket kan du välja vad som delas om dig?</i> <i>II Hur mycket kan du välja vad du tar del av?</i> <i>III Har appen något verktyg för att begränsa tiden du använder i appen?</i>
7. *	Gå in på reklamrelaterade inställningar. <i>I Kan du kontrollera vilken typ av reklam du tar emot?</i>
8. *	Använd applikationen för dess syfte (minst tio minuter): <i>I Beskriv appens naturliga flöde - vad använde du den för?</i> <i>II Kunde du använda appen som du ville eller blev dina interaktioner "guidade"?</i> <i>III Hur lätt var det att bli distraherad och om så var fallet, vad distraherade dig?</i>
9. *	Ta bort ditt konto. <i>I Gick det att ta bort det helt eller går det att återställa?</i>
10. **	Stäng av skärminspelningen och spara den.

För att se hur lång tid utförandet skulle ta och ifall det fanns möjlighet att utveckla uppgifterna utfördes en pilotstudie. Vissa uppgifter kom att utökas med frågor. Dessa återges kursiverade i Tabell 1.

3.2.4 Dokumentstudie

Då innehåll och utseende på SNS förändras dokumenterades dåvarande version för att säkerställa studiens reproducerbarhet. Dokumentation gjordes med skärminspelning på de mobiltelefoner som användes eftersom DP kan innefatta dynamiska element och därmed vara svåra att identifiera på statiska bilder (Di Geronimo et al., 2020). Utöver video spelades också ljud in då *think-aloud*-metoden (Jaspers et al., 2004) användes för att komplettera videomaterialet med förklaringar bakom författarnas handlingar. Detta för att lättare kunna identifiera förekomna DP.

Samtliga SNS undersöktes separat av båda författare under cirka 30 minuter per SNS. Arbetet att undersöka alla nio SNS tog fem dagar och resulterade i 60 minuters inspelat material per SNS. För samtliga SNS blev det totalt cirka åtta timmar material.

3.2.5 Tematisk analys

Utförandet av analysen skedde i enlighet med de steg som definierats av Braun och Clarke (2006). Då tematisk analys påbörjas under datainsamlingen hade initiala idéer om koder redan påbörjats.

Analysen av videomaterialet utfördes med båda författare närvarande för att minska risken för *DP-blindhet*. Ett fenomen som innebär att även experter kan missa förekomsten av DP, speciellt i videoinspelningar (Di Geronimo et al., 2020). Författarna gick igenom samtliga videoinspelningar och identifierade dataextrakt för att generera koder. Detta gjordes med stöd av ontologin av Gray et al. (2024), taxonomin av Monge Roffarello et al. (2023), samt författarnas definition av DP. Utifrån behov pausades inspelningarna för diskussion och kodande. De koder där ingen klassificering initialt kunde göras antecknades dessa instanser som frågetecken. Samtliga koder skrevs ned i ett Google Docs-dokument, sorterade efter författare och SNS. De gavs kontext, förklaring samt tidsstämplar för att dokumentera förekomsterna i inspelningarna. För de som mappades till ontologin av Gray et al. (2024) eller taxonomin av Monge Roffarello et al. (2023) markerades detta. I nästa iteration kom samtliga koder att sorteras efter SNS. Detta för att lättare kunna sammanfatta och analysera förekomster av identifierade DP. Skärmdumpar togs av de koder som inte var dynamiska, medan de dynamiska dokumenterades genom att klippa ut videoklipp från inspelningarna. Samtliga skärmdumpar och videoklipp namngavs efter de DP som de motsvarade.

I nästa steg analyserades samtliga koder och mappades till ontologin av Gray et al. (2024) och definitionerna av Monge Roffarello et al. (2023). Braun & Clarke (2006) rekommenderar att använda visuell representation för att representera de teman som identifieras, vilket gjordes med hjälp av programmet Draw.io.

När koderna matchats med Gray et al. (2024) överhängande teman inleddes granskningen. Först granskades samtliga koder för koherens gentemot sina teman. Om sådan inte fanns kodades de om eller flyttades till ett annat tema. Sedan bedömdes validiteten av de matchningar som gjorts utifrån datasetet. Här fanns möjligheten att gå igenom samtliga teman för att se om de kunde omformuleras eller förbättras utifrån dataextrakt som tidigare kan ha missats i datasetet. Omkodning kan ske ad infinitum och detta togs i beaktande. Omkodningen avslutades därmed när den inte längre ansågs ge förbättringar.

I nästa steg definierades och förfinades samtliga kategorier och underteman, alltså låg- och mesonivåer, som hittats. Varje nytt DP gavs en detaljerad förklaring om vad de fångade upp utifrån hur DP i ontologin av Gray et al. (2024) formulerats. Samtliga ansågs vara tydligt definierade om det gick att beskriva deras omfattning och innehåll utifrån ontologins formuleringar, i annat fall förfinades de ytterligare.

Slutligen skrevs den tematiska rapporten. Målet var att berätta historien bakom studiens data på ett sätt som övertalar läsaren om validiteten av analysen. Utifrån detta inkluderades bland annat tydliga exempel av de koder (statiska DP) som framkommit i form av skärmdumpar. Dessa demonstrationer av funna koder användes också för att argumentera för de kategorier och underteman som identifierats.

3.2.6 Forskningsetiska aspekter

Skärmdumpar och videoklipp av upptäckta DP maskerades för att kunna användas i uppsatsen utan att riskera individers integritet eller använda upphovsrättsskyddat material.

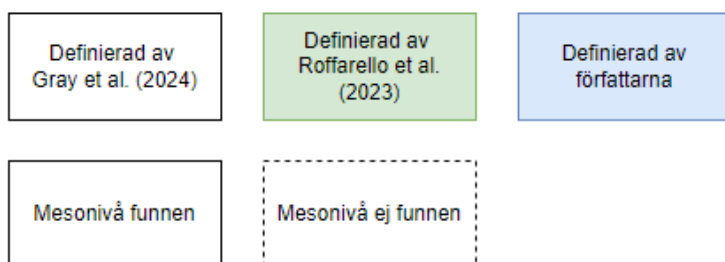
Materialet som samlats in kommer bara användas i syftet att besvara forskningsfrågan (Denscombe, 2014). Det lagras på en extern hårddisk och sparas i tre månader efter avslutad studie.

4 Resultat och analys

Nedan presenteras först resultatet av den tematiska analys som utförts på det material som samlats in om Sveriges nio mest populära mobilapplikationer för sociala medieplattformar (SNS) och de *Dark Patterns* (DP) som förekommer där. Därefter presenteras analysen och dess slutsatser som avser att besvara studiens frågeställningar. I bilaga B finns en tabell av samtliga unika DP som hittades, inklusive de DP som definierades av författarna.

4.1 Förekomsten av Dark Patterns på nio av de mest populära SNS-mobilapplikationerna i Sverige

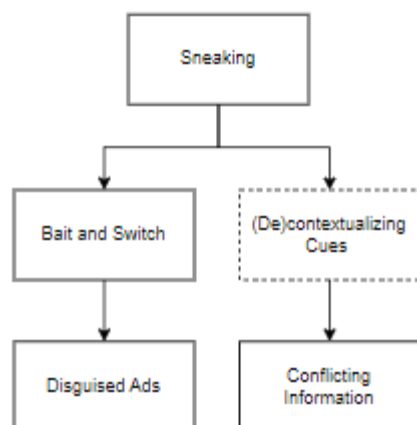
Nedan redovisas samtliga funna meso- och lågnivåer av DP, organiserade i underrubriker döpta efter högnivåerna i ontologin av Gray et al. (2024). Varje underrubrik har även en visualisering av funna DP för gällande högnivå. I dessa tillhör vita rektanglar ontologin av Gray et al. (2024), gröna tillhör Monge Roffarello et al. (2023) och blåa är identifierade och formulerade av författarna. Mesonivåer med heldragen kontur innebär att en iakttagelse av denna gjorts. De med streckad kontur innebär att endast lågnivåer av denna identifierats (Figur 1). En mesonivå med heldragen kontur och tillhörande lågnivå innebär således att iakttagelse gjorts av meso- och lågnivå(er).



Figur 1 Visualisering av de olika typerna av funna Dark Patterns

4.1.1 Sneaking

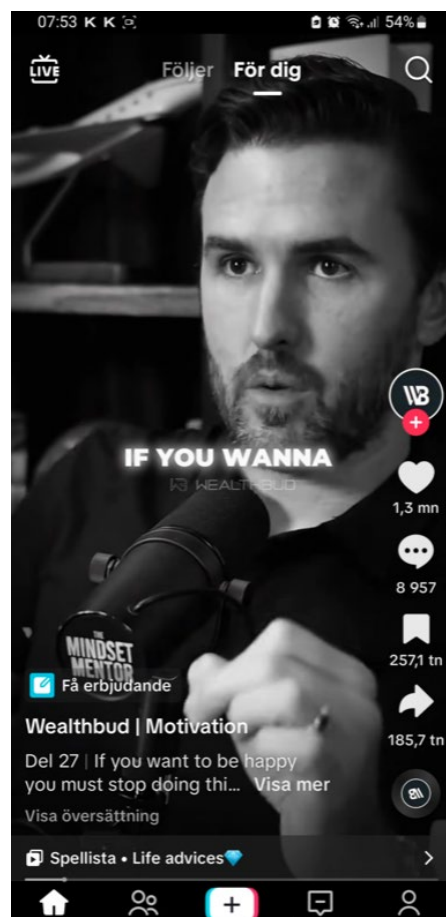
Högnivån *Sneaking* från ontologin av Gray et al. (2024) beskriver en strategi som försenar, förklarar eller gömmer viktig information från användaren. Detta kan leda denna att omedvetet ta felaktiga beslut. Totalt hittades tre unika instanser av DP under denna högnivå (Figur 2).



Figur 2 Visualisering av högnivån Sneaking med samtliga funna meso- och lågnivåer

Bait and Switch handlar om hur systemet lurar användaren att tro att deras handling kommer leda till ett önskvärt resultat – en handling som i stället resulterar i något oönskat. Exempel på detta är hur Youtube leder användaren att tro att de kan ladda hem videos gratis men när användaren klickar på “Ladda ner” öppnas i stället en ruta för att börja abonnera på Youtube Premium. Under denna mesonivå hittades även ett DP på lågnivån *Disguised Ads* som beskriver hur annonser, eller annat partiskt innehåll, illustreras på sätt som lurar användaren att tro att det handlar om vanligt innehåll. Denna typ av DP hittades på samtliga SNS utom Snapchat. Figur 3 illustrerar ett exempel från TikTok där en video i flödet ser ut att vara en vanlig video, men texten “Få ett erbjudande” gör detta tvetydigt. Denna lågnivå kan leda till att användaren engagerar sig med innehåll i tron att det är opartiskt när det i själva verket är sponsrat och kan vara vinklat med ensidig information.

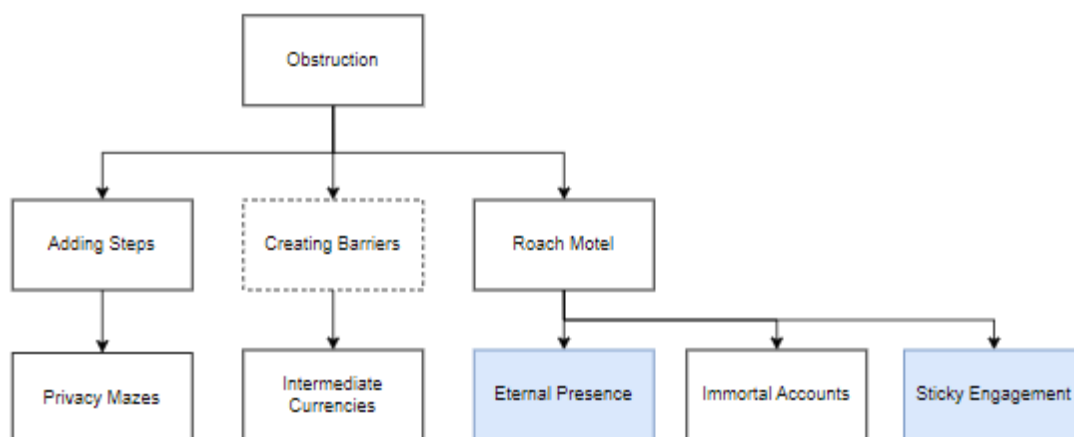
Det sista exemplet som identifierades var lågnivån *Conflicting Information*. Den representerar hur systemet inkluderar två motstridiga källor till information för att förvirra användaren, vilket kan leda till att användaren är mer mottaglig för standardinställningar som inte nödvändigtvis är i deras intresse. Endast en instans av denna iakttogs och detta var när Facebook frågade författaren om de ville tillåta Facebook att använda *cookies*. På grund av formuleringen var det oklart att förstå vad valen kunde komma att leda till.



Figur 3 Exempel på lågnivån Disguised Ads på TikTok

4.1.2 Obstruction

Obstruction är ännu en högnivå i den ontologi som definierats av Gray et al. (2024). Strategin som brukas här är att störa användarens uppgiftsflöde genom att försvåra vissa interaktioner med målet att övertala användaren att inte ta vissa handlingar. När det kommer till underordnade variationer av detta DP på meso- och lågnivå hittades sju unika instanser, varav två definierades av författarna (Figur 4).



Figur 4 Visualisering av högnivån Obstruction med samtliga funna meso- och lågnivåer

En mesonivå som identifierades var *Adding Steps* där användarens förväntningar om att en uppgift ska ta så få steg som möjligt i stället undergrävs genom att de måste utföra flera nödvändiga, men onödiga steg. Exempelvis både Pinterest och LinkedIn involverade onödigt många steg för användaren att ta bort sitt konto, däribland obligatoriskt delgivande om varför användaren vill ta bort det. Ett annat exempel är hur YouTube gjorde det onödigt komplicerat att stänga av *autoplay* för videos på mobilapplikationen, genom att tvinga användaren att gå in i inställningarna. Det var således tydligt att flera SNS försökte göra vissa handlingsalternativ oattraktiva för användaren genom att lägga till flera steg för att åstadkomma det de vill göra.

En underordnad DP på lågnivå som identifierades var *Privacy Mazes*, vilket innebär att alternativ och information gällande användares personliga information döljs genom att tvinga dem att navigera genom flera olika sidor för att få en helhetsbild av det hela. YouTube och Facebook stod ut då placeringen av alternativ och information på dessa SNS gjorde det komplicerat att få en överblick hur användarens information används, samt vilka alternativ det finns för att styra denna användning. Det fanns således flera SNS där det fanns ett betydande motstånd för en användare att utforska och styra hur deras personliga information egentligen används.

Mesonivån *Creating Barriers* består av att användarens förväntningar om att relevanta handlingar ska stödjas av gränssnittet, i stället abstraheras, kompliceras eller omöjliggörs vissa handlingar av dess design. Inga självständiga instanser av denna nivå identifierades, men instanser av den underordnade lågnivån *Intermediate Currencies* identifierades på flera SNS, däribland Instagram, Reddit, Snapchat och TikTok. Denna DP handlar om att användarens möjlighet att bedöma den sanna kostnaden av en transaktion försvåras genom att begära att användaren först betalar för en virtuell valuta. Det finns således en möjlighet att de utför köp som de aldrig skulle ha gjort ifall den verkliga kostnaden kommunicerades till dem.

Den sista mesonivån som identifierades för *Obstruction* var *Roach Motel*, vilket är när användarens förväntningar om att en handling ska vara lika enkel att gå tillbaka som den är att utföra undergrävs. Endast ett fåtal instanser av denna nivå för sig själv identifierades, däremot så identifierades tre underordnade låg-nivå DP på flera olika SNS, varav två definierades av författarna. Den lågnivå DP som redan existerade sen tidigare var *Immortal Accounts*, vilket handlar om att det görs mycket svårt, eller till och med omöjligt för användaren att ta bort ett konto när de väl har skapat ett, vilket leder till att de kan skapa ett eller dela information med det falska antagandet att de enkelt kan ta bort det senare. Dessa svårigheter skulle således kunna övertala dem att låta kontona vara, även fast de skulle

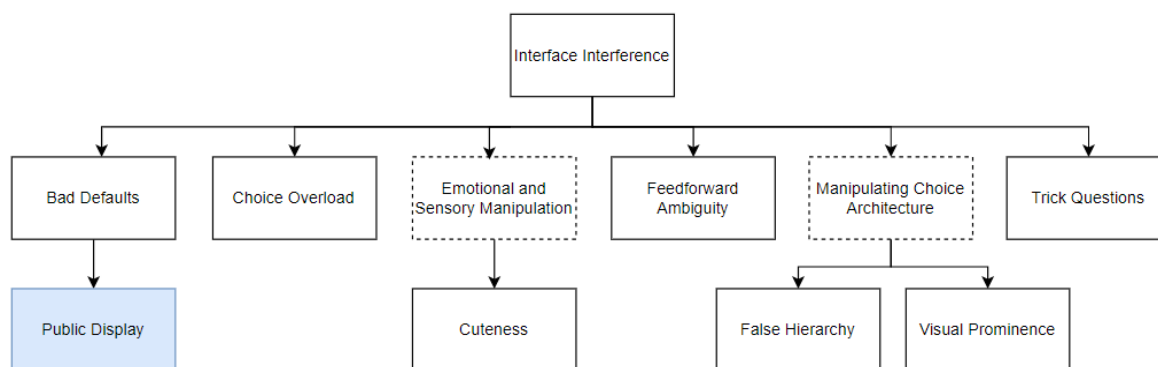
vilja ta bort dem. Detta DP identifierades på bland annat Facebook, Instagram, Pinterest och Snapchat. På samtliga fann författarna det svårt att hitta vart man kunde ta bort sitt konto.

När det kommer till DP som författarna själva definierade under *Roach Motel* var den första *Eternal Presence*. Författarna fann att det på flera SNS var mycket svårt att hitta exakt vart man loggade ut, och *Roach Motel* ansågs inte helt kunde fånga upp detta fenomen. Således definierade författarna ett nytt DP på lågnivå kallat *Eternal Presence*, vilket handlar om att det görs mycket svårt för användaren att logga ut när de väl skapat ett konto och loggat in för första gången. Denna DP identifierades på flera SNS, däribland på Reddit, Youtube, TikTok och X. På Reddit och TikTok manifesterades detta DP genom att det var mycket svårt att hitta exakt vart man kunde logga ut, exempelvis på Reddit doldes alternativet bakom en ikon som inte omedelbart kommunicerade detta. På YouTube rörde svårigheten att inloggningen var kopplad till Google-kontot, vilket innebär att en utloggning från Youtube på mobilen även innefattar en utloggning från Google-kontot. På X handlade det om att ingen av författarna kunde hitta hur de skulle logga ut. Här avråds således användarna från att logga ut på grund av de svårigheter som är involverade med att uppnå den uppgiften.

Sticky Engagement var det andra DP som författarna definierade under *Roach Motel*. Ett vanligt mönster på SNS var att det var enklare att följa innehåll än att avfölja det och återigen ansågs inte *Roach Motel* helt kunna fånga upp detta. *Sticky Engagement* definierades således som att det är enkelt för en användare att följa socialt innehåll, medan det är svårare att avfölja det, vilket innebär att användare kan bestämma sig att följa socialt innehåll med det falska antagandet att det är lika enkelt att sluta följa det. Exempelvis på Facebook fann författarna det svårt att hitta var man kunde avfölja sidor, konton och hur man skulle gå till väga för att gå ut ur grupper.

4.1.3 Interface Interference

Högnivån *Interface Interference* i ontologin av Gray et al. (2024) är en strategi som försöker övertyga användaren genom att privilegiera vissa handlingar över andra i användargränssnittet. Detta kan förvirra användaren och därigenom begränsa deras möjligheter att upptäcka handlingar som kan vara relevanta för dem. När det kommer till underordnade instanser av denna DP hittades åtta unika sådana både på meso- och lågnivå, varav en definierades av författarna (Figur 5).



Figur 5 Visualisering av högnivån *Interface Interference* med samtliga funna meso- och lågnivåer

En mesonivå som identifierades var *Bad Defaults*, där användarens förväntningar om att standardinställningarna ska vara i deras bästa intresse undergrävs. Användaren måste i stället ta aktiva steg för att ändra dessa för att undvika potentiell skada eller oavsiktligt avslöjande av information. På Facebook, Reddit, Snapchat, TikTok och LinkedIn var standardinställningarna att anpassa reklamen utifrån användarens personliga information. På Reddit och Snapchat var standardinställningarna att använda sådan information från utomstående parter för att anpassa den. Ett annat exempel är hur

Facebook har som standardinställning att andra användare kan hitta en via angiven e-postadress eller telefonnummer. Detta kan ses som problematiskt då det inte är givet att användaren vill att en individ som har tillgång till deras e-postadress eller telefonnummer ska kunna hitta deras Facebook. Det var tydligt att standardinställningar på SNS ofta var inställda på sätt som tjänade plattformen, men där användaren, om möjlighet till reflektion givits, skulle valt andra alternativ.

En underordnad DP till *Bad Defaults* som definierades av författarna var *Public Display*.

Standardinställningar som privilegierade synligheten av personlig information eller skapat innehåll var allmänt förekommande på SNS, och *Bad Defaults* ansågs inte kunna fånga upp detta fenomen på ett tillfredställande sätt. Således formulerades *Public Display* på lågnivå och beskriver hur standardinställningar används för att göra användarens information och skapade innehåll offentligt synliga. Detta kan resultera i att användaren oavsiktligt avslöjar mer information än önskvärt om sig själva till individer eller grupper. Ett exempel är hur Facebook, Pinterest och Reddit har som standardinställning att användarens profil är indexbar av sökmotorer, vilket gör dem enklare att hitta. Ett annat exempel är hur skapat innehåll på TikTok, X och LinkedIn har som standardinställning att vara offentligt. Det var således tydligt att offentliggörande av både information och skapat innehåll av användaren privilegierades på flera SNS.

En DP på mesonivå som identifierades var *Choice Overload* som beskriver hur användarens förväntningar om att valen de har ska kunna förstås och jämföras men i stället förses med för många alternativ. På grund av volymen av alternativ uppmuntras användaren att förbise relevant information. Ett exempel på detta var Instagram, där författarna fann att alternativen på inställningarna var extremt oorganiserade. Detta gjorde det svårt att orientera och utforska de givna handlingsalternativen.

En mesonivå DP som inte förekom av sig själv men som hade underordnade förekomster var *Emotional or Sensory Manipulation*. Denna handlar om hur användarens förväntningar att de ska kunna uppnå sina mål utan manipulation undergrävs. För detta används olika sorters språk, färger eller andra designelement för att skapa en viss känsla eller för att manipulera sinnena för att övertala användaren till en viss handling. Den underordnade lågnivån som identifierades var *Cuteness* som handlar om hur vissa element eller signaler används i designen av ett robotgränssnitt eller andra formfaktorer. I fallet med en robot kan detta leda till att användaren sätter otillbörlig tillit till den och delar information som de annars inte skulle delat. Ett exempel av detta är hur Snapchat har en chattbot med en "söt" avatar (Figur 6).

Ännu en mesonivå som identifierades var *Feedforward Ambiguity* som handlar om hur resultatet av användarens val förväntas kunna förutsägas. På grund av diskrepans mellan tillgänglig information och handlingar resulterar det i stället i något oväntat för användaren. Ett exempel av detta hittades på Facebook där en knapp med texten "Titta nu" inte ledde till någon förväntad form av intern livetjänst,



Figur 6 Exempel på lågnivån *Cuteness* på Snapchat

utan istället ledde till en extern sida. Ett annat exempel identifierades på Reddit, där reklamens format av ett vanligt inlägg missvisande kommunicerade att delar av det skulle vara klickbara utan att ta användaren till någon extern länk.

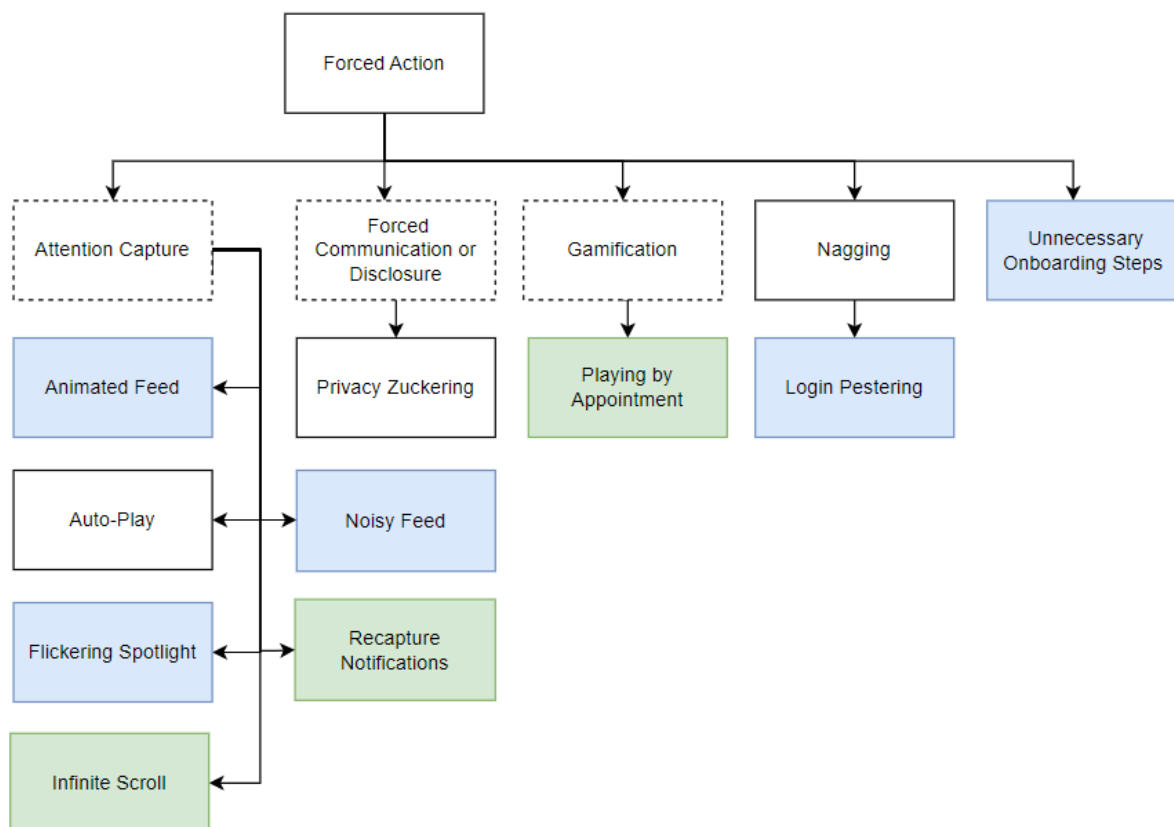
Mesonivån *Manipulating Choice Architecture* förekom inte av sig själv, men två underordnade DP identifierades. *Manipulating Choice Architecture* handlar om hur användarens förväntningar om att givna handlingsalternativ stödjer deras mål undergrävs, genom att istället inkludera en ordning eller struktur till dem som gör andra utfall mer troliga. Ett av de två DP som identifierades under denna var *False Hierarchy*, vilken handlar om att ett eller flera alternativ ges visuellt och interaktivt företräda över andra, speciellt när de egentligen borde presenteras likvärdigt. Detta kan leda till att användaren missförstår eller ej lyckas att helt förstå de handlingsalternativ de har tillgängliga och således fattar beslut utifrån en falsk eller inkomplett grund. Ett exempel är hur både Facebook och Instagram privilegierar valet att spara ens inloggningsinformation, då alternativens hierarkiska position drar användarens uppmärksamhet till valet att spara. Ett annat är hur när Snapchat frågar användaren ifall de kan använda dess position för olika funktioner är alternativen arrangerade så att användaren först och främst dras till alternativet för att godkänna det, vilket är problematiskt då användarens GPS-position kan vara en känslig uppgift.

Det andra DP som identifierades på denna nivå är *Visual Prominence* och kan anses vara nära relaterat till *False Hierarchy*. Här så handlar det dock om att de element eller handlingsalternativ som är relevanta för användaren sätts i visuell konkurrens med andra mer distraherande och synliga saker. Detta kan leda till att användaren distraheras från, eller till och med helt glömmer bort vad deras syfte var till att börja med. Exempelvis på Reddit när användaren skapar ett konto så ombeds de att engagera i en avatar-funktion som kommuniceras på ett mycket distraherande och framträdande sätt jämfört med att bara gå vidare i kontoskapande processen. Sammanfattningsvis var både *False Hierarchy* och *Visual Prominence* vanligt förekommande på SNS och ibland förekom de även i kombination med varandra då vissa handlingsalternativ kunde privilegieras samtidigt som distraherande element var närvarande.

Den sista mesonivån som identifierades var *Trick Questions*. Denna DP undergräver användarens förväntningar om att givna instruktioner kommer att vara skrivna på ett enkelt och direkt sätt, i stället så används förvirrande eller ledande språk för att manipulera användarens val. Ett exempel är hur Facebook försöker uppmåna användaren att fylla i sin profil, genom att göra dialogrutan för att avbryta processen tvetydigt formulerad.

4.1.4 Forced Action

Forced Action är en högnivå från ontologin av Gray et al. (2024) som övertygar användaren till fler och/eller ovidkommande handlingar för att få åtkomst till, eller kunna fortsätta använda specifik funktionalitet. Om sådana handlingar inte utförs nekas användaren fortsatt interaktion med systemet. Totalt hittades elva unika DP under denna högnivå, varav fem definierades av författarna (Figur 7).



Figur 7 Visualisering av högnivån Forced Action med samtliga funna meso- och lågnivåer

Från ontologin hittades mesonivån *Nagging* som handlar om hur användaren tror sig ha rationell kontroll över interaktionen med systemet men i stället blir upprepat distraherad av denna att utföra handlingar eller ta beslut som de inte vill göra. Denna mesonivå återfanns på Facebook, Reddit, TikTok och LinkedIn. Som ny lågnivå till denna skapade författarna *Login Pestering* som beskriver hur systemet försöker övertala användaren att spara sina inloggningsuppgifter. Det var ett återkommande tema som återfanns på Facebook, Instagram och Snapchat. Definitionen beskriver hur denna används för att sänka barriären för engagemang med systemet genom att ta bort en möjlighet för användaren att reflektera över sitt val att engagera vid nästa inloggning. SNS tycks använda tjtande för att nöta ned användarens motstånd, om sådant finns, och få dem att agera i enlighet med systemets mål.

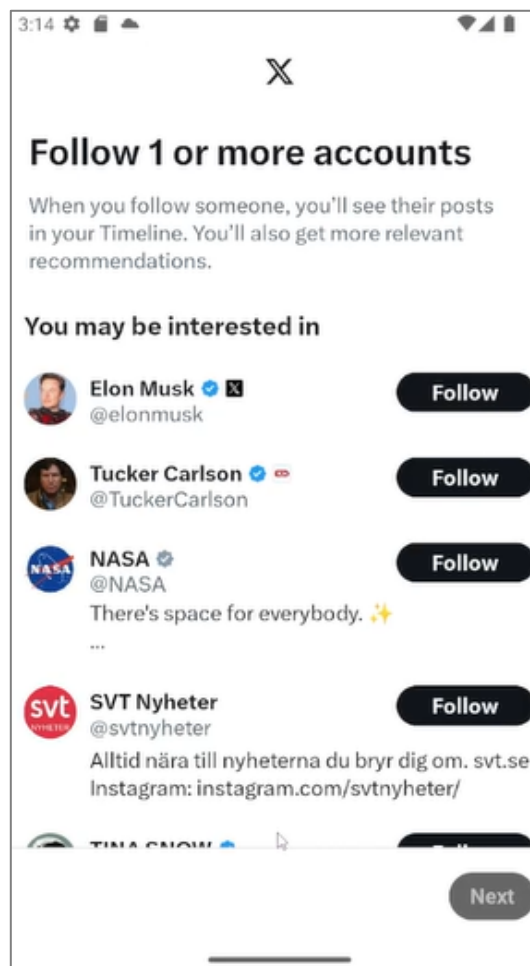
Två lågnivåer från ontologin hittades också – *Auto-Play* och *Privacy Zuckering*. Den första tillhör mesonivån *Attention Capture* som handlar om hur användaren luras tro att de har kontroll över mängden tid, eller andra resurser, de spenderar på systemet. *Auto-Play* beskriver sedermera hur användaren kan tappa kontrollen på sitt tittande genom att videos automatiskt fortsätter spelas upp efter varandra. Detta kan även leda till att användaren tittar på oväntat eller skadligt innehåll och en instans av denna lågnivå hittades på Youtube. *Privacy Zuckering* tillhör mesonivån *Forced Communication or Disclosure* som beskriver användarens förväntan att systemet bara efterfrågar nödvändig information men i stället lurar dem att dela onödig information om sig själva som sedan kan användas i syften som användaren inte godkänt. Lågnivån *Privacy Zuckering* manipulerar användaren att dela med sig av personlig information som inte är nödvändig för att använda tjänsten. Information som sedan används i andra syften, såsom att säljas vidare. Instanser av denna DP återfanns på Facebook och LinkedIn.

Resterande tre nya lågnivåer hamnade under den tidigare definierade mesonivån *Attention Capture*. Den första av de nya lågnivåerna kallas *Animated Feed* och beskriver hur videos spelas upp automatiskt i systemets flöde för att fånga användarens uppmärksamhet. Detta kan i sin tur leda till att användaren förlorar kontrollen över tiden de spenderar där. Den skiljer sig således från *Auto-Play* då den distraherar användaren i dess utforskande av flödet. *Animated Feed* återfanns i samtliga SNS utom LinkedIn och kombineras med författarnas *Noisy Feed* av Facebook, Instagram, Pinterest, Reddit, Snapchat, Youtube och TikTok. *Noisy Feed* beskriver hur ljud spelas upp automatiskt i flödet och kan precis som *Animated Feed* användas för att fånga användarens uppmärksamhet och leda till att användaren förlorar kontrollen över sin tid. Det tycks vara vanligt förekommande hos SNS att försöka fånga användarens uppmärksamhet med audiovisuella element som automatiskt spelas upp.

Den sista av de nya lågnivåerna heter *Flickering Spotlight* och fångar användarens uppmärksamhet genom plötsliga förändringar i gränssnittet för att markera specifik funktionalitet. Detta kan leda till att användarens uppmärksamhet guidas från sin ursprungliga uppgift och i stället leds att interagera med vad systemet anser vara lämpligt. En instans av denna DP iaktogs på Pinterest. Genom att hastigt ändra den neutrala svarta "Dela"-ikonen till en neongrön sådan för meddelandeapplikationen WhatsApp uppmuntras användaren att dela med sig av videon till andra.

Författarnas nya mesonivå *Unnecessary Onboarding Steps* beskriver hur användaren luras att tro att de kan börja använda systemet på egna villkor efter nyregistrering men i stället tvingas genomgå handlingar som kan anses vara onödiga för att kunna börja använda systemet. Plattformen X tvingar användaren att välja mellan ett fåtal förutbestämda offentliga personer och företag att följa innan de kan använda tjänsten som de vill (Figur 8). På liknande sätt tvingar Reddit användaren att välja minst ett forum att följa. SNS tycks därmed använda *Unnecessary Onboarding Steps* och *Flickering Spotlights* för att leda användarens uppmärksamhet och handlingar.

Slutligen hittades tre unika förekomster av DP definierade av Monge Roffarello et al. (2023) – *Infinite Scroll*, *Recapture Notifications* och *Playing by Appointment*. Den första beskriver hur innehåll ändlöst fortsätter laddas när användaren skrollar nedåt i flödet. Används främst på SNS och kan framkalla användning utan tidsbegränsningar. Återfanns på Facebook, Instagram, Pinterest, Reddit, Snapchat, Youtube och TikTok. *Recapture Notifications* beskriver hur notifikationer skickas från systemet för att återfå användarens uppmärksamhet och övertala dem att påbörja en ny användningssession. Används främst på SNS, streamningstjänster och meddelandeapplikationer. Hittades på Pinterest, Youtube, TikTok och X. Sista lågnivån *Playing by Appointment* handlar om hur användare tvingas använda tjänsten under specifika tider som definierats av tjänsten för att inte förlora poäng eller prestationer. Används främst på spelsidor på SNS och SNS i allmänhet. Denna lågnivå hittades endast på Snapchat. De två första lågnivåerna placerades under Gray et al. (2024) mesonivå

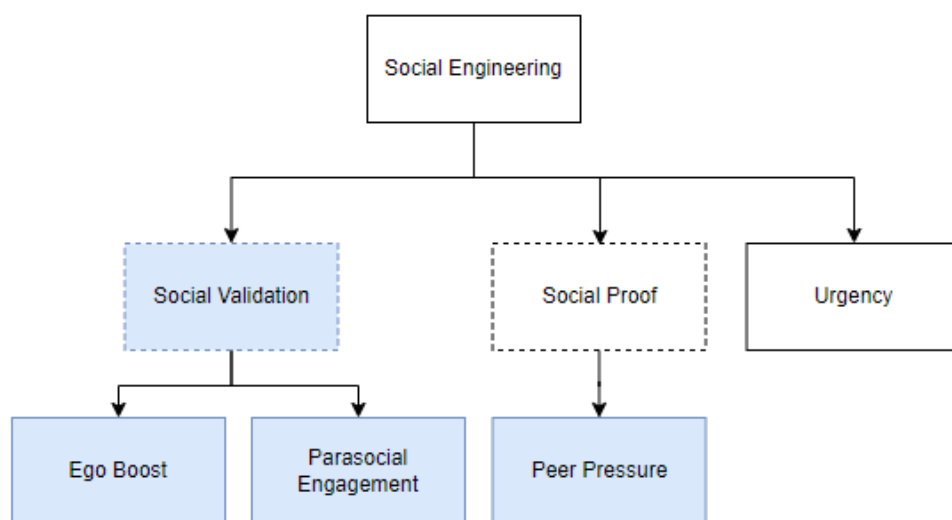


Figur 8 Exempel på mesonivån *Unnecessary Onboarding Steps* på X

Attention Capture på grund av deras fokus på hur systemet stjälar användarens uppmärksamhet. Den tredje sattes under Gray et al. (2024) mesonivå *Gamification* på grund av dess fokus på hur systemet övertalar användaren till upprepad (och ibland oönskad) användning av delar av tjänsten. Ett exempel på detta är Snapchats funktion *snaps streaks* som uppmuntrar användaren att fortsätta skicka bildmeddelanden varje dag för att inte förlora sin *streak*. Samtliga DP definierade av Monge Roffarello et al. (2023) tycks därmed syfta till att fånga och upprätthålla användarens engagemang med plattformen.

4.1.5 Social Engineering

Social Engineering är en högnivå av DP som utnyttjar individens önskan att följa förväntade eller ålagda sociala normer. Detta genom att alternativ eller information presenteras på sätt som guidar individens handlande (Gray et al., 2024). Totalt hittades fyra unika DP inom denna strategi. De tillhör en mesonivå och tre lågnivåer. En av dessa mesonivåer är definierad i ontologin av Gray et al. (2024). Resten hittades och definierades av uppsatsens författare. Samtliga förekomna DP illustreras i Figur 9 nedan.



Figur 9 Visualisering av högnivån *Social Engineering* med samtliga funna meso- och lågnivåer

Mesonivån *Urgency* handlar om att manipulera användarens uppfattning att något erbjudande är limiterat och därigenom påskynda deras beslutsfattningsprocess för att få dem att handla (Gray et al., 2024). Exempel på detta är hur användaren kan se hur många som sökt en tjänst på LinkedIn, vilket kan leda till att användaren känner en brådska att agera.

En ny mesonivå och tre nya lågnivåer definierades av författarna. Den första mesonivån döptes till *Social Validation* och handlar om hur användarens behov av social validering utnyttjas av systemet. Till denna skapades två nya lågnivåer: *Ego Boost* och *Parasocial Engagement*. *Ego Boost* beskriver hur användarens önskan av social validering i form av gilla-markeringar, kommentarer, delningar etcetera, utnyttjas av systemet för att uppmuntra dem att skapa mer innehåll, något som leder dem till att spendera mer tid på plattformen. *Parasocial Engagement* handlar om hur användare uppmuntras till att interagera med kreatörer i hopp att få uppmärksamhet från- eller kunna inleda ömsesidiga relationer med dem. En önskan som utnyttjas av systemet och får användaren att spendera resurser de annars kanske inte skulle gjort. Ett exempel på detta är hur användaren kan köpa rosor och andra symboliska gåvor på TikTok att ge till kreatörer under livesändningar.

Den tredje nya lågnivån döptes till *Peer Pressure* och placerades under Gray et al. (2024) mesonivån *Social Proof*. *Peer Pressure* handlar om hur systemet övertalar användaren att interagera med innehåll på grund av dess upplevda popularitet (exempelvis gilla-markeringar och antal delningar). Detta kan leda till att användaren spenderar mer tid på plattformen med innehåll som kanske inte är av deras intresse.

4.1.6 Odefinierad hög- eller mesonivå

Slutligen hittades en ny trolig mesonivå, *Retention Blackmail*, som inte anses passa in i någon av Gray et al. (2024) högnivåer. Den observerades två gånger på Facebook och handlar om hur systemet manipulerar användaren att fortsätta använda dess tjänster på samma sätt om de vill behålla sina tidigare köp och prestationer. Ett exempel är när användaren ska börja använda Marketplace och ställs inför frågan om de vill använda en "anpassad upplevelse" eller ej. Med "anpassad" menas att Facebook tillåts nyttja användarens Facebook-uppgifter för att anpassa upplevelsen av tjänsten. Skulle användaren vilja byta till den "oanpassade" upplevelsen ges ultimatumet att all tidigare historik i den anpassade versionen kommer att raderas, såsom mailkonversationer. Då användaren inte *tvingas* till detta ansågs inte högnivån *Forced Action* lämplig. *Social Engineering* anses inte heller passande då *Retention Blackmail* inte har något med sociala normer att göra. *Retention Blackmail* handlar i stället om att användaren behåller satta inställningar på grund av att de räds att förlora något.

4.2 Analys

För högnivån *Sneaking* identifierades tre unika underordnade DP. Strategin försöker manipulera användaren att ta vissa handlingar genom att undanhålla, förkläda eller försena information. Detta undergräver användarens autonomi genom att de inte tillåts ta välinformerade beslut. Att medvetet undanhålla information på detta sätt är ett tydligt exempel på hur SNS-företagen exploaterar användare och visar att de saknar respekt gentemot dem (Bhargava & Velasquez, 2021).

Sju DP hittades för *Obstruction*, varav två formulerades av författarna. Den övergripande strategin som används här är att manipulera de val som användaren gör genom att göra vissa alternativ svårare än nödvändigt. Genom *Adding Steps* och *Privacy Mazes* avråds användaren från att fatta vissa beslut, speciellt beslut relaterade till hur deras personliga information ska användas. Med *Creating Barriers* och *Intermediate Currencies* skymms och kompliceras vissa uppgifter för att hindra användaren från att försöka åstadkomma dem. *Roach Motel*, *Eternal Presence*, *Immortal Accounts* och *Sticky Engagement* försöker se till att låsa in användaren när det kommer till vissa beslut genom att göra dem svåra att ångra. Det görs således aktiva försök att inskränka användares autonomi genom att försvåra vissa handlingar, och därigenom avråda användare från att välja dem.

För *Forced Action* hittades elva unika underordnade DP, varav fem definierades av författarna. Strategin försöker övertyga användaren till vissa handlingar genom att tvinga dem att medvetet eller omedvetet utföra vissa handlingar. Här används olika tillvägagångssätt, varav tjat (*Nagging*) och uppmärksamhetsfångande (*Attention Capture*) är de vanligaste. Systemet försöker bryta ned användarens egen vilja och undergräver därmed deras autonomi. *Forced Action* är också den högnivå som innehåller flest identifierade undernivåer. Det var här författarna identifierade flest nya DP och placerade tre funna *Attention-Capture Damaging Patterns* (ACDP) formulerade av Monge Roffarello et al. (2023) (*Infinite Scroll*, *Recapturing Notifications* och *Playing by Appointment*). Tre lågnivåer som alla ämnar stjäla användarens uppmärksamhet och hålla dem kvar på plattformen, bland annat genom att ta bort naturliga stoppsignaler (Bhargava & Velasquez, 2021).

Under *Interface Interference* identifierades åtta unika instanser av DP, varav en definierades av författarna. Strategin försöker utnyttja kognitiva begränsningar för att leda eller förvirra användaren genom att manipulera gränssnittet. Detta görs bland annat via plattformens standardinställningar som är inställda på att användarens information och innehåll är publikt (*Bad Defaults* och *Public Display*). Strategin försöker även manipulera användarens beslut genom att framhålla vissa alternativ, distrahera eller överväldiga användaren med information (*Manipulating Choice Architecture*, *False Hierarchy*, *Visual Prominence* och *Choice Overload*). Manipulativt språk, färger och andra designelement kan också användas (*Trick Questions* och *Emotional and Sensory Manipulation*). Denna kognitiva manipulation menar Bhargava och Velasquez (2021) utnyttjas av SNS. *Interface Interference* används således för att undergräva användares autonomi genom att framhålla vissa handlingar och därigenom leda användaren gentemot vissa handlingsalternativ.

Från högnivån *Social Engineering* identifierades fyra unika instanser av DP, varav tre formulerades av författarna. Denna högnivå utnyttjar användarens kognitiva sårbarheter för att få dem att följa förväntade eller satta normer i samhället. Användning av funktioner som utgör belöningssystem som utnyttjar det mänskliga behovet av validering och gemenskap är något Bhargava och Velasquez (2021) lyfter fram som vanligt på SNS. Genom *Urgency* och *Peer Pressure* försöker systemet manipulera användaren att varor eller innehåll är värdefulla för dem. Författarnas mesonivå *Social Validation* beskriver hur systemet utnyttjar användarens behov av validering och samhörighet i form av synliga bevis av uppskattning (*Ego Boost*), men försöker även belysa hur användares önskan om validering kan leda till att de spenderar resurser på kreatörer i hopp på uppmärksamhet (*Parasocial Engagement*). Därmed undergrävs användarens autonomi genom att deras mänskliga sociala behov exploateras för att få dem att agera på ett specifikt sätt.

5 Diskussion och slutsatser

Nedan diskuteras studiens resultat utifrån den analys som har gjorts och sammanfattas sedan i slutsatser. Vidare diskuteras begränsningar för studien, förslag för framtida forskning, resultatetiska konsekvenser och vilka tekniska verktyg som använts under studien.

5.1 Diskussion

När SNS använder sig av adaptiva algoritmer för att anpassa innehållet för att behålla användarens uppmärksamhet, eller utveckla mer effektiva sätt att tjata på användaren, kan högnivåstrategin *Forced Action* anses vara den mest beroendeframkallande och därav mest moraliskt förkastliga högnivån då den bland annat ämnar hålla användarens uppmärksamhet på plattformen. Som Bhargava och Velasquez (2021) poängterar är det mängden *tid* som den internetberoende spenderar på teknologin som avgör hur mycket skada de erfar. Kanske är det därför många unika instanser utav denna identifierades eftersom det är tidsaspekten som är viktigast för uppmärksamhets-ekonomin och de SNS-företag som opererar inom den. Bhargava och Velasquez (2021) betonar vikten av att respektera användarens autonomi, och *Forced Action* tillsammans med *Obstruction* verkar vara de mest uppenbara exemplen på högnivåstrategier som tydligt motverkar denna princip.

En vanlig företeelse var att uppgifter gjordes mer komplicerade än nödvändigt, vilket uppnåddes genom högnivåstrategin *Obstruction*. Det är viktigt att notera att detta kan bero mer på dålig design än på ett medvetet försök att manipulera användaren. Bland alla högnivåstrategier kan *Obstruction* vara den där detta framträder tydligast, eftersom det är relativt enkelt att komplicera saker genom dålig design, jämfört med exempelvis att framhäva vissa handlingar över andra som i *Interface Interference*, där det sannolikt är gjort med avsikt. Oavsett om de fall som identifierades under *Obstruction* var medvetet designade eller inte, klassificeras de som DP. Detta betonas av Gray et al. (2024), som framhåller att DP är designmönster som underminerar användarens beslutsförmåga och autonomi, oavsett deras avsikt.

Att undanhålla information kan betraktas som en effektiv strategi för att manipulera användarens beteende, vilket gör det anmärkningsvärt att högnivåstrategin *Sneaking* hade minst antal underordnade unika och identifierade DP av alla högnivåer. En möjlig förklaring till detta är att ontologin enligt Gray et al. (2024) delvis inkorporerar aspekter av att undanhålla information under andra högnivåstrategier. Till exempel faller *Roach Motel* under högnivåstrategin *Obstruction*, vilken beskriver situationer där användaren luras att utföra handlingar under förespeglingen att dessa ska vara enkla att återkalla – således kan bristen på information betraktas som en bidragande faktor. En ytterligare anledning till att så få DP identifierades under denna högnivå kan vara de specifika former i vilka de manifesterar sig, vilket gör dem svåra att identifiera och särskilt bedrägliga.

Högnivåstrategin *Interface Interference* var den med näst flest identifierade DP. Mesonivån *Bad Defaults* och dess underordnade *Public Display* var bland de mest noterade DP på de nio SNS som undersöktes. De flesta av dessa mönster återfanns i plattformarnas standardinställningar, vilket tyder på att SNS-företagen vill uppmuntra användare att dela information eller godkänna vissa villkor utan möjlighet till reflektion. Användare ges inte möjlighet att göra aktiva val från början utan förväntas acceptera omfattande delning av personlig information; annars måste de aktivt söka upp och ändra

inställningarna. Samma princip gäller för notifikationer, som alltid verkar vara påslagna för samtliga alternativ. Detta antyder att plattformarna vill maximera användarnas tillgänglighet både i termer av synlighet på plattformen och genom notifikationer.

Vad gäller notifikationer, var mesonivån *Manipulating Choice Architecture* och dess underordnade DP också vanligt förekommande. Detta pekar på att direkt manipulation av användargränssnittet kan vara ett kraftfullt verktyg för att övertyga användare att fatta vissa beslut.

De tidigare identifierade DP för *Social Engineering* i Gray et al. (2024) ontologi är endast delvis tillämpliga på SNS. Därför valde författarna att inkludera SNS-specifika definitioner enligt Monge Roffarello et al. (2023). Trots dessa anpassningar fann författarna ett tomrum när det gäller hur social validering utnyttjas för beroendeframkallande design på SNS. Detta går i linje med tidigare uttalanden om att mer forskning om DP på SNS behövs. De DP som identifierades och definierades under *Social Engineering* i denna studie kan ses utgöra grundläggande funktioner hos SNS – såsom gilla-markeringar och antal visningar.

Redström (2006) talar om hur det vanligtvis pågår en sorts dialog mellan designer och användare där användaren kan välja att acceptera, neka eller modifiera en design efter eget tycke. Med bakgrund i uppmärksamhets-ekonomin och hur SNS individanpassas med användning, samt hur DP medvetet spelar på användares psykologiska sårbarheter för att manipulera och utnyttja dem tycks dialogen i stället vara en monolog. Som Bhargava och Velasquez (2021) skriver skulle SNS-företagen kunna använda sin beroendeframkallande design för det motsatta – att uppmuntra användare till hälsosam SNS-användning. Som Baughan et al. (2022) beskriver finns ett flertal verktyg för detta som SNS-företagen skulle kunna använda sig av för att öka användarnas digitala välmående – detta gäller såväl interna som externa interventioner. Ett verktyg som författarna noterade på flera ställen var det som Monge Roffarello och De Russis (2019) lyfter om tidsbruksspårning. Dock var detta, och möjligheten att pausa notifikationer, det enda verktygen för digitalt välmående som noterades. Användarna behöver dessutom aktivt söka efter dem. *Digital nudging*, som tas upp av Purohit et al. (2020), skulle kunna anammas mer av SNS-företag för att stärka digitalt välmående. Många av de DP som identifierats skulle kunna göras till medvetna val, likt det gränssnitt som Purohit et al. (2020) undersökte i sitt tillägg. Om användaren ges möjligheten att aktivt välja eller neka sådana DP som *autoplay* och *infinite scroll* vid varje användningstillfälle skulle deras negativa effekter kunna dämpas. Ansvar ligger delvis hos gränssnittsdesigners men i slutändan ligger det hos de som anställer dem – de som aktivt väljer att exploatera sina användare på ett moraliskt förkastligt sätt – nämligen SNS-företagen.

5.2 Slutsatser

Syftet med denna studie var att undersöka förekomsten av *Dark Patterns* (DP) på sociala medieplattformar (SNS), vilka övertygande argument deras design innehåller, samt vad DP har för påverkan på användares autonomi. Antalet unika DP som identifierades var 34 över nio olika SNS. Av dessa definierades tretton av författarna. De tretton nya DP som identifierades gavs nya definitioner eftersom de var svåra att placera i ontologin av Gray et al. (2024). Resten identifierades med hjälp av tidigare definitioner. Följande unika DP per högnivå identifierades:

- *Sneaking*: en mesonivå och två lågnivåer.
- *Obstruction*: två mesonivåer och fem lågnivåer.
- *Interface Interference*: fyra mesonivåer och fyra lågnivåer.
- *Forced Action*: två mesonivåer och nio lågnivåer.

- *Social Engineering*: en mesonivå och tre lågnivåer.

De övertygande argumenten hos DP, som studien ämnade undersöka, återfinns i deras definitioner. Ett exempel på detta är mesonivån *Bad Defaults*, där standardinställningarna förväntas vara i användarens bästa intresse. Användaren övertygas således att tro detta och därmed kanske inte skapar en egen uppfattning. DPs och deras övertygande argument används således i olika syften för att undergräva användarens autonomi – däribland för att göra det svårare för användaren att fatta beslut, försvåra deras handlingar, få dem att stanna kvar på plattformen eller på annat sätt utnyttja användarens psykologiska sårbarheter för systemets vinning.

5.3 Begränsningar

Eftersom validitet baseras på teoretisk förståelse och tidigare forskning (Hjerm et al., 2014) bedöms studien ha en god vetenskaplig grund i ontologin av Gray et al. (2024). Författarnas begränsade kunskap i ämnet kan dock ha påverkat validiteten i hur väl DP identifierades. Ett DP som inte identifierades är *Guilty Pleasure Recommendations*, definierad av Monge Roffarello et al. (2023), som handlar om att användare får förslag på liknande innehåll när de konsumerar innehåll. Detta kan ha att göra med omständigheterna kring hur studien utfördes, bristande kunskaper eller DP-blindhet.

Datainsamlingens tidsbegränsning på 30 minuter per SNS kan också påverkat validiteten negativt vad gäller vilka DP som identifierades. Exempelvis kräver *Recapturing Notifications*, definierad av Monge Roffarello et al. (2023), att användaren varit inaktiv på plattformen – något som inte tilläts ske då samtliga applikationer avinstallerades direkt efter användning.

För att stärka studiens reliabilitet kunde ytterligare en datainsamlingsmetod använts för bättre förståelse av området och frågeställningarna (Denscombe, 2014). Dokumentstudier kunde exempelvis kombinerats med intervjuer eller enkäter för att få tillgång till användares förståelse för DP och hur de påverkar dem. Reliabiliteten kom dock att stärkas genom nyttjandet av det protokoll som tidigare formulerats och använts av Di Geronimo et al. (2020) och Mildner et al. (2023a) i studier av DP på SNS. Vissa punkter i protokollet kom även att förtydligas av författarna.

5.4 Framtida forskning

För framtida forskning föreslås en långtgående kartläggning, alternativt fallstudie, för att kunna fånga upp DP som inte visar sig vid kort användning. Att involvera personer med expertis i gränssnittsdesign, psykologi och kognitionsvetenskap skulle också kunna vara givande, då det skulle ge bättre förutsättningar för att kunna identifiera DP, hur de övertalar, samt vilka effekter de kan ha på individer.

Med tanke på att 92% av barn och unga (8–19 år) använder SNS dagligen föreslås forskning på mobilapplikationer som är populära i denna ålderskategori. För att titta på förekomsten av DP där samt för att undersöka hur barn och unga kan skyddas från beroendeframkallande design.

5.5 Resultatetiska konsekvenser

De resultat som framkommit i denna studie förväntas inte ha några förutsebara tillämpningar eller konsekvenser av militär natur. Detsamma gäller de processer eller kontrollmekanismer som står i konflikt med internationell rätt.

Vad gäller samhällliga följder av omvälvande karaktär, såsom en kraftigt minskad arbetsmarknad eller andra negativa sysselsättningseffekter, kan inga sådana effekter identifieras som en följd av detta arbete. Inga negativa effekter på individers försörjningsmöjligheter kan heller skönjas.

Slutligen, beträffande personlig integritet och yttrandefrihet, finns det inga förutsägbara negativa konsekvenser baserade på studiens resultat.

5.6 Tekniska verktyg

Ett flertal tekniska verktyg användes i studien. Användningen underlättade arbetet och möjliggjorde en optimerad process.

- ChatGPT användes för översättning och för att hitta lämpliga artiklar (något som Google Scholar visade sig vara bättre på).
- Discord användes för kommunikation mellan författarna.
- Draw.io användes för att visualisera data under den tematiska analysen.
- Google Docs användes för att organisera idéer samt skriva och revidera uppsatsen.
- Google Drive användes för att organisera och dela studiens dokument.
- Google Scholar användes för att söka efter relevant litteratur.
- Samsungs inbyggda skärminspelningsverktyg användes för datainsamlingen.

Källförteckning

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5: e uppl.). American Psychiatric Publishing, Inc. <https://doi.org/10.1176/appi.books.9780890425596>
- Andersson, J., Blomdahl, F. & Bäck, J. (2023). *Svenskarna och internet*. Internetstiftelsen. <https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2023/>
- Baughan, A., Zhang, M. R., Rao, R., Lukoff, K., Schaadhardt, A., Butler, L. & Hiniker, A. (2022). “I Don’t Even Remember What I Read”: How Design Influences Dissociation on Social Media. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI ’22)*, 1-13. <https://doi.org/10.1145/3491102.3501899>
- Bhargava, V. R. & Velasquez, M. (2021). Ethics of the Attention Economy: The Problem of Social Media Addiction. *Business Ethics Quarterly*, 31(3), 321–359. <https://doi.org/10.1017/beq.2020.32>
- Braun, V. & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi-org.ezp.sub.su.se/10.1191/1478088706qp063oa>
- Brignull, H. (2023). *Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You*. Testimonium Ltd.
- Buchanan, R. (1989). Declaration by Design: Rhetoric, Argument, and Demonstration in Design Practice. I Margolin, V. (Red). *Design Discourse; History, Theory, Criticism* (91–109). The University of Chicago Press.
- CPRA. (2022). California Privacy Rights Act. https://cppa.ca.gov/meetings/materials/20220608_item3.pdf [2024-04-04]
- Denscombe, M. (2014). *The Good Research Guide: For Small-Scale Social Research Projects*. (5:e uppl.). Open University Press.
- Di Geronimo, L., Braz, L., Fregnan, E., Palomba, F., & Bacchelli, A. (2020). UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3313831.3376600>
- Dunne, A. (2005). *Hertzian Tales; Electronic products, aesthetic experience and critical design*. The MIT Press.
- Gray, C. M., Bielova, N., Santos, C., & Mildner, T. (2024). An Ontology of Dark Patterns: Foundations, Definitions, and a Structure for Transdisciplinary Action. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI ’24)*, 22 sidor. <https://doi.org/10.1145/3613904.3642436>
- Gray, C. M., Santos, C., Tong, N., Mildner, T., Rossi, A., Gunawan, J. & Sinderson, C. (2023). Dark Patterns and the Emerging Threats of Deceptive Design Practices. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA ’23)*. <https://doi.org/10.1145/3544549.3583173>
- Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The Dark (Patterns) Side of UX Design. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 534, 1–14. <https://doi.org/10.1145/3173574.3174108>

- Hjerm, M., Lindgren, S., & Nilsson, M. (2014). *Introduktion till samhällsvetenskaplig analys*. (2:a uppl.). Gleerups. ISBN: 9789140686121
- Jaspers, M. W., Steen, T., van den Bos, C. & Geenen, M. (2004). The think aloud method: a guide to user interface design. *International journal of medical informatics*, 73(11–12), 781–795. <https://doi.org/10.1016/j.ijmedinf.2004.08.003>
- Johannesson, P. & Perjons, E. (2014). *An Introduction to Design Science*. Springer Cham. <https://doi.org/10.1007/978-3-319-10632-8>
- Judd, T. (2014). Making sense of multitasking: The role of Facebook. *Computers and Education*, 70, 194–202. <http://dx.doi.org/10.1016/j.compedu.2013.08.013>
- Kane, L. (2019). *The Attention Economy*. Nielsen Norman Group. <https://www.nngroup.com/articles/attention-economy/> [2024-04-06]
- Kojima, R., Sato, M., Akiyama, Y., Shinohara, R., Mizorogi, S., Suzuki, K., ... & Yamagata, Z. (2019). Problematic internet use and its associations with health-related symptoms and lifestyle habits among rural Japanese adolescents. *Psychiatry and Clinical Neurosciences*, 73(1), 20–26. <https://doi.org/10.1111/pcn.12791>
- Lanier, J. (2018). *Ten arguments for deleting your social media accounts*. New York: Henry Holt.
- Lu, L., Xu, N.J., Ge, X., Yue, W., Su, W.J., Pei, G. & Ma, L. (2002). Reactivation of morphine conditioned place preference by drug priming: role of environmental cues and sensitization. *Psychopharmacology*, 159, 125–132. <https://doi.org/10.1007/s002130100885>
- Luguri, J., & Strahilevitz, L. J. (2021). Shining a light on dark patterns. *Journal of Legal Analysis*, 13(1), 43–109. <https://doi.org/10.1093/jla/laaa006>
- Lukoff, K., Lyngs, U., Zade, H., Liao, J. V., Choi, J., Fan, K., Munson, S. A. & Hiniker, A. (2021). How the Design of YouTube Influences User Sense of Agency. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, 1–17. <https://doi.org/10.1145/3411764.3445467>
- Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark Patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on human-computer interaction*, 3(CSCW), 1–32. <https://arxiv.org/abs/1907.07032>
- Mathur, A., Mayer, J. & Kshirsagar, M. (2021). What Makes a Dark Pattern... Dark?: Design Attributes, Normative Considerations, and Measurement Methods. *CHI Conference on Human Factors in Computing Systems (CHI '21)*, 1–18. <https://doi.org/10.1145/3411764.3445610>
- Mildner, T., Freye, M., Savino, G-L., Doyle, P. R., Cowan, B. R. & Malaka, R. (2023a). Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. *Designing Interactive Systems Conference (DIS '23)*, 2362–2374. <https://doi.org/10.1145/3563657.3595964>
- Mildner, T., Savino, G-L., Doyle, P. R., Cowan, B. R. & Malaka, R. (2023b). 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 (CHI '23)*, 1-15. <https://doi.org/10.1145/3544548.3580695>
- Mirsch, T., Lehrer, C. & Jung, R. (2017) Digital Nudging: Altering User Behavior in Digital Environments. *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*, 634-648. <https://wi2017.ch/images/wi2017-0370.pdf>

- Monge Roffarello, A. M. & De Russis, L. (2019). The Race Towards Digital Wellbeing: Issues and Opportunities. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, 386, 1–14. <https://doi.org/10.1145/3290605.3300616>
- Monge Roffarello, A. M. & De Russis, L. (2023). Nudging Users Towards Conscious Social Media Use. *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction. (MobileHCI '23 Companion)*, 1–7. <https://doi.org/10.1145/3565066.3608703>
- Monge Roffarello, A. M., Lukoff, K. & De Russis, L. (2023). Defining and Identifying Attention Capture Deceptive Designs in Digital Interfaces. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, 1–19. <https://doi.org/10.1145/3544548.3580729>
- Narayanan, A., Mathur, A., Chetty, M. & Kshirsagar, M. (2020). Dark Patterns: Past, Present, and Future. *Communications of the ACM*, 63(9), 42–47. <https://doi.org/10.1145/3397884>
- Purohit, A. K., Barclay, L., & Holzer, A. (2020). Designing for Digital Detox: Making Social Media Less Addictive with Digital Nudges. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–9. <https://doi.org/10.1145/3334480.3382810>
- PwC. (2018). *IAB Internet Advertising Revenue Report*. Corporate Report.
- Redström, J. (2006). Persuasive Design: Fringes and Foundations. IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (Red.) *Persuasive Technology. PERSUASIVE 2006. Lecture Notes in Computer Science*, 3962. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11755494_17
- Scott, J. (1990). *A Matter of Record*. Cambridge: Polity.
- Vanden Abeele, M. M. (2021). Digital wellbeing as a dynamic construct. *Communication Theory*, 31(4), 932–955. <https://doi.org/10.1093/ct/qtaa024>
- Vetenskapsrådet. (2017). God forskningssed. *Rapport VR1708*. Vetenskapsrådet. <https://www.vr.se/analys/rapporter/vara-rapporter/2017-08-29-god-forskningssed.html>
- Ward, A. F., Duke, K., Gneezy, A. & Bos, M. W. (2017). Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity. *Journal of the association for consumer research*, 2(2), 140–154. <http://dx.doi.org/10.1086/691462>
- Wood, A. (2005). Exploitation. I T. Honderich (Red.), *The Oxford Companion to Philosophy*, (1:a uppl. 283–84). Oxford: Oxford University Press.
- Wu, T. (2016). *The Attention Merchants: The Epic Scramble to get Inside our Heads*. New York: Alfred A. Knopf.
- Zagal, J. P., Björk, S., & Lewis, C. (2013). Dark patterns in the design of games. *Foundations of Digital Games*.
- Zhao, Y., Li, T. & Sobolev, M. (2024). Digital Wellbeing Redefined: Toward User-Centric Approach for Positive Social Media Engagement. *IEEE/ACM 11th International Conference on Mobile Software Engineering and Systems (MOBILESoft '24)*, 1–4. <https://doi.org/10.48550/arXiv.2403.05723>

Bilaga A – Vald ontologi

I tabellen nedan listas ontologin av Gray et al. (2024) med dess hög-, meso- och lågnivåer.

Högnivå	Mesonivå	Lågnivå
Sneaking	Bait and Switch	Disguised Ads
	Hiding Information	Sneak into Basket
		Drip Pricing, Hidden Costs, or Partitioned Pricing
		Reference Pricing
	(De)contextualizing Cues	Conflicting Information
		Information without context
Obstruction	Roach Motel	Immortal Accounts
	Creating Barriers	Price Comparison Prevention
		Intermediate Currencies
	Adding Steps	Privacy Maze
Interface Interference	Manipulating Choice Architecture	False Hierarchy
		Visual Prominence
		Bundling
		Pressured Selling
	Bad Defaults	-
	Emotional or Sensory Manipulation	Cuteness
		Positive or Negative Framing
	Trick Questions	-
	Choice Overload	-
	Hidden Information	-
	Language Inaccessibility	Wrong Language
		Complex Language
	Feedforward Ambiguity	-

Forced Action	Nagging	-
	Forced Continuity	-
	Forced Registration	-
	Forced Communication or Disclosure	Privacy Zuckering
		Friend Spam
		Address Book Leeching
		Social Pyramid
	Gamification	Pay-to-Play
		Grinding
	Attention Capture	Auto-Play
Social Engineering	Scarcity or Popularity Claims	High Demand
	Social Proof	Low Stock
		Endorsements and Testimonials
		Parasocial Pressure
	Urgency	Activity Messages
		Countdown Timers
		Limited Time Messages
	Personalization	Confirmshaming

Bilaga B – Funna Dark Patterns

Nedan listas den samtliga unika *Dark Patterns* som identifierats, samt deras beskrivningar översatta till svenska från engelska. Dessa innefattar definitioner av Gray et al. (2024) och Monge Roffarello et al. (2023). De *Dark Patterns* som definierats av författarna är också inkluderade.

Samtliga mesonivåer för identifierade lågnivåer finns också med, även om de inte specifikt identifierades.

Sneaking är en strategi som döljer, förklarar eller fördröjer avslöjandet av viktig information som, om den gjordes tillgänglig för användarna, skulle få dem att oavsiktligt vidta en åtgärd de sannolikt skulle motsätta sig.

- ***Bait and Switch*** undergräver användarens förväntningar att deras val kommer att resultera i en önskad åtgärd, och leder i stället till ett oväntat och oönskat resultat.
 - ***Disguised Ads*** använder *Bait and Switch* och *Sneaking* för att utforma gränssnittselement så att de inte tydligt markeras som en annons eller annan partisk källa. Som ett resultat luras användarna att klicka på gränssnittselementet eftersom de antar att det är en relevant och viktig interaktion, vilket leder till att de omedvetet interagerar med annonsinnehåll.
- ***(De)Contextualizing Cues*** undergräver användarens förväntan att den tillhandahållna informationen kommer att vägleda användaren till att göra ett informerat val, istället förvirrar användaren och/eller förhindrar dem från att hitta relevant information på grund av sammanhanget där informationen presenteras.
 - ***Conflicting Information*** använder *(De)Contextualizing Cues* och *Sneaking* för att inkludera två eller fler informationskällor som motsäger varandra. Som resultat är användaren osäker på vad konsekvenserna av deras handlingar kommer att bli och är mer benägen att acceptera standardinställningar som kanske inte är i deras bästa intresse.

Obstruction är en strategi som försvårar användarens arbetsflöde och gör en interaktion mer besvärlig än vad den egentligen behöver vara, vilket avskräcker användaren från att vidta en viss åtgärd.

- ***Adding Steps*** undergräver användarens förväntan att en uppgift kommer att ta så få steg som tekniskt möjligt, i stället skapas ytterligare punkter av onödig men nödvändig användarinteraktion för att utföra en uppgift.
 - ***Privacy Mazes*** lägger till steg och använder *Obstruction* för att kräva att en användare navigerar genom många sidor för att få relevant information eller kontroll utan en omfattande och uttömmande översikt. Som ett resultat hindras användaren från att lätt upptäcka relevant information eller handlingsmöjligheter, vilket gör att de inte kan fatta informerade beslut gällande deras sekretess.

- **Creating Barriers** undergräver användarens förväntan att relevanta användaruppgifter kommer att stödjas av gränssnittet, i stället förhindrar, abstraherar eller på annat sätt komplicerar en användaruppgift för att avskräcka användaren från att agera.
 - **Intermediate Currencies** skapar hinder och använder *Obstruction* för att dölja den verkliga kostnaden för en produkt eller tjänst genom att kräva att användaren spenderar riktiga pengar för att köpa en virtuell valuta som sedan används för att köpa en produkt eller tjänst. Som ett resultat är användaren inte i stånd att fastställa den verkliga monetära kostnaden för en produkt eller tjänst, vilket leder till att de gör ett oinformerat köpbeslut baserat på en fördold kostnad.
- **Roach Motel** undergräver användarens förväntning att en åtgärd ska vara lika lätt att ångra som den är att utföra, och skapar i stället en situation som är lätt att hamna i, men svår att ta sig ur.
 - **Eternal Presence** skapar ett *Roach Motel* och använder *Obstruction* för att göra det svårt att logga ut från ett användarkonto när det väl har skapats och loggats in. Som resultat kan användaren spendera mer tid på plattformen på grund av notifikationer och den lägre tröskeln för att surfa.
 - **Immortal Accounts** skapar ett *Roach Motel* och använder *Obstruction* för att göra det svårt eller omöjligt att radera ett användarkonto när det väl har skapats. Som ett resultat kan användaren skapa ett konto eller dela data med den falska förutsättningen att de senare kan radera denna information, även om kontot och/eller datan sedan inte kan tas bort av användaren.
 - **Sticky Engagement** skapar ett *Roach Motel* och använder *Obstruction* för att göra det svårt för användaren att avsluta prenumerationen på socialt innehåll. Som resultat kan användaren prenumerera på en innehållsleverantör med den felaktiga uppfattningen att de senare kan avsluta prenumerationen lika enkelt.

Interface Interference är en strategi som gynnar specifika handlingar över andra genom manipulation av användargränssnittet, vilket förvirrar användaren eller begränsar upptäckten av relevanta handlingsmöjligheter.

- **Bad Defaults** undergräver användarens förväntning att standardinställningar är till deras bästa, och kräver i stället att användarna aktivt ändrar inställningar som kan orsaka skada eller oavsiktlig avslöjande av information.
 - **Public Display** använder *Interface Interference* och *Bad Defaults* för att göra användarens information och genererade innehåll helt synligt för andra. Som ett resultat kan användaren oavsiktligt dela mer om sig själv än förväntat.
- **Choice Overload** undergräver användarens förväntan att de val de gör ska vara förståeliga och jämförbara, i stället erbjuds för många alternativ för att jämföra eller uppmuntra användare att förbise relevant information på grund av mängden val som erbjuds.
- **Emotional and Sensory Manipulation** undergräver användarens förväntan att utformningen av webbplatsen kommer att tillåta dem att uppnå sitt mål utan manipulation, i stället förändras

språk, stil, färg eller andra designkomponenter för att framkalla en känsla eller manipulera sinnen för att övertyga användaren att utföra en viss handling.

- **Cuteness** använder *Emotional and Sensory Manipulation* och gränssnittsstörning för att infoga attraktiva signaler i utformningen av ett robotgränssnitt eller formfaktor. Som ett resultat kan en användare sätta orimligt mycket tillit till roboten, vilket leder till att användaren felaktigt eller ofullständigt bedömer riskerna med att interagera med roboten.
- **Feedforward Ambiguity** undergräver användarens förväntan att deras val sannolikt kommer att resultera i en åtgärd de kan förutse, istället tillhandahålls en diskrepans mellan information och tillgängliga åtgärder för användare som resulterar i ett resultat som är annorlunda än vad användaren förväntar sig."
- **Manipulating Choice Architecture** undergräver användarens förväntan att de alternativ som presenteras kommer att stödja deras önskade mål, i stället inkluderas en ordning eller struktur av alternativ som gör ett annat resultat mer sannolikt."
 - **False Hierarchy** manipulerar valarkitekturen, genom att använda gränssnittsstörning för att ge ett eller flera alternativ visuell eller interaktiv framträdande över andra, särskilt där objekt borde vara parallella snarare än hierarkiska. Som ett resultat kan användaren missförstå eller inte kunna jämföra sina alternativ korrekt, vilket leder till att de gör ett val baserat på en falsk eller ofullständig valarkitektur.
 - **Visual Prominence** manipulerar valarkitekturen, genom att använda gränssnittsstörning för att placera ett element relevant för användarens mål i visuell konkurrens med ett mer distraherande och framträdande element. Som ett resultat kan användaren glömma bort eller distraheras från sitt ursprungliga mål, även om det målet var deras primära avsikt.
- **Trick Questions** undergräver användarens förväntan att information som tillhandahålls om rabatter eller ett tidsbegränsat erbjudande för en produkt är korrekt, istället påskyndas användarens beslutsprocess genom att kräva omedelbar eller tidskänslig handling.

Forced Action är en strategi som kräver att användare medvetet eller omedvetet utför en ytterligare och/eller tangentiell handling eller ger information för att få tillgång till (eller fortsätta att ha tillgång till) specifik funktionalitet, vilket hindrar dem från att fortsätta sin interaktion med systemet utan att utföra denna handling.

- **Attention Capture** undergräver användarens förväntning att de har rationell kontroll över den tid de spenderar med att använda ett system, och i stället lurar dem att spendera mer tid eller andra resurser för att fortsätta använda det längre än de annars skulle.
 - **Animated Feed** använder *Forced Action* och *Attention Capture* för att automatiskt spela upp videor i ett flöde. Som resultat kan användaren förlora kontrollen över sin upplevelse av sociala medier, vilket leder till att de spenderar en oönskad mängd tid på plattformen.
 - **Auto-Play** använder *Attention Capture* som en typ av *Forced Action* för att automatiskt spela upp en ny video efter att en befintlig video har avslutats. Som ett

resultat kan användaren förlora kontrollen över sin tittarupplevelse, vilket leder till att de tittar på mer innehåll än de avsåg eller att de tittar på innehåll som är oväntat eller skadligt.

- ***Flickering Spotlight*** använder *Forced Action* och *Attention Capture* för att göra plötsliga förändringar i användargränssnittet för att framhäva specifika funktioner. Som resultat kan användaren bli distraherad och engagera sig med det, vilket avleder dem från den uppgift de höll på med och leder deras uppmärksamhet till något annat.
- ***Infinite Scroll (Monge Roffarello et al. 2023)*** är ett designmönster där, när användaren scrollar ner en mobilapp eller en webbplats på sin dator, mer innehåll automatiskt och kontinuerligt laddas längst ner. Trots sina fördelar kan *infinite scroll* bli en 'skadlig funktion' eller ett 'anti-mönster' som främjar oändliga användningssessioner.
- ***Noisy Feed*** använder *Forced Action* och *Attention Capture* för att automatiskt spela upp ljud i ett flöde. Som resultat kan användaren förlora kontrollen över sin upplevelse av sociala medier, vilket leder till att de spenderar en oönskad mängd tid på plattformen.
- ***Recapturing Notifications (Monge Roffarello et al. 2023)*** är notifikationer som medvetet skickas för att återfå uppmärksamheten hos en användare som har lämnat en digital tjänst under en period. Detta görs till exempel för att få dem att starta en ny användningssession.
- ***Forced Communication or Disclosure*** undergräver användarens förväntan att ett system endast kommer att begära information som behövs för att slutföra deras önskade mål, istället luras de att dela mer information om sig själva eller att deras information används för ändamål som de inte önskar.
 - ***Privacy Zuckering*** använder *Forced Communication or Disclosure* som en typ av *Forced Action* för att lura användare att dela mer information om sig själva än de avser eller skulle gå med på om de var fullt informerade. Som resultat antar användaren att den information de ombeds att tillhandahålla är avgörande för att använda tjänsten, även om denna information används eller säljs för andra ändamål.
- ***Gamification*** undergräver användarens förväntan att system funktionaliteten är baserad på överensstämmelse med användarens mål och behov, istället tvingas dem att få tillgång till aspekter av en tjänst genom upprepade (och kanske oönskad) användning av aspekter av tjänsten.
 - ***Playing by Appointment (Monge Roffarello et al. 2023)*** är ett skadligt mönster som genom *Gamification* tvingar användare att använda en digital tjänst vid specifika tider som definieras av tjänsten, snarare än av användaren.
- ***Nagging*** undergräver användarens förväntning att de har rationell kontroll över de interaktioner de gör med ett system, och i stället distraherar användaren från en önskad uppgift de fokuserar på för att framkalla en handling eller få dem att fatta ett beslut de inte vill göra genom att upprepade gånger avbryta användaren under normal interaktion.

- **Login Pestering** använder *Forced Action* och *Nagging* för att framkalla att användaren sparar sina inloggningsuppgifter på plattformen. Som ett resultat sänks barriären för att använda det, och användaren förlorar en möjlighet att aktivt reflektera över sitt val att engagera sig med plattformen.
- **Unnecessary Onboarding Steps** undergräver användarens förväntningar att de kan börja engagera sig med ett system på sina egna villkor, och i stället tvingar dem att slutföra vissa handlingar som inte borde vara nödvändiga innan de kan fortsätta.

Social Engineering är en strategi som presenterar alternativ eller information som gör att en användare är mer benägen att utföra en specifik åtgärd baserat på deras individuella och/eller sociala kognitiva fördomar, och därmed utnyttjar användarens önskan att följa förväntade eller pålagda sociala normer.

- **Social Validation** undergräver användarens förväntningar att deras behov av social bekräftelse inte kommer att utnyttjas, och exploaterar i stället detta behov för att få användaren att engagera sig mer med systemet.
 - **Ego Boost** använder *Social Engineering* och *Social Validation* för att få användaren att få uppskattning för sitt innehåll på plattformen. Som resultat uppmuntras användaren att skapa mer innehåll, vilket i sin tur leder till att de spenderar mer tid på plattformen.
 - **Parasocial Engagement** använder *Social Engineering* och *Social Validation* för att driva användaren till att engagera sig med innehållsskapare i hopp om att få uppmärksamhet eller starta en ömsesidig relation. Som resultat kan användaren spendera resurser som de annars inte skulle ha gjort på plattformen.
- **Social Proof** undergräver användarens förväntning att det beteende som andra visar i en specifik situation är korrekt eller önskvärt, och påskyndar i stället användarens beslutsfattande och uppmuntrar användaren att lita på felaktiga slutsatser genom den tillhandahållna informationen.
 - **Peer Pressure** använder *Social Engineering* och *Social Proof* för att övertala användaren att engagera sig med innehåll på grund av dess kommunicerade popularitet. Som ett resultat kan användaren känna sig tvingad att spendera tid på innehåll som egentligen inte är av intresse för dem.
- **Urgency** undergräver användarens förväntan att information som tillhandahålls om rabatter eller ett tidsbegränsat erbjudande för en produkt är korrekt, istället påskyndas användarens beslutsprocess genom att kräva omedelbar eller tidskänslig handling.

Retention Blackmail undergräver användarens förväntning att tidigare prestationer eller köp inte är beroende av att fortsätta engagera sig med systemet på ett specifikt sätt, och utpressar i stället användaren att fortsätta använda systemet som tidigare av rädsla för att förlora sina tidigare framsteg.