PATTERNS IN THE DARK

DECEPTIVE PRACTICES IN ONLINE INTERACTIONS

University of South Australia

Australian Research Centre for Interactive and Virtual Environments

A Report to the Data Standards Chair

Landscape Assessment: Dark Patterns

A Report to the Data Standards Chair

James Baumeister
Ji-Young Park
Andrew Cunningham
Stewart Von Itzstein
Ian Gwilt
Aaron Davis
James Walsh

Acknowledgements
Graphic Design: Benjamin Altieri
Illustration: Neven ElSayed

About UniSA's Australian Research Centre for Interactive and Virtual Environments

The University of South Australia's Australian Research Centre for Interactive and Virtual Environments (IVE) is a unique alignment of computer science, engineering, psychology and cognition, neuroscience, art, architecture, and design. Founded in 2O19 as a unification of a number of individual areas of expertise, the Centre explores multidisciplinary problems, where the human is at the centre of the solution. The Centre is inspired by the challenges of industry and society to achieve impactful outcomes through delivering world-leading research, developing global research talent, and top-performing PhD students. In collaboration with our industry partners, IVE investigates and combines world expertise in all digital and virtual environments, with computer science, engineering, psychology, neuroscience, art, architecture, and design to solve real-world problems.

Increasingly the problems being encountered in our digital lives are no longer solely technical problems, but problems that touch at the heart of human cognition, emotion, and basic human responses to stimuli. Good digital systems are no longer just created by software developers, rather holistic teams of software developers paired with designers, psychologists, and neuroscientists. Modern applications and web pages are now designed to leverage the user's biological response to stimulus, feeding people's need to infinitely scroll or engage.

IVE's expertise and contribution lies not just in researching and developing solutions for academic problems and industry, but also providing consultation and advice, offering the capability to generate grounded, evidence-based reports and whitepapers, as well as performing grounded, multi-disciplinary objective research focused on the fundamentals of human factors, and how that impacts our relationship with technology.

Contact **IVECentre@unisa.edu.au**

University of South Australia | Australian Research Centre for Interactive and Virtual Environments

UniSA Capability

The University of South Australia (UniSA) is Australia's University of Enterprise and has extensive experience in working with industry, and Defence. UniSA is the largest university in the state with 35,000 students, 2,900 staff, 220,000 alumni and 2,500 partnerships with global universities, research bodies, organisations and industry.

Internationally, UniSA is ranked within the world's top 50 universities under 50 years old, with a five-star rating for World Universities by QS World University Rankings. UniSA is globally recognized as the number 1 young university in Australia for industry collaborations.

In the Excellence in Research for Australia (ERA) assessment, 100% of UniSA research was rated at or above world-class. The university is ranked Number 1 in Australia for industry research impact and engagement and UniSA Business is ranked in the top 1% worldwide.

UniSA is agile and astute, and recognised internationally for relevance, equity and excellence. UniSA educates and prepares global learners from all backgrounds, instilling professional skills, knowledge and a capacity and drive for lifelong learning. UniSA is committed to excellence: excellence in learning, ongoing improvement and innovation, community building, leading effective organisation and management.

The University of South Australia is meeting future challenges through cutting-edge research and the education of tomorrow's professionals.

Disclaimer

This report is not intended to be read or used by anyone other than the Department of the Treasury.

The University of South Australia (UniSA) prepared this report solely for the Department of the Treasury's use and benefit in accordance with and for the purpose set out in the Order of Service with The Department of the Treasury dated 2 February, 2024. In doing so, UniSA acted exclusively for The Department of the Treasury and considered no-one else's interests.

UniSA accepts no responsibility, duty, or liability:

- to anyone other than the Department of the Treasury in connection with this report, or
- to the Department of the Treasury for the consequences of using or relying on it for a purpose other than that referred to above.

UniSA makes no representation concerning the appropriateness of this report for anyone other than the Department of the Treasury. If anyone other than the Department of the Treasury chooses to use or rely on it, they do so at their own risk.

This disclaimer applies:

- to the maximum extent permitted by law and, without limitation, to liability arising in negligence or under statute; and
- even if we consent to anyone other than the Department of the Treasury receiving or using this report, including publication.

Purpose Statement

This report was commissioned pursuant to an Order of Work between the University of South Australia and the Department of the Treasury dated 2 February 2024. This report is specifically tailored to the requirements of the Data Standards Chair (Chair) and is to be read within the context of the Consumer Data Right (CDR). The purpose of this report was "to identify Dark Patterns that are relevant and likely to be used in the CDR, specifically in relation to the provision of consumer consent and consent management". Consideration of the regulatory environment was not requested.

Intended audience

The Chair is the primary owner and audience of this report. The report is also intended to be published and shared with external stakeholders as part of the Chair's requirements to consult.

Executive Summary

Dark patterns, otherwise known as deceptive patterns, are deceptive and manipulative tactics, present in online websites and applications (apps), that can be used to negatively influence a person's decision to perform a particular action. Shown to be present in all of the top social media services, 95% of the world's most popular mobile apps, and over 11% of the world's top shopping websites, the scale and reach that deceptive patterns have to exert their manipulative motivations is frightening. Driven by the desire for increased quality and quantity of personal data, commercial entities are targeting deceptive patterns and the personalisation afforded by artificial intelligence to supercharge deceptive patterns toward consumers. This report examines the ways in which deceptive patterns have infiltrated our online services and what harms they are causing.

In **Part I** of this report we examine how deceptive patterns exploit the vulnerabilities exposed by cognitive biases present in our fundamental human psychology. By understanding how our decision making processes can be subconsciously influenced, we can become aware of why deceptive patterns have managed to be as successful as they are. Whereas deceptive patterns rise and fall in popularity in response to technological advances, legislative reform, and website and app design trends, our cognitive biases are constant. Understanding the unchanging, foundational basis of these patterns provides a stronger footing to influence policy.

We present the IVE deceptive patterns typology in **Part II,** which forms a comprehensive overview of the myriad deceptive patterns identified by researchers in the field. We consolidate these deceptive patterns into a model that focuses on the protection of consumer autonomy. The typology serves as a directory of deceptive patterns, useful for regulators, software developers, and the general public as a reference for what constitutes a deceptive pattern and what not to do when influencing consumers. The use of a model for categorising the patterns gives this reference stability for the future, as it is unlikely that the underlying model will change as new deceptive patterns emerge.

In **Part III** we explore the landscape of deceptive pattern research. We show how deceptive patterns have pervaded the online and mobile app spaces, influencing the behaviour of consumers in their consent to data access, consumption of social media, and engaging in online shopping. We examine the research through the lens of the rising concern of artificial intelligence, envisioning how emerging technologies relating to large language methods, mass data aggregation, and user profiling could shape a new generation of even more powerful and effective deceptive patterns.

The deceptive pattern landscape has shown that commercial entities and consumers are locked in an adversarial relationship over individual privacy, autonomy, and data rights. Tensions raised by manipulative tactics in physical retail stores, with one-in-four Australians reportedly confused about promotional price tags in stores, are also present online, with 40% of Australians reporting annoyance when using a website. While deceptive patterns can provide commercial entities with a pathway to short term profit, research has shown that larger long term benefits can be gained by fostering trustworthiness and reliability through

Contents

Introduction

We live in an online world. Data has become the central powerhouse for almost all of the online services that we utilise daily. This reality has led to commercial entities developing a craving for this data, as it is pivotal in monetising their business strategies. These entities employ an array of tactics, both overt and covert, to gather as much information as possible. The primary driving mechanism behind these data gathering activities is maximising profit through a better understanding of their potential customers. Their profit-oriented mindset often means that consumer protection may not necessarily be a primary concern of their operations. Moreover, these tactics are not static. They are continuously evolving, often becoming more sophisticated and increasingly harder to detect and defeat. The most nefarious of these tactics are referred to as deceptive or manipulative patterns, **deceptive patterns**[1] for short. These are manipulative strategies used to trick consumers into sharing more data than they would willingly or knowingly share, or engage in ways they would not have otherwise engaged.

The state of regulation in respect to deceptive patterns presents a complex picture. The most blatant and egregious deceptive patterns are protected against under Australian consumer law, enshrined in the legislation such as the Competition and Consumer Act [141] and Spam Act [142].

Some examples include commercial entities misleading customers about their products and services, presenting misleading pricing strategies, and preventing users from unsubscribing from marketing communications. When it comes to more subtle patterns, however, the regulatory landscape becomes less clear. Among these more subtle patterns are emotional manipulation tactics such as **Confirmshaming**, **Fake Social Proof**, and **Safety Blackmail**. As these commercial entities are not breaking any specific laws, they are given passive permission to exploit these loopholes. They cleverly use these tactics to deceptively and manipulatively achieve their goals, often at the expense of the consumer. As such, the current state of regulation, while covering the more obvious deceptive patterns, still leaves room for these more subtle manipulative strategies to thrive.

This report explores the landscape of deceptive patterns both within Australia and on a global scale. It is intended to enlighten Australian policy makers about the nature of deceptive patterns and the reasons why there should be a cause for concern and act as a call to action. The potential harm of these deceptive patterns is discussed, along with an analysis of the different types of deceptive patterns, how they function, and their prevalence. The report delves into the various approaches towards investigating and taking actions against these deceptive patterns within the regulatory bodies, academic community, and news media. **Figure 1** visualises where deceptive patterns fit within the current regulatory environment.

This landscape assessment has three main aims:

1. Inform the reader on the existence and dangers of deceptive patterns.
2. Provide a clear method of identifying and classifying deceptive patterns so the reader is aware of them and policy makers can work to prevent them.
3. Describe the current state of deceptive pattern research.

1 Deceptive patterns are more commonly referred to as "dark" patterns. In recognition that the usage of "dark" in this way is non-inclusive, UniSA prefers deceptive patterns, which is also a more descriptive term.

The remainder of this report is divided into three parts, with each part respectively corresponding to the aims.

Part I introduces deceptive patterns as deceptive and/or manipulative tactics, describing how they are modern extensions of preexisting psychological vulnerabilities that have been exploited in other domains.

Part II presents the IVE deceptive patterns typology, which is a comprehensive list and categorisation method for the currently identified deceptive patterns. It details why a new typology was required for this report and how it was developed.

Part III outlines the current state of deceptive pattern research. It details how a systematic literature review was conducted and the main themes of this review are presented. It concludes with some overall observations about the impacts deceptive patterns have in different domains and what is and can be done about them.

Landscape of Deceptive Patterns

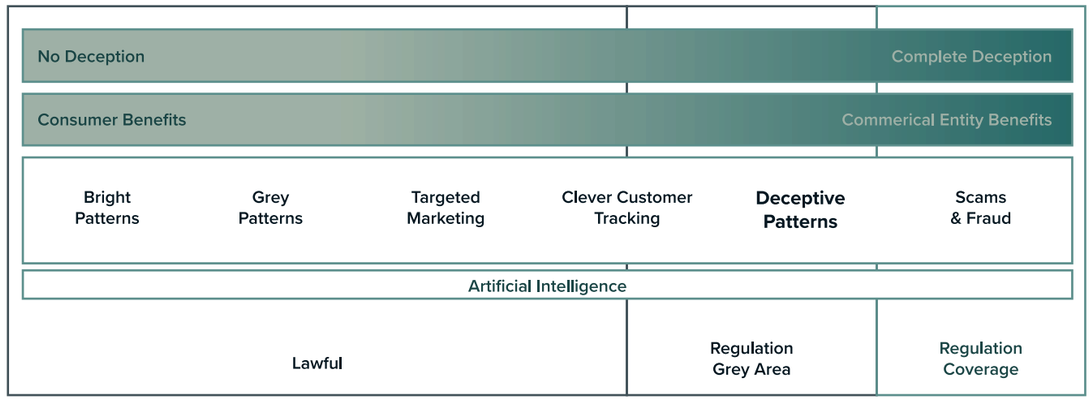
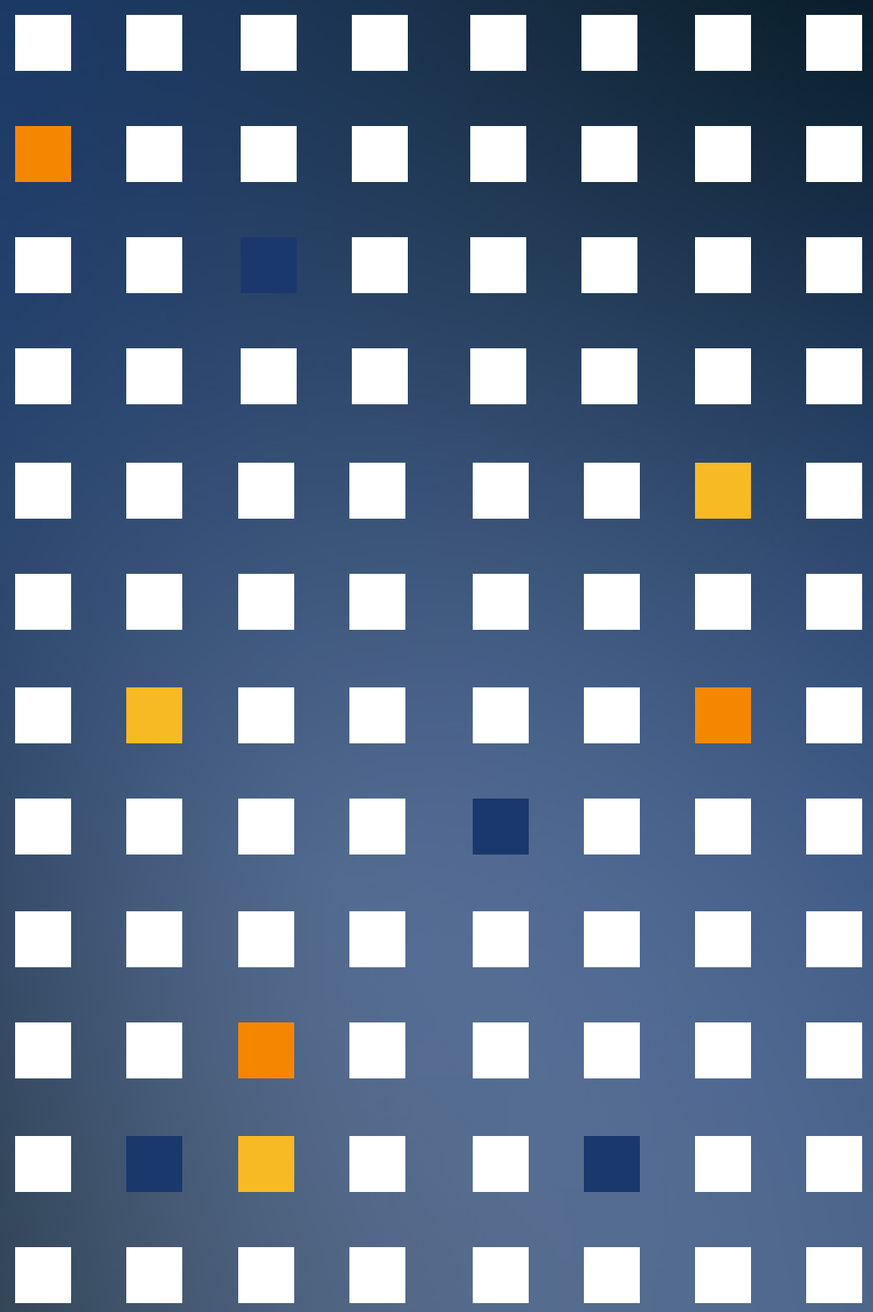


Figure 1. This visualisation shows that deceptive patterns exist in a spectrum of practices that include a degree of deception, benefit and legality.



Part I:
Deceptive Patterns

Deceptive Patterns

While deceptive patterns refer to deceptive and manipulative tactics in the digital realm, they were adopted into that realm from real-world patterns. Consider a real-world situation of shopping in a department store. A long history of analysing customer behaviour has taught the retail giants that consumers respond very well to specific colours, smells and packaging. You may have noticed that the most frequently purchased products are located at the back of the store ensuring the consumer is navigated through a maze of aisles, specials, and other products to find the items they are really looking for (see **Figure 2**). This is a deliberate tactic aimed to:

- Maximise the time spent in the store;
- Expose the consumers to visually and emotionally appealing colours, fonts, merchandising standards, and packaging; and,
- Allow other tactics, such as exposure to freshness, eye-level products, and end-of-aisle fixtures, to work their devious magics on the consumer.

Most recently such tactics have been coming to light within the Australian shopping experience as part of submissions to the Senate Select Committee on Supermarket Prices, with one-in-four Australians having difficulty identifying whether promotional price tags actually represent any saving [**102**, **137**].

The benefit of these tactics are apparent; the more consumers are exposed to marketing and sales tricks, the more products will be sold and the more money the retailer will make. The real-world has factors that limit the prevalence of these deceptive patterns. In the department store example, the physical space needs to be designed,

Figure 2. Department stores use manipulative strategies to maximise customer exposure to products across almost every interaction a customer has with them.

constructed, merchandised, and maintained, which is an expensive and complex process.

The seminal work from Thaler and Sunstein [**118**], discusses the concept of nudges and choice architecture. This concept stems from the observation that all individuals make numerous decisions each day, with these decisions often being made from a range of presented options. These options are often presented by a third party, which Thaler and Sunstein refer to as a choice architect. For example, when voting in local elections, the choices are presented on a ballot. The ballot is specifically designed to present the choices in an impartial manner, thus the ballot designer is acting as the choice architect. In another example, a doctor presents a range of treatment options to a patient, usually weighted according to efficacy research. The doctor thus acts as the choice architect, presenting the options in a best to worst range. A final example can be found in online content, such as a post about the best antivirus software written on a popular antivirus company's website. The company, as the choice architect, presents various antivirus choices, listing pros and cons, and ranks its own software as the best.

In all these examples, the third party has acted as a choice architect. Choices are presented to the individual, but the manner in which those choices are presented also involves some degree of thought and intent. This method of presentation is what Thaler and Sunstein [**118**] refer to as a nudge. Nudges serve to influence individuals to make a particular decision. In the case of the ballot, the nudge is simply to select one of the candidates, with electoral regulations ensuring that bias is eliminated to the greatest possible extent. In the case of the doctor, the nudge ideally leads the patient to the treatment that is most likely to be effective and with the least potential harm. There is potential, however, for the doctor's nudge to be influenced by other factors, such as promoting medication that results in the most commission from the pharmaceutical company. In the case of the anti-virus company, the nudge is clearly self-serving. The company desires for readers to conclude that their software is the best and therefore purchase it. The nudge is designed to push the reader toward becoming a customer of the company.

The nudge concept helps us understand what department stores are doing and why it is effective. Now consider online environments, namely websites and mobile applications (apps). Unlike the department store, a website can be rapidly created and updated frequently to adjust to new customer data and to target new populations of relevance. In order to use a website, visitors agree (implicitly or explicitly) for data about themselves to be collected. This can vary from as little as their IP address[2], to full demographic information and data about other visited websites.

2 An Internet Protocol (IP) address is a numerical label assigned to every device connected to the internet and can be considered personally identifiable information.

A physical retail store's ability to collect data is limited to the behaviours exhibited by the shopper in the store. Armed with a website and app, however, a company can collect *much* more information, even when the customer is not actively shopping. When you visit a website or use a mobile app, the site is performing clandestine data collection. They are logging and analysing how long you spend on a particular item, what you clicked after viewing that page, what types of items you have in your cart and what that might mean about you, and even what you type into an unsubmitted online form[3]. Every action you perform can have meaning and the service is employing all its technological capacity to ascertain and commercialise your behaviour.

The value of collecting more data means more insights, better targeting, and ultimately more sales. In addition, the advent of data brokers has created companies that operate purely on capturing and selling data about users. In a competitive environment, the need to gather as much insight about potential customers as your competitors also becomes a pressure to bend the principles of even ethical companies. This is especially true with the proliferation of new off-the-shelf artificial intelligence (AI) tools designed to enhance customer recruitment, retention, recommendations, etc. This use of personalised, algorithmic or AI-assisted nudging has been defined as a hypernudge [**138**].

The pursuit of the required data to achieve better outcomes can result in bad-actors compelling people to release unintended personal information. In the digital space, actors wishing to maximise the effectiveness of their data collection goal may employ manipulative tactics to convert website visitors into members of an ostensibly innocent newsletter subscription, giving said actors at least an email address to target with advertising and use as a point of data matching with data acquired from data brokers. The advertising may then promote a rewards program that enables access to specific member-only specials. Signing up to a rewards program is free, and just requires an account with some mandatory demographic data; that gives basic information about age and geographic location. Being an online member gives the customer the ability to place orders online. When they do so, they can opt into receiving mobile phone notifications via an app to alert the customer to when the delivery truck is nearby; that installs an app on the customer's phone and enables push notifications and potential location tracking.

Being an online member gives the customer the ability to place orders online. When they do so, they can opt into receiving mobile phone notifications via an app to alert the customer to when the delivery truck is nearby; that installs an app on the customer's phone and enables push notifications and potential location tracking.

From simply walking into a department store, our customer has now enabled a rich, informative, and very valuable profile to be built about themselves. This is visualised in **Figure 3**. It is easy to see that most companies are incentivised to convince their customers to provide as much data as possible. When these tactics of persuasion are deceptive and/or manipulative in the digital space, they are called deceptive patterns.

3 There is a whole market of form analytics services that developers can implement into their websites to keep track of forms and, in the most invasive case, follow-up with the potential client if contact details were provided in the partial completion.

Before we discuss deceptive patterns, it is important that we present a definition. For the purposes of this report and the overall goal to inform and protect against deceptive patterns, it is important that our definition focus on the involuntary nature of deceptive pattern influences and their impact on consumer autonomy. As such, this report will endorse the definition provided by the European Union *(see highlight)*.

Deceptive patterns on online interfaces of online platforms are practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions. Those practices can be used to persuade the recipients of the service to engage in unwanted behaviours or into undesired decisions which have negative consequences for them [**145**].

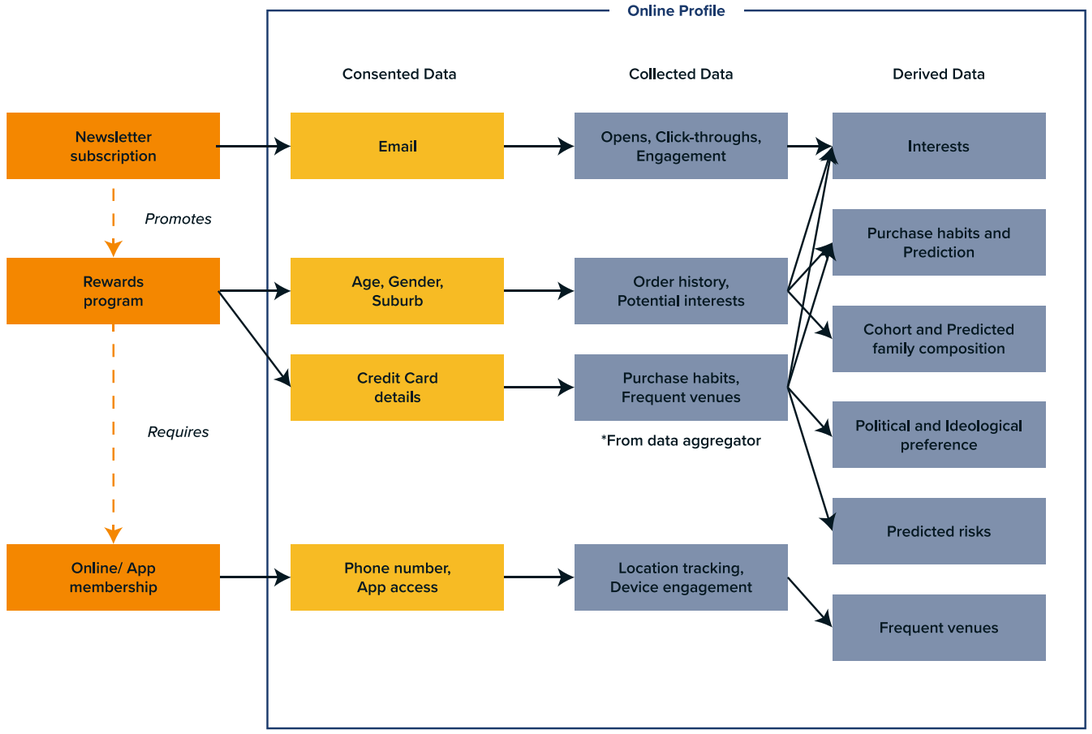


Figure 3. By consenting to provide access to a small amount of personal data, a rich online profile of the user can be derived.

Why are Deceptive Patterns used?

Deceptive patterns are commonly used because of their efficacy in influencing consumer behaviours. Many of these influence tactics might not be purposeful or deceitful and designers may unknowingly implement a deceptive pattern with good intention. Unfortunately, in many other cases, deceptive patterns are being deliberately implemented—this report focuses on these cases. It is clear that deceptive patterns promote conversion. This can range from turning a website visitor into an email list subscriber, a subscriber into an account holder, a free account into a paid account, a free trial user into a monthly subscriber, or a user providing minimal data into one providing maximal data for “improvements", and so on. These conversions can lead to more profit for commercial entities, both directly in selling more goods and services to that consumer, but also then the on-sale of that information to other commercial entities (i.e. data brokers).

The motivation of commercial entities is to achieve their business goals, which may clash with the goals of the consumer. One such motivation is most often a benign goal, such as the improvement of services. Many online services will measure their productivity and profitability through metrics such as daily visits and usage time. These metrics are often set by executives and designers aim to maximise the service’s performance against these measures. The designers implement new features and user interfaces and can test their performance against the measure with A/B testing[4].

Incentivised by producing improved metrics, designers might implement deceptive patterns, improving their performance at the cost of consumer data safety or autonomy. Secondly, market pressures and competition also play a role in the use of deceptive patterns. Commercial entities may employ deceptive patterns to boost their service’s performance to secure investments, create a larger audience, and display boosted click-through rates and other key performance indicators. Lastly, the use of deceptive patterns can be attributed to incomplete regulatory protection. Simply, they are used because they can be. If there is profit to gain and open avenues, commercial entities may take those avenues.

It is important to note that the line between what is considered a benevolent nudge and a deceptive pattern is not clearly defined. To better distinguish this line, three terms can be utilised: bright pattern, grey pattern, and deceptive pattern. A bright pattern is a nudge that is transparent in its motive and mechanism[5]. It provides the user with the autonomy to choose whether or not they accept the nudge and its influence, and it encourages positive behaviour. As an example, a bright pattern named 'simple consent' is where a website consent dialog provides a clear explanation of how the data will be used and gives the user easy methods for opting in and out. It is still not this simple, unfortunately.

While adhering to the golden rule of nudging in ways that promote the most help and the least harm is a good start, people need and want nudges in different ways [**118**]. Thaler and Sunstein suggest that nudges are great for choices that require memory, those that are difficult, and for when connection between the choice and the resulting experience are unclear. From a commercial entity’s perspective, they could offer a bright pattern with the intent of giving a positive nudge, but many factors relating to individual needs and preferences, and market influences can move these nudges out of the bright pattern territory.

A deceptive pattern, as defined earlier, is the opposite of a bright pattern. It is opaque in purpose and mechanism, strips a user of autonomy, and promotes a negative behaviour as it does not serve the nudge recipient’s best interest. Many examples of deceptive patterns are presented in **Part II** of this report and in the IVE deceptive pattern typology (see **Appendix**).

A grey pattern is one that straddles the line between bright and deceptive, featuring some components of both [**60**]. An example given by Potts and Mahnke [**106**] is Twitter’s throttling of post rate for users under investigation for breaches of terms and conditions. This throttle exhibits deceptive characteristics as it alters the system's operation for the user, while maintaining the appearance of normalcy.

On the other hand, it is considered 'bright' as it does notify the user that their posts have limited visibility (but not that the user is throttled), which is not a common practice among other platforms. Therefore, this example qualifies as a grey pattern. Grey patterns can either be transparent or hidden, aim to improve the service experience, and promote an ongoing behaviour.

As we have discussed, there is a temptation for commercial entities to prioritise deceptive over bright and grey patterns if they are driven by profit rather than consumer welfare.

4 In A/B testing, researchers present distinct versions of a product–perhaps a website or mobile application–to users in order to determine which one performs better. The label “A” corresponds to the original design, while “B” represents the variation of that design.

5 A website dedicated to promoting bright patterns and sharing examples of them is available at https://brightpatterns.org/bright-patterns-collection

Why should we care?

There are a significant number of identified deceptive patterns, and more are constantly added[6]. The different patterns deceive in various ways, some by hiding important information and others by pulling at emotions. The tactics employed by these patterns range from benign to highly manipulative. This range presents difficulty for understanding the breadth of deceptive patterns in their entirety. Additionally, the landscape of these patterns is constantly changing as new technology leads to new opportunities for data collection. With this constant evolution, it is not possible to know every individual deceptive pattern. It is crucial, however, to understand how they work and the reasons why we should care about the damage they can do to the consumer and commercial entity.

If we care about the protection of a consumer's online privacy and autonomy, we need to understand that nobody wants to be tricked, frustrated, or misled when using online services. Even if most deceptive patterns appear benign, or not something the consumer is actively aware of or concerned about, the deceptive tactics of commercial entities can lead to consequences for consumers' personal, sensitive data that should be protected on the consumers' behalf.

As an example that happens all too often, web services with poor security can be breached and the consumers' data can be combined with other leaked data, and used for targeted scams or fraud or even complete identity theft. It is conceivable that unnecessary data could be obtained by the online service via deceptive patterns without the consumer's awareness. We should try to both educate consumers about deceptive patterns and their effect on data capture, and act to protect the unaware public.

Considering the perspective of the commercial entity, the use of deceptive patterns may have a short-term benefit to the company in terms of increased profit.

Over the longer term, however, as consumer and media awareness of deceptive pattern usage is exposed, this erodes trust and transparency in the commercial entities that engage in these practices. Some websites offer a name-and-shame of commercial entities that use deceptive patterns [**152**], and recent legal challenges have successfully won millions of dollars in damages against Amazon's deceptive pattern usage [**26**, **34**]. Researchers [**9**] have found that a service's frequency of deceptive pattern use is correlated with a user's level of frustration. If we care about protecting the fairness, competitiveness, and trustworthiness of business, then deceptive patterns are not in their long-term interest and this should be shown.

Finally, that deceptive patterns are online in nature means that their manipulative abilities can scale up to a level of consumer access previously unattainable by the types of tactics from our department store example. With very little effort commercial entities can deploy deceptive patterns to their wide user base, and rapidly modify them for better efficacy in response to their testing. This unprecedented scale amplifies the other discussed concerns. With an understanding of why we need to focus our attention on deceptive patterns, we now need to understand how they work.

6 https://darkpatterns.uxp2.com/patterns/, https://www.deceptive.design/hall-of-shame

The AI Multiplier

Regular deceptive patterns are used to manipulate the user experience and influence user behaviours. For effective manipulation, the system must possess knowledge about how the user will perceive and respond to the user interface. Traditionally, these manipulations are enacted on a group scale, with the designer and system attempting a best-fit approach for all the service's users.

The introduction of AI has dramatically changed this landscape. AI-based deceptive patterns are highly capable of dynamically adapting to individual user's preferences. These manipulative approaches are made possible by two main factors: data and machine learning algorithms. AI algorithms enable mass data harvesting and aggregation. They also have the capability to analyse high-volume, high-dimensional data, derive insights on users, and build comprehensive user profiles. These capabilities are further enhanced by the system's ability to consume other sensor data, particularly from phones. This includes face recognition systems, voice recognition applications, and emotion detection algorithms.

Generative AI, such as ChatGPT, can also power deceptive business practices. Fake reviews and false ratings are pervasive, providing deceptive information on product quality to users and disrupting users' decision-making processes. Generative AI is capable of generating highly believable and persuasive fake reviews.

Dynamic pricing refers to the practice of optimising the prices of services or products based on various market factors, including demand, supply, customer demographics, and competitors' pricing. Numerous cutting-edge technologies have been adopted in dynamic pricing, and it has been demonstrated that AI technologies significantly improve business profit. As if deceptive patterns were not concerning enough, the increase in their effectiveness with AI considerably multiplies this concern.

What can our psychology tell us?

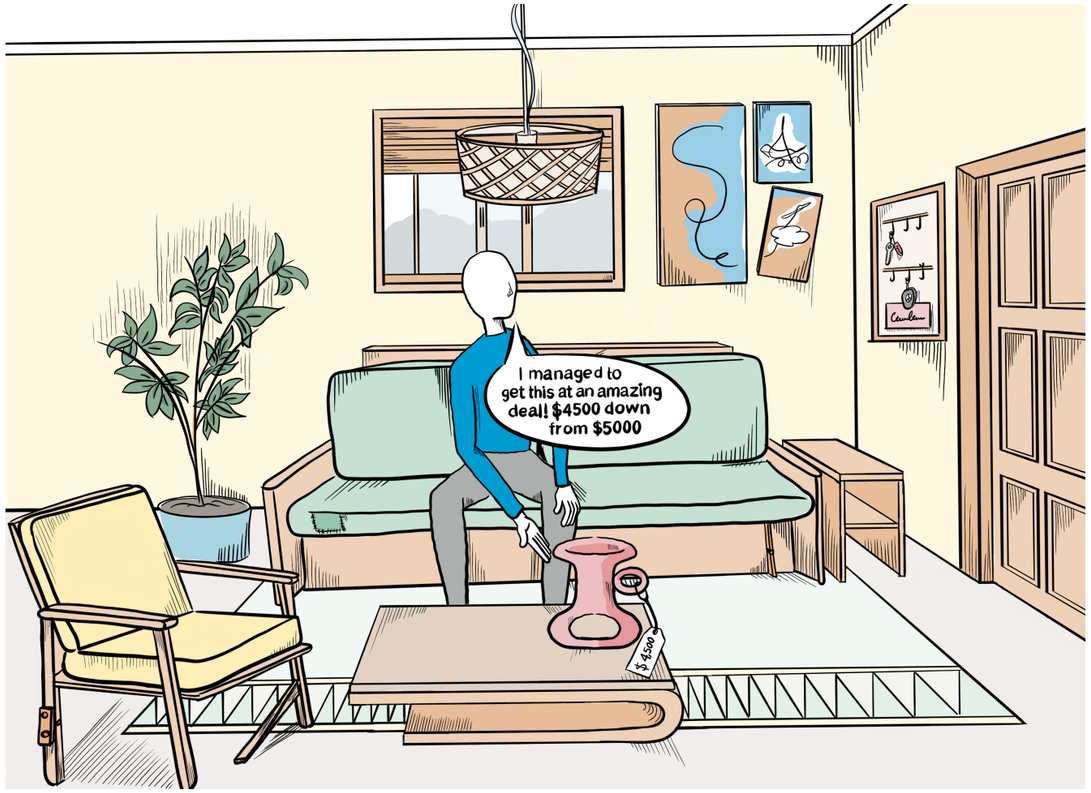
We all rely on mental shortcuts, known in the psychological literature as judgement heuristics [114], to function in a world filled with an overwhelming amount of information. Our mental shortcuts are available to us as part of an "intuitive, rapid, and automatic system" [114] that helps us to reduce the cognitive load that comes with calculating probabilities and predicting values from this huge set of data. Mental shortcuts, therefore, make our judgments much simpler to reach. While this reduction in cognitive load and brain processing time is beneficial, the shortcuts can lead to what are known as cognitive biases or fallacies in our reasoning that can be used against us.

As mentioned, we need mental shortcuts in order to comprehend the vast amount of information available to us at all times. The problem emerges when knowledge of the shortcuts leads to identification of the ways in which they can be exploited to suggest, manipulate, or deceive a person into making a particular decision without their awareness. Marketers, sales people, website designers, and many others can use cognitive biases to deploy tactics that influence us in ways that serve their objectives all without us noticing.

Concerningly, even when we are aware of the tactics and our own biases, their effects are so strong that we might not even be able to resist when they happen to us [38]. These cognitive biases are the psychological mechanisms that are predominantly targeted for manipulation by deceptive patterns. The following sections will illustrate some of the many cognitive biases that are most relevant to this report.

As a whole, researchers across many fields, including psychology and business, have identified many cognitive biases [2]. Not all are related to deceptive patterns, but the following can help us understand how deceptive patterns operate and which vulnerabilities they exploit in order to be effective in their deceptive and manipulative goals. We present the cognitive bias here, and later link them to deceptive patterns.

To help us understand how our cognitive biases impact our everyday decisions, the illustrations in this section follow the daily decisions made by our protagonist, "Casey". Follow the journey through their house renovation and see how the decisions have been shaped by underlying psychology.



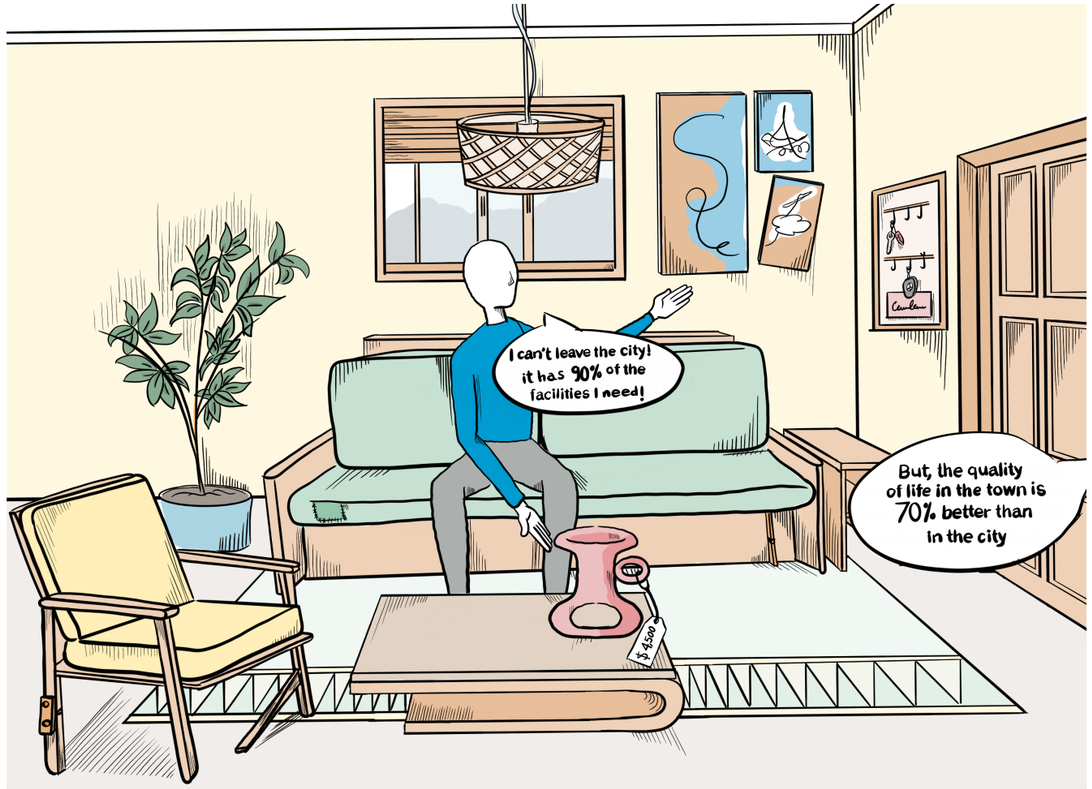
Anchoring

"When making a decision, the first presented value has an undue influence on the final decision"

The anchoring bias leads us to use the first presented value as a referent for our decision [38]. For example, consider our protagonist Casey looking to buy a car. When assessing an offer price for a car on a lot, the price on the windscreen is the first value Casey sees. Casey sets, unconsciously, that initial price as their anchor. When negotiating a sale price with the salesperson, getting $1000 off that initial price seems like a great deal to Casey given the initial valuation. The truth, however, is that the initial price may have been so grossly overvalued that even the reduced price was overpaying. The anchor has the power to reshape our conception of the car's true value.

Figure 4. Casey is impressed at the ability to negotiate the price from the salesperson's first offer. In reality, Casey was anchored to the first offer, meaning any lesser value seems great by comparison.

Experiments in psychology have shown that this anchoring effect even resists expertise and prior knowledge. As an example of a common experimental method, participants might be asked to guess the average June temperature in Germany, a country and climate with which they have no familiarity. The participants would be presented an initial temperature and then asked to guess the true answer. Participants who are presented with a high anchor will generally guess higher than those with a low, or differently to those who have no anchor [22]. In fact, even when participants are warned that the anchor is not indicative of a true answer, participants are drawn to the anchor, guessing a value closer to the anchor than those without [122].



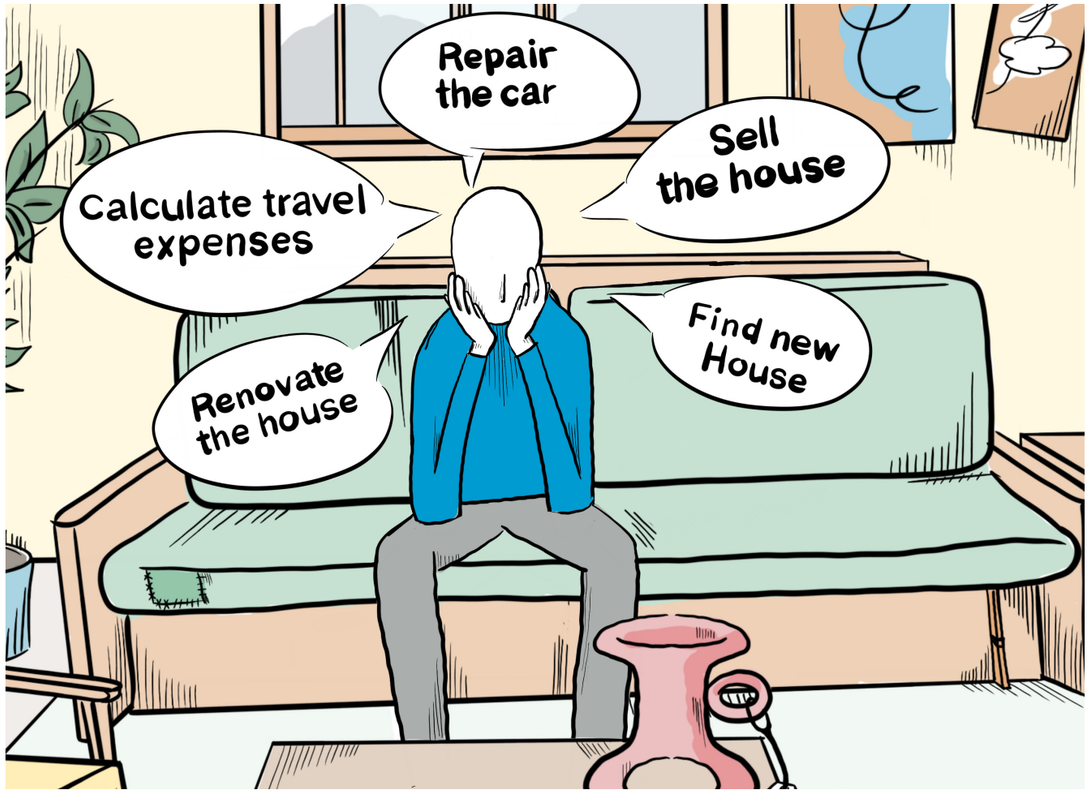
Framing

"How information is delivered to us can influence how we decide to act"

Formally, the framing effect consists of four components: problem definition, causal analysis, moral judgement, and remedy promotion [35]. These four components work together to encourage us to think, feel, and decide in a particular way. Basically, the same information can be presented in different ways.

As an example, our protagonist Casey is comparing two energy plans for their house. Both plans have the exact cost. Energy company number one advertises that they will save $100 on Casey's current energy bill. Energy company number two advertises that Casey will lose $100 by not choosing them. It is the same result from each, but advertising a gain is generally more effective than advertising a loss [123].

Figure 5. The choice of how information is presented can influence important decisions. Casey is presenting information that indicates that they should not move out of the city, but someone could just as easily present different information that suggests moving to a small town.



Status quo

"We tend to prefer our existing states of being over changes."

Quite rationally, the status quo bias can serve us well. Changes can be costly both in terms of money and time, and also can require significant effort. Sometimes, a good enough solution is worth maintaining in order to avoid those costs [**33**]. More often, however, we have strong feelings toward loss aversion, regret avoidance, preference stability, and cost that keep us in an undesirable position and make it hard to seek genuinely beneficial change.

An example of this bias in effect in politics is the incumbency advantage, where the politician who is currently elected is more likely to win again than their challenger. From business, an example is when Coca Cola created "New Coke," which people preferred when trying in a blind taste, but low sales led to the discontinuation of the product, likely due to regret avoidance [**153**].

Figure 6. When contemplating the choice of moving home, Casey imagines all the hurdles that this would present. Even if the decision is the right one, the status quo bias can make it difficult.

Priming

"Being presented with leading values can influence us to come to a particular conclusion."

Similar to framing, the priming effect centres around prior exposure to information influencing a decision. In experimental psychology, this effect is often tested by presenting participants with a series of words or images that have related traits.

For example, all presented stimuli might be fruits. Following the presentation of these primes, the participants may be asked to guess the next in the sequence. The participant is more likely to make a choice based on trait-similarity to the primes than something irrelevant [**30**]. A common exploitation of this effect is exemplified by YouTube advertising. The strategies of many companies when advertising on YouTube is to flood the space with ads for a particular brand; for a virtual private network software brand, for example. If viewers decide they later want a virtual private network, they are already primed to recognise and trust the brand they have seen so many times.

A less overt example is using colour in subtle ways in a website design to prime the user to get used to the colour and therefore trust the company's logo that features the same colour. In Australia, political parties have previously admitted to utilising the purple colour associated with the Australian Electoral Commission to imply a level of authorisation from them when targeting non-English speakers with voting material [**103**].

Figure 7. Casey mistakenly put soap powder in the kitchen pantry. The priming bias meant that the obscured text surely meant soup when coupled with the other items. If it had been in the laundry cupboard, this mistake would not have been made.

Figure 8. Casey purchased this water filter because it was popular, but that does not mean that it is actually the best.

Bandwagon

"We tend to agree with the viewpoint of the majority, even if it disagrees with our own."

Related to another cognitive bias, named group think [130], this bias stems from a variety of underlying psychological traits, such as the desire for conformity [91]. Brands use this effect by claiming that a particular product of theirs is popular and therefore an individual should buy it.

The effect, truthfully or not, implies that the popularity is a reflection of the product's usefulness or quality, or that the purchaser might gain some prestige or social credit by being part of the crowd [10].

Other examples can include political elections where people are more likely to vote for the candidate they perceive is winning, and social media groups such as anti-vaccine movements where the desire to be in this in-group resulted in unnecessary disease outbreak.

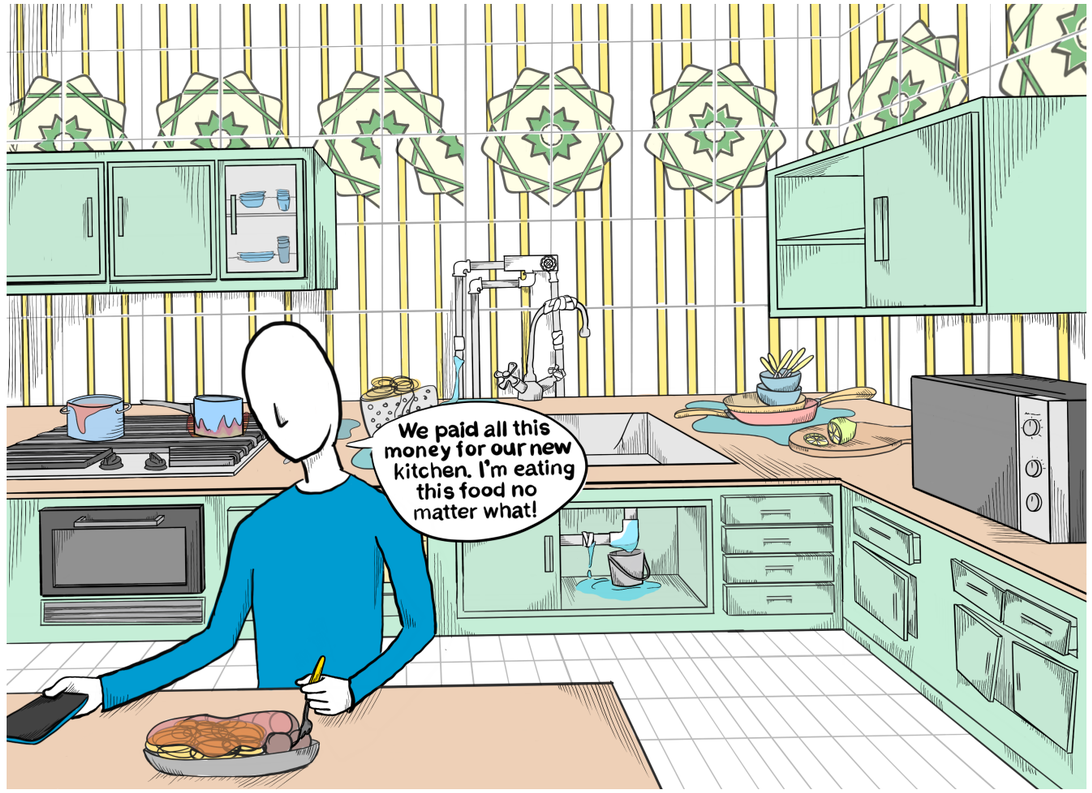


Figure 9. Despite the water filter not actually doing a very good job of providing clean water, Casey stubbornly refuses to change it, instead making the quality of all future meals suffer.

Sunk Cost

"We tend to continue to invest in an endeavour for which a prior investment of time, effort, or money has been made."

The sunk cost effect is similar to the status quo effect in that the avoidance of unnecessary costs can be rational. The irrationality, however, is that the prior investment should not influence the decision to continue investing [6]. Further, it does not necessarily follow that more investment will complete or resolve the endeavour.

An Australian example is criticism of Melbourne's decision to cancel hosting the 2026 Commonwealth Games. The criticism levelled at the government focuses on the investment in the Games before being cancelled and the remaining contractual obligations that need to be paid after cancelling. If the future benefit of the project does not outweigh the sunk cost, then the reasoning based on the sunk cost is mistaken.

Part II:
Types of Deceptive Patterns

Types of Deceptive Patterns

In **Part I** we introduced some of the predominant cognitive biases that we are susceptible to as humans, as a tool for understanding how deceptive patterns can influence our behaviour. The growth in the number of identified deceptive patterns[7] has increased to the point where it has become increasingly difficult to be both aware of and protect against these emerging patterns. It is even more challenging to defend against future and emerging deceptive patterns. Furthermore, due to the constantly changing landscape it is hard to gauge the relative risk of each pattern, especially in the context of data standards.

Part II categorises the current deceptive patterns for the sake of clarity and explanation. Introducing categories allows us to provide a way of classifying new deceptive patterns that may emerge and suggest possible approaches that are not confined to naming specific deceptive patterns. It is less likely that a new deceptive pattern will require a completely new category, than it is that new deceptive patterns will emerge within existing categories. Thus, we identify deceptive patterns from the literature, place them into a model with strongly defined categories, and ultimately produce a list of deceptive patterns with a consistent definition style.

In this section, we will describe our method, the creation of our typology[8] (the IVE deceptive pattern typology), the selection of a categorisation model, and the discussion of the categories with a selection of deceptive patterns within those categories. We aim for this body of work to be a reference for identifying and understanding deceptive patterns.

Using the IVE Deceptive Pattern Typology

The IVE deceptive pattern typology was created with a number of uses in mind. For regulators, it serves two main purposes:

1. A means by which to identify areas of concern, as represented by the four quadrants of the model shown in **Figure 10**; and
2. A comprehensive list of known deceptive patterns, designed to be referenced when identifying deceptive patterns.

For user experience designers, the typology has the same purposes as for regulators, but it gives many examples of what not to do when considering how to nudge their users, if nudges are even necessary.

For the general public, the typology serves to increase awareness so people can defend against the negative influence that deceptive patterns exert.

7 For continuously updated lists, see https://deceptive.design and https://darkpatterns.uxp2.com

8 A typology is a classification based on types or categories. In this report, the IVE deceptive pattern typology categorises deceptive patterns.

Method

Deceptive pattern researchers have identified a multitude of deceptive patterns and formed several typologies[9]. With a goal of understanding the current landscape of deceptive patterns in mind, the first step was to identify all existing typologies as part of a larger systematic literature review (detailed in Part III). Then, we conducted a qualitative review of all these typologies, which allowed us to extract the deceptive patterns present in each typology into a larger corpus.

Due to the many typologies having different styles of identifying and defining each deceptive pattern, our final step involved rephrasing the definition of each deceptive pattern and placing them into a specific category for easier understanding and categorisation.

Creating the IVE Deceptive Pattern Typology

The first step was to compile a list of deceptive patterns from existing review publications. Through a combination of a systematic literature review (detailed in **Part III**) and a directed search, we identified 19 source typologies for the IVE deceptive pattern typology (see **Appendix**).

From those typologies, we extracted a total of 157 unique deceptive patterns (see Appendix). We then rephrased the original author's definition into a consistent form, beginning with the phrase "the user". This stems from our belief that since the deception and manipulation is targeted toward the user, and since they bear the brunt of the harm caused by deceptive patterns, the primary focus should be on the user. This form also reflects a "user story" description that is applied in user experience design, enabling discussion with designers and implementers of front end systems.

As previously mentioned, this list of 157 deceptive patterns is subject to expansion as new tactics and manipulative opportunities arise. In addition, such a long list is too cumbersome to be of use. As such, we sought a categorisation model that would work for the purposes of this report.

Choosing a Categorisation Model

As previously mentioned, it is important to understand the deceptive patterns from a model perspective. A model allows an understanding of how deceptive patterns impact consumers at a deeper level than a list of individual patterns would. There are many models that have been created by deceptive pattern researchers.

We were specifically searching for one that places its primary focus on the impact to users, rather than the design features of the patterns. When considering a source model for the IVE typology, we selected the Leiser model [**72**], designed with Unfair Commercial Practices Directive (UCPD) [**146**] in mind. The UCPD is European Union legislation that aims to protect consumers' economic interests from commercial entities that may otherwise violate them in service of their own interests.

Leiser's categories take this goal of protecting the consumer and apply it to deceptive patterns, creating a model dividing deceptive patterns into two categorisation levels, information asymmetry and free choice repression. Together, these two broad categories represent how consumer autonomy and decision-making can be

9 These are often referred to as taxonomies in the literature. As collecting and defining dark patterns is a manual, qualitative process, they are more correctly referred to in this report as typologies.

covertly manipulated. The benefit of the Leiser model over others is that it focuses on how the deceptive patterns impact the user, rather than group them together based on superficial characteristics (such as how they obstruct, or how they are styled). This focus on the user leads to a model with less overlap between categories, and also more strongly aligns with the aims of this report.

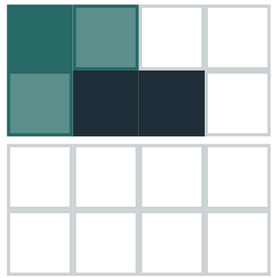
Figure 10. Leiser deceptive pattern categorisation model [**72**]. At level 1, patterns are split into either information asymmetry or free choice repression. The four level 2 categories are shown in the corners, with the eight level 3 categories attached.

Deceptive Patterns

Information Asymmetry

The first of Leiser's level-one categories, *information asymmetry*, encompasses manipulation tactics that prey on consumers' lack of available information.

The asymmetry is that the commercial entity controls the information flow to the consumer, meaning that the information can be changed, hidden or delivered in such a way that the consumer does not have the full picture. Additionally, the commercial entity is able to ingest the consumer's data, sometimes without the consumer's knowledge. The information asymmetry level-one category is further divided into level-two categories of *active misleading actions* and *passive misleading omissions*.



Active Misleading Actions

Active misleading actions include those where the commercial entity shows information to the consumer that actively (that is, deliberately) deceives or manipulates. It is further divided into two level-three categories, misleading information and misleading presentation.

Misleading Information

Misleading information comprises actions that "provide false, confounding, deceiving, or exaggerated information actively to mislead consumers" [**72**]. Examples of deceptive patterns in this category include **Hidden Legalese Stipulations**, **Just Between You and Us**, and **Loss-gain Framing**.

Hidden Legalese Stipulations, as shown in **Figure 11**, displays an apparently normal terms and conditions consent dialog. If the user does not carefully read the legal language, they may not notice that the web developer has incorporated some strange requests in the text, including that the user surrenders their "immortal soul" to the commercial entity. This is in fact a real example that the US company Gamestation displayed as an April Fool's joke [**150**], fooling all 7500 customers who made purchases that day. While this is a silly joke, it exemplifies the ease with which information that is not in the consumer's interest can be buried in complex legal language.

In the **Just Between You and Us** deceptive pattern (shown in **Figure 12**) the user is encouraged to provide extra information, with a stipulation that the information will not be visible to others but it is in their best interest to provide it to give a better overall experience. Social media services are known to employ this kind of deceptive pattern, with the proviso that it is only to help better connect with relevant people and marketing.

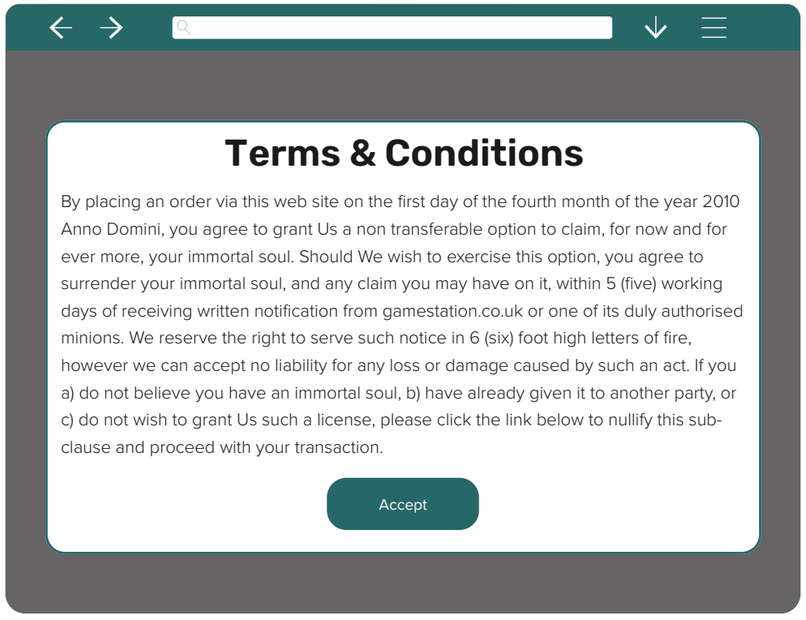


Figure 11. Hidden Legalese Stipulations

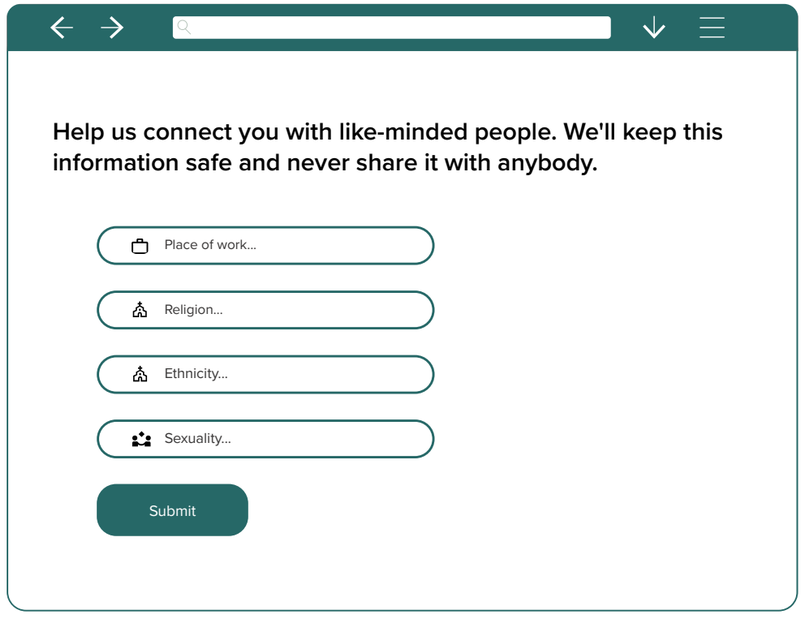


Figure 12. Just Between You and Us

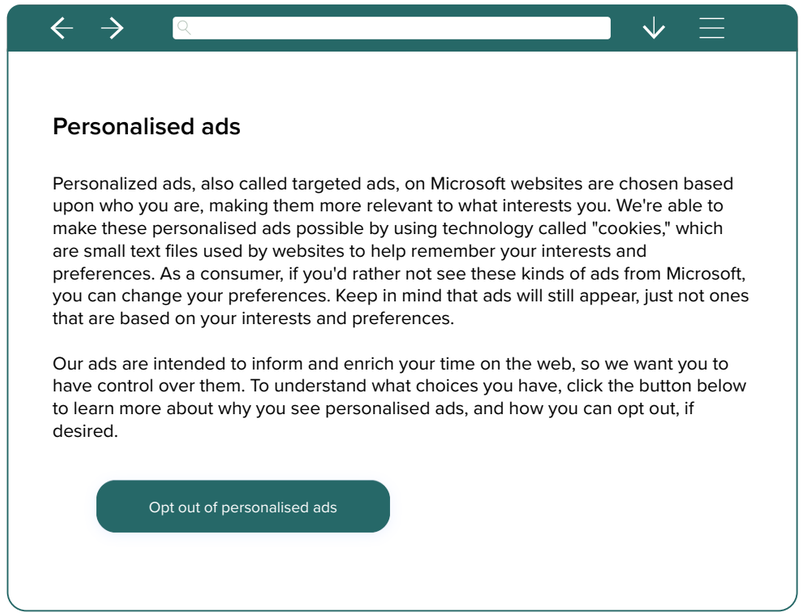


Figure 13. Loss-gain Framing

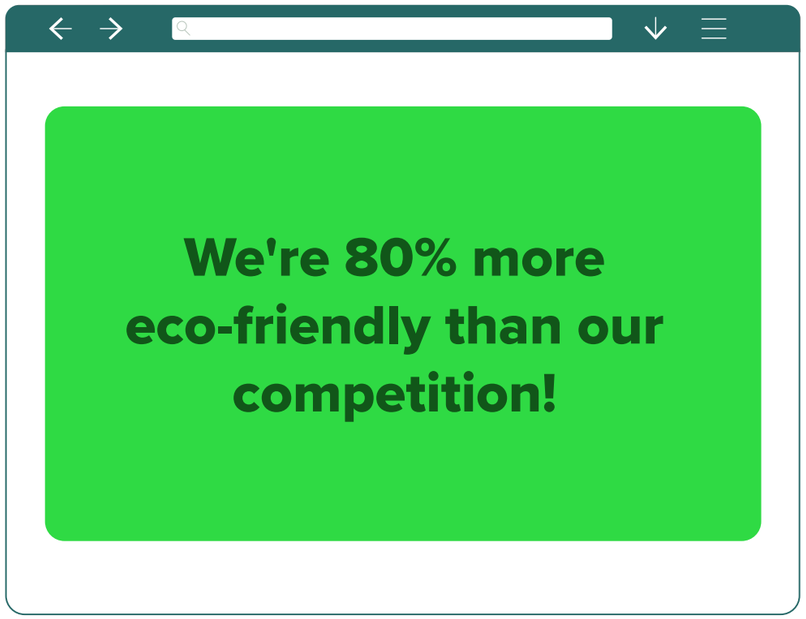


Figure 14. Colour

With clever use of language, the **Loss-gain Framing** pattern (shown in **Figure 13**) presents options in such a way that the commercial entity's preferred result is the obvious decision. A common example of this pattern appears in relation to personalised advertisements. On one of Microsoft's pages[10], they frame the option to select personalisation based on the ads being "intended to inform and enrich your time on the web" and that the ads "are chosen based upon who you are, making them more relevant to what interests you". This makes it seem like you will have a much worse experience if you opt out of personalisation. The actual benefit of doing this is restricting Microsoft's ability to share personal data with marketers.

10 https://about.ads.microsoft.com/en-us/resources/policies/personalized-ads

Misleading Presentation

A misleading presentation is where a commercial entity aims to “present information in a misleading manner” [**72**]. Some deceptive patterns that fall into this category include **Colour**, **Trick Question** and **Visual Interference**.

In **Figure 14**, an interface for a company that wants to promote their product to eco-minded people can use the colour green, which has been shown to be associated with being eco-friendly. Other examples include mainland Chinese customers strongly identifying with the colour red meaning something positive, where Western customers might have that identification with the colour green [**55**].

Commercial entities may use **Trick Question** to deceive consumers into selecting a particular option by way of confusion or accident. In **Figure 15**, an example from The Washington Post is shown where the user is presented with a phrase and a checkbox that make it appear to be asking if the user wants to receive a special edition. Other phrasing on the interface asks the user to “uncheck” the box if they do not want the post. This phrasing could lead consumers into thinking they were supposed to check the box to not receive.

By using **Visual Interference**, designers can hide or conceal information that they would prefer the consumer misses. In **Figure 16**, an example is shown of a Tesla app interface where the customer is able to purchase a software upgrade.

Hidden in concealed text is the information that the upgrade is non-refundable. It is plausible that some customers would expect a satisfaction guarantee and would refrain from purchasing if they noticed that disclaimer.

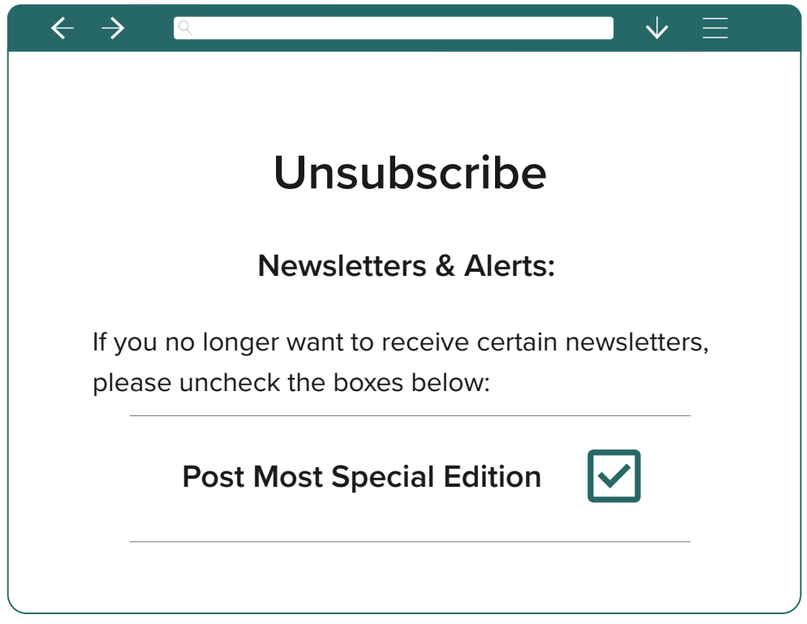


Figure 15. Trick Question

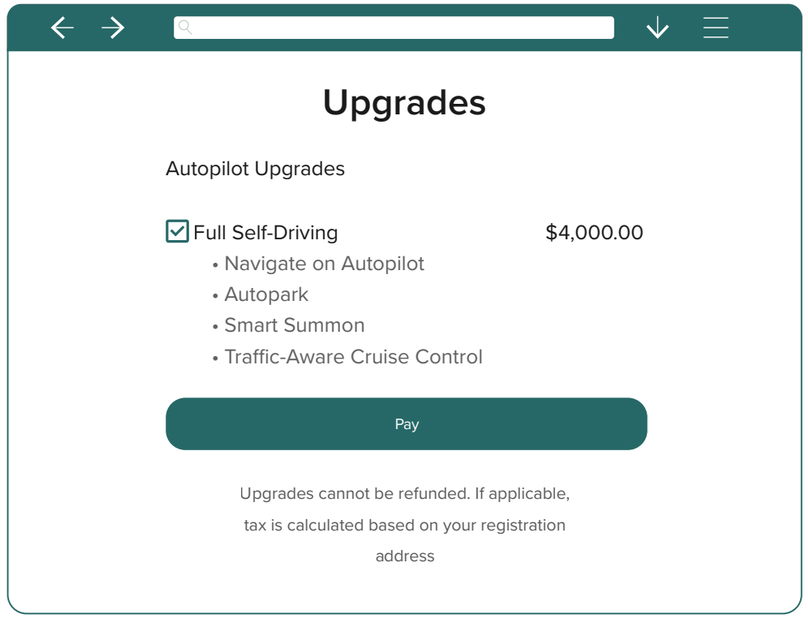
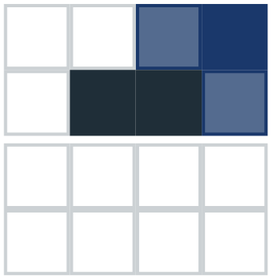


Figure 16. Visual Interference

Information Asymmetry



Passive Misleading Omissions

Unlike active misleading actions, passive misleading omissions deceive and manipulate consumers by hiding or delaying specific information that may not be in the commercial entity's interest for the consumer to see. Rather than being actively manipulated, the consumer is left unaware of the information that could, or perhaps should, be shown to them. This category is further divided into two level-three categories, hiding information and delaying provision.

Hiding Information

Hiding information involves when a commercial entity "fail[s] to provide or omit[s] necessary information" [**72**]. The category is represented by the patterns **Hidden Information** and **Immortal Accounts**.

In **Figure 17**, a common example of **Hidden Information** is shown where a service will hide important information by utilising an interface component, in this case a collapse field. The consumer is required to perform an action to reveal the hidden information, then uncheck the box in order to see and deselect the default option that may not be in their best interest. Another method of implementing the **Hidden Information** deceptive pattern is via complicated interface navigation.

As an example, in Apple's iOS 6, user's had the option to opt out of ad tracking. To do this, they had to navigate to settings, then general, then about, then advertising, and then toggle the ad tracking switch. The option was hidden in a different manner than the example in **Figure 15**, but the effect is the same.

The **Immortal Accounts** deceptive pattern relates to commercial entities collecting personal information from their service account holders, enabling account deletion, but not deleting the information along with the account.

Figure 18 shows an example of this where the Animal Crossing Community website openly acknowledges that unless the user takes specific actions to remove personal information before deleting their account, the data will be kept by the service. The drive to collect data is so strong that commercial entities can find value even in data relating to people who have deleted their account.

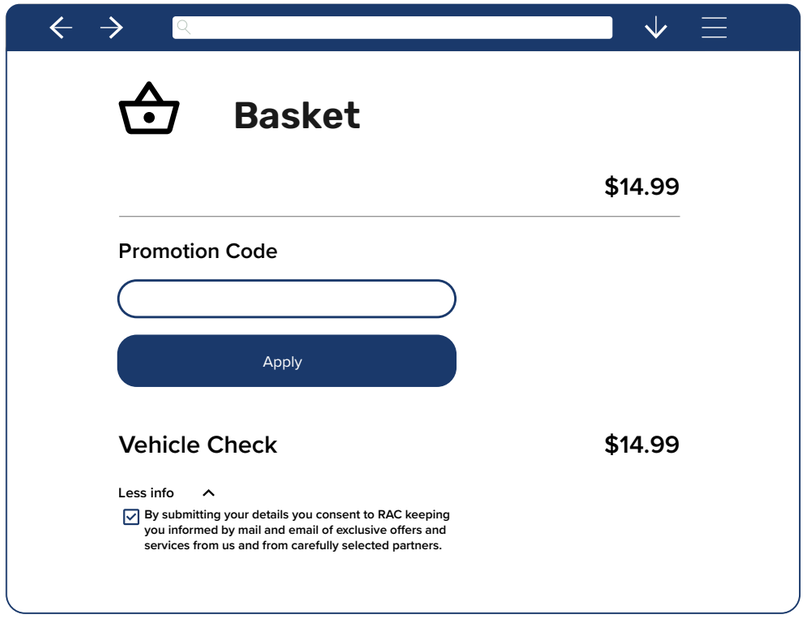


Figure 17. Hidden Information

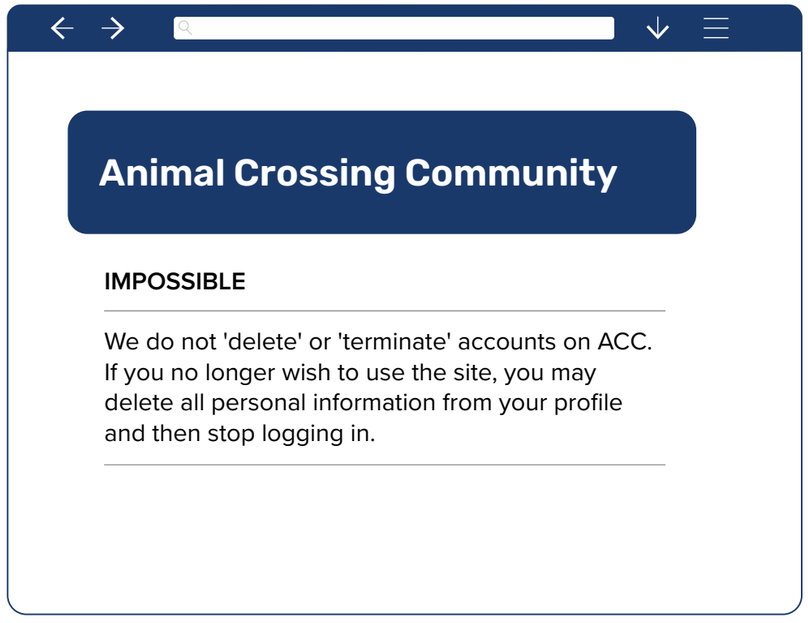


Figure 18. Immortal Accounts

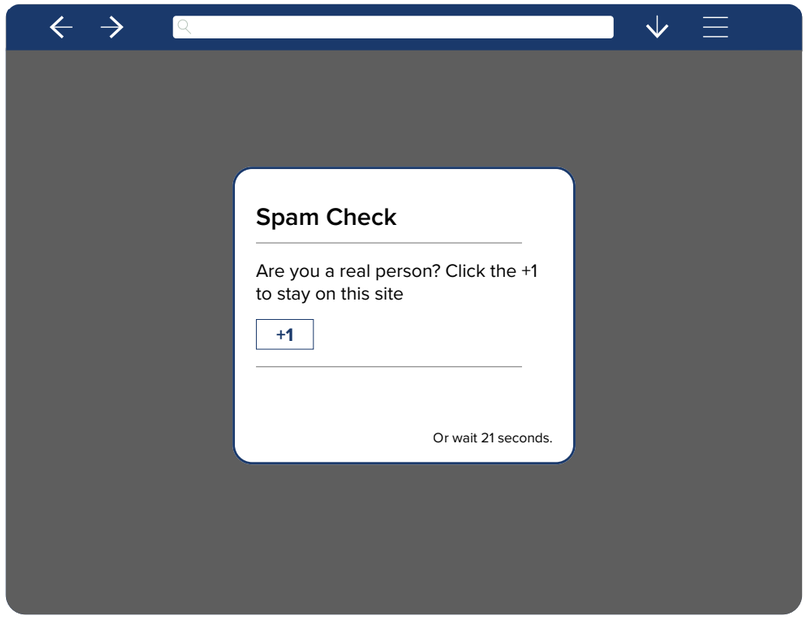


Figure 19. Delay User's Work Effort

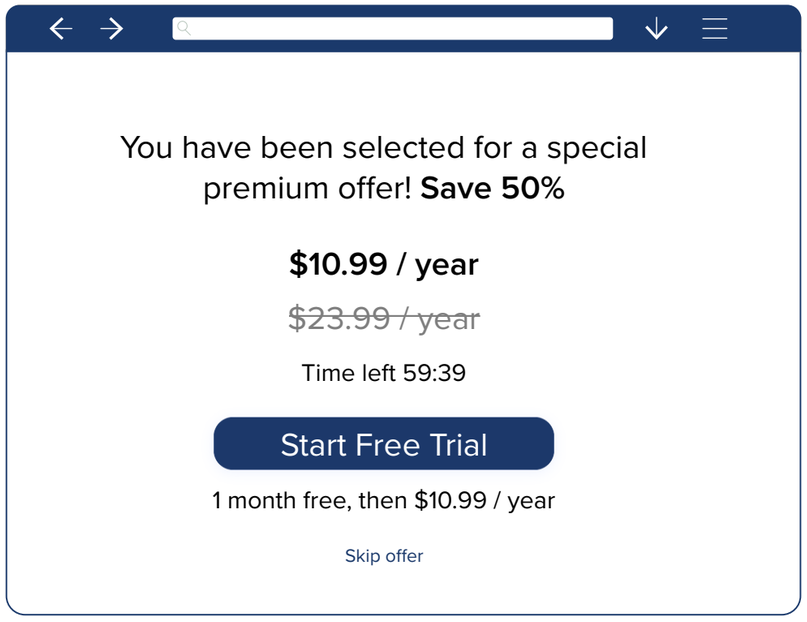


Figure 20. Hidden Costs

Delaying Provision

The final level-three category in information asymmetry is delaying provision, which is described as when the commercial entity aims to "delay the provision of information" [**72**]. Two examples of this pattern include **Delay User's Work Effort** and **Hidden Costs**.

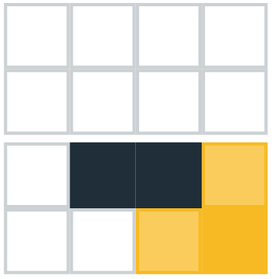
Figure 19 shows an example of the **Delay User's Work Effort** deceptive pattern. This pattern displays some user interface elements that interfere with regular interaction. The delaying component refers to making the user wait for a period of time (in this example 21 seconds), before regular interaction is resumed. If the user does not want to wait, they can perform an action that is in the service's interest, clicking the component, which most likely would redirect the user to an advertisement, other website, or download.

The **Hidden Costs** deceptive pattern lures the consumer into an attractive deal, only to later reveal that additional costs were hidden in the initial engagement. Pictured in **Figure 20** is an example where it looks like the user is being given a free trial, based on the button text. If the user accepts this offer, the service will automatically commence an annual subscription. If the user is not paying attention, they will be charged.

Free Choice Repression

The second of Leiser's two top-level categories, *free choice repression*, refers to the ways in which commercial entities can impose restrictions or barriers to performing actions that do not align with their interest. The resultant outcome for consumers is that their desired actions may be hindered or outright blocked, negatively impacting their autonomy. Commercial entities are incentivised to reduce their consumers' choices, as directing them to perform only actions that contribute to the commercial entity's profit margin is highly motivating.

The free choice repression level-one category is further divided into level-two categories of *undesirable imposition* and *undesirable restriction*.



Undesirable Imposition

A pattern in the level-two category of undesirable imposition has features where the commercial entity is forcing some action or workflow upon the consumer as antithetical to the consumer's desires. It is further divided into two level-three categories of pressure imposing and forced acceptance.

Pressure Imposing

In this category, deceptive patterns include tactics by commercial entities that "impose burdens or pressures on users" [**72**]. Two examples include **Confirmshaming** and **Safety Blackmail**.

Confirmshaming shaming employs emotive language to promote a feeling of guilt that prevents a consumer from making a choice that does not align with the commercial entity's goals. In **Figure 21**, the service is attempting to gain consent from the user for displaying notifications. The Confirmshaming deceptive pattern is used in the "no" response, which is unnecessarily emotive and hyperbolic.

The **Safety Blackmail** deceptive pattern takes advantage of consumer desires to be safe and secure online. Shown in **Figure 22** is an example of this desire being exploited. The service claims that the mobile phone number is being collected as a safety feature, akin to multi-factor authentication that consumers might be familiar with, but the terms and conditions state that the number will be passed onto third parties for advertising purposes.

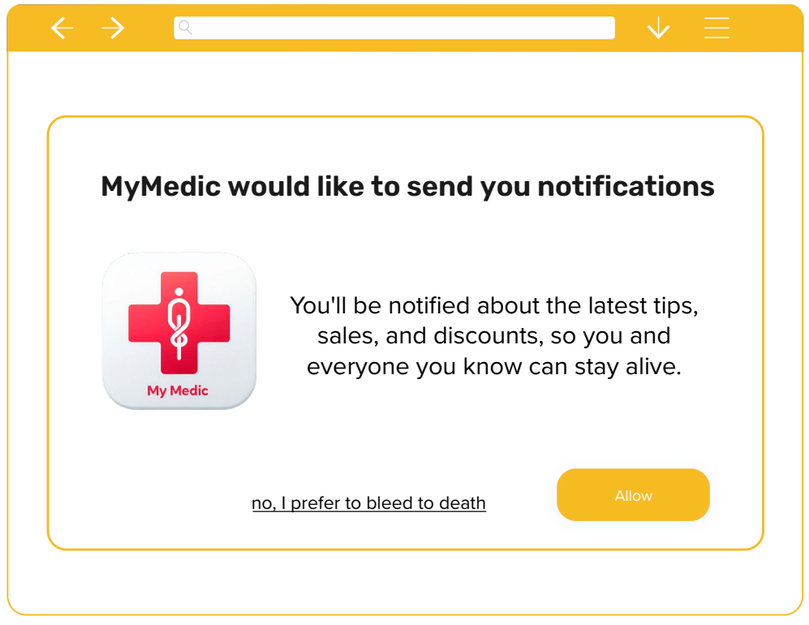


Figure 21. Confirmshaming

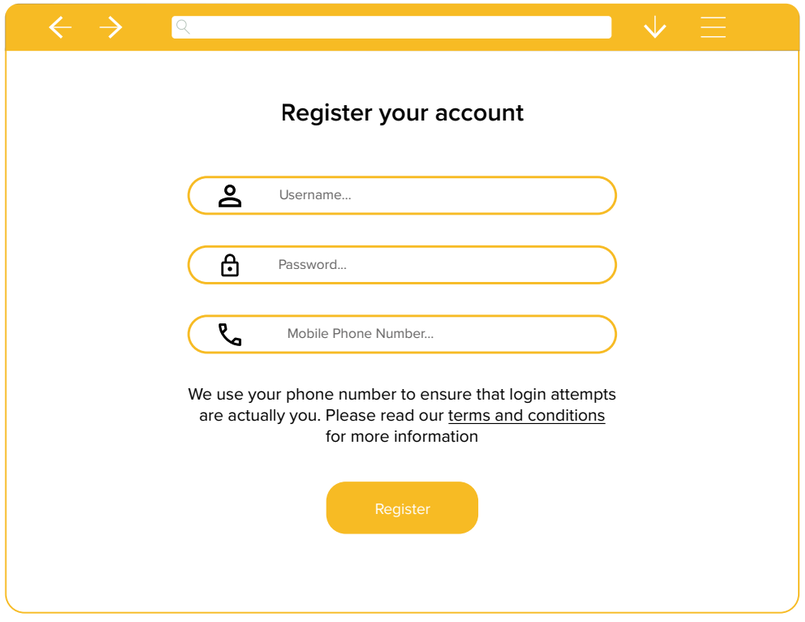


Figure 22. Safety Blackmail

Forced Acceptance

The second level-three category is forced acceptance, which involves patterns that "induce consumers to accept or retain an undesirable entity such as product sneaking into their shopping carts" [72]. **Bait and Switch**, **Forced Consent**, and **Illusion of Control** are all deceptive pattern examples that fit in this category.

In the **Bait and Switch** deceptive pattern, the consumer is led to believe that an action will have a particular result, but it instead causes some other, likely undesired result. The example in **Figure 23** presents a lot of text relating to the cancellation of a service, and the colour-filled button seems to indicate a confirmation or accept button. In fact, the cancel and back buttons have the same result; not proceeding with the cancellation.

In **Figure 24**, the **Forced Consent** deceptive pattern is used to force a user to agree to both the terms of service and cookie policy in order to use the service. Even if the user would prefer to amend the data the cookie policy is collecting, the service provides no way to do so, except not using the service entirely.

In a similar manner to the **Safety Blackmail** deceptive pattern, **Illusion of Control** takes advantage of a consumer's desire to protect their personal data. In **Figure 25**, the service shows a comforting message about how the consumer will have full control over their data and can "easily" change how the service uses their data. In reality, this ease is an illusion, as the process for finding and modifying the settings is tedious and difficult.

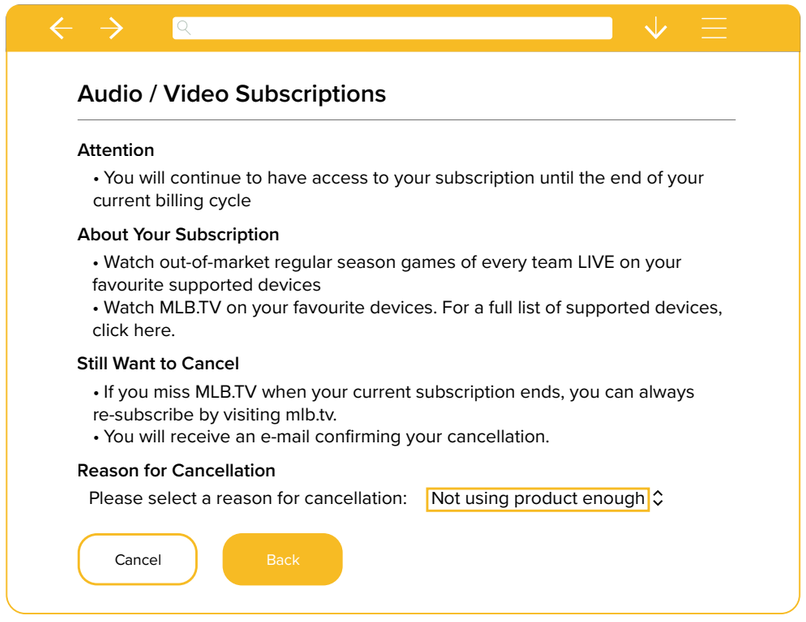


Figure 23. Bait and Switch

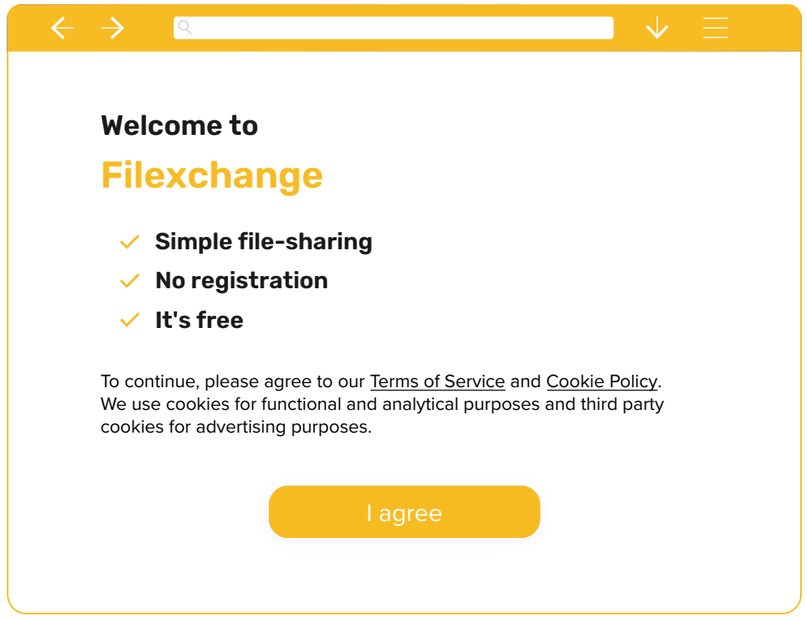


Figure 24. Forced Consent

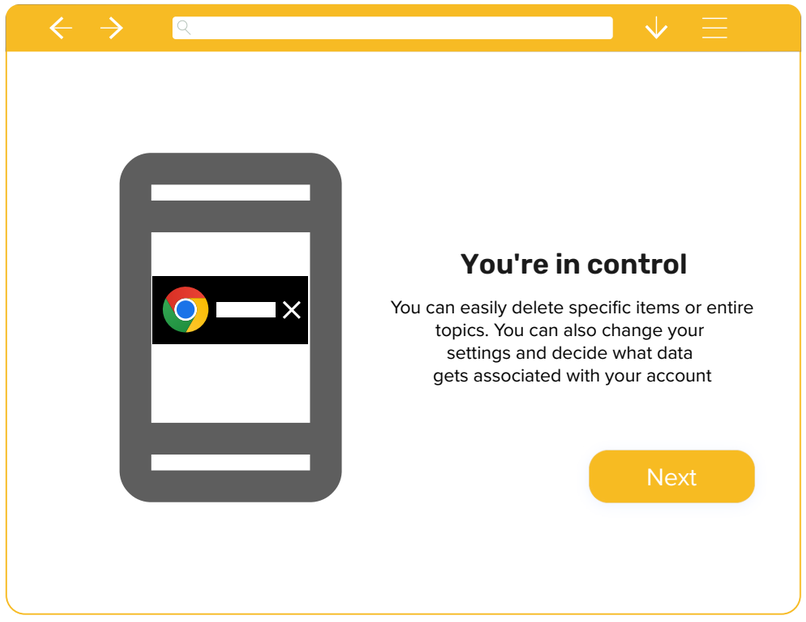
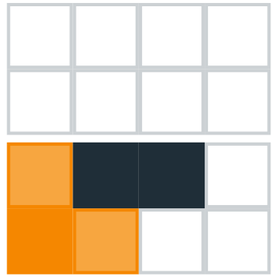


Figure 25. Illusion of Control

Free Choice Repression



Undesirable Restriction

Contrary to impositions, undesirable restrictions place unfair limitations or obstructions on the consumer's actions. Again, this category is split into two level-three categories, restricting specific users and restricting specific actions.

Restricting Specific Users

In restricting specific users, patterns aim to make "certain functionalities unavailable or challenging to use for specific groups of users" [**72**]. Two examples are **Nickling-and-diming**, and **Pressure to Receive Marketing**.

When using the **Nickling-and-diming** deceptive pattern, the commercial entity is trying to squeeze more money from the consumer than the consumer realises. **Figure 26** shows an email that a credit card company sent to the card holder. The text casually suggests that since the consumer has available funds, they should be spent. The commercial entity is hoping that by convincing the consumer to spend more, this opens up further opportunities for fees and interest charges.

Many commercial entities are keen to access consumer email addresses as this creates an easy avenue for marketing communications. The most nefarious of commercial entities will then sell these addresses to third parties who can use them for scam campaigns. Often, the **Pressure to Receive Marketing** deceptive pattern can be presented like in **Figure 27** where the consumer must provide an email address in order to use the service. It is not clear to the consumer why the email address is necessary, but the pressure to provide it exists.

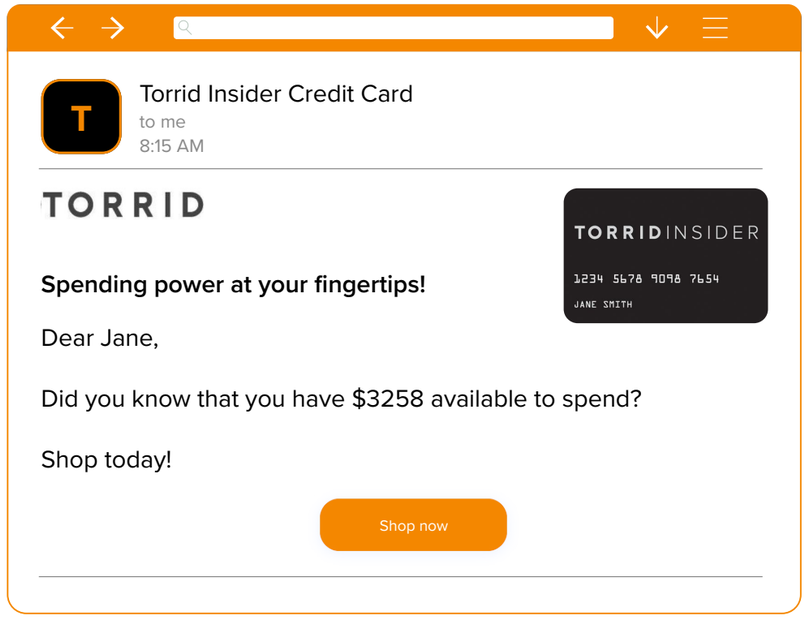


Figure 26. Nickling-and-diming

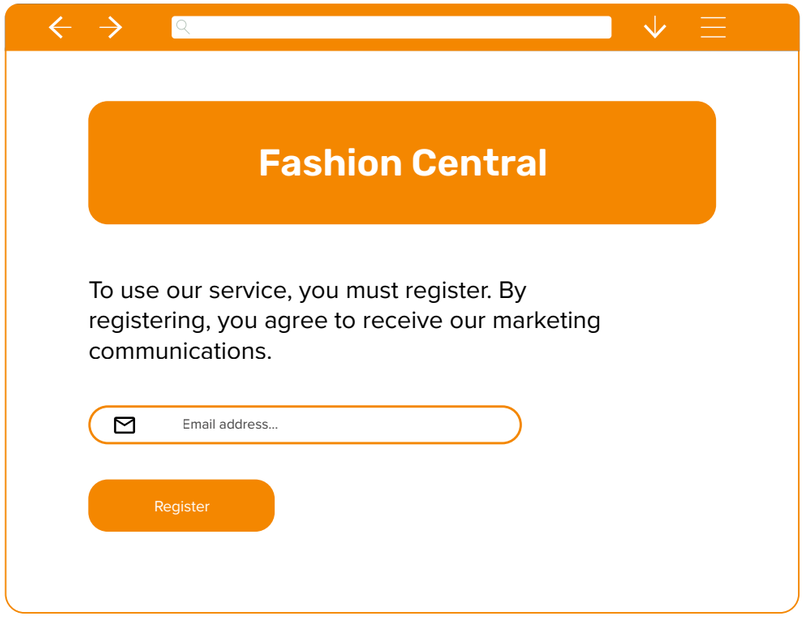


Figure 27. Pressure to Receive Marketing

Restricting Specific Actions

The final category, restricting specific actions, involves patterns “setting restrictions or obstacles on specific actions for all users. These include making access to the service or the options to unsubscribe more complicated than [it] needs to be” [**72**]. Examples include **Decision Uncertainty**, **Forced Explanation**, and **Roach Motel**.

Decision Uncertainty is represented in **Figure 28**, where a dialog is presented on the TikTok social media app. The background, still visible around the outside of the dialog is playing a video with sound. The dialog is presenting information pertaining to a choice the user should make about whether or not they want to receive personalised advertisements. The language used makes
it unclear about what the user should choose, but the easiest response is to select the “Accept and continue” option. This quick and easy response is coupled with the deliberate playing of the background video, encouraging the user to quickly dismiss the dialog by accepting and returning to the TikTok content.

Figure 29 shows how difficult deleting an account or unsubscribing from a service can be. Commercial entities are incentivised to keep users, so it is often not in their best interest to offer a convenient mechanism for users to cancel.

The **Forced Explanation** deceptive pattern presents a barrier to the cancellation process that many consumers will find very unattractive; communicating directly with a service staff member in order to make a request. The example shown involves a real-time chat service where a consumer asks a support staff member. In most cases, the staff member would be instructed to ask follow up questions and suggest alternatives to deletion that the consumer would have to rebuff to proceed. It is also quite common for this deceptive pattern to present as a requirement to email or phone the support staff.

The **Roach Motel** deceptive pattern humorously refers to the online maze that consumers are often forced to navigate in order to find the setting or process they require. As discussed, commercial entities are incentivised to make actions counter to their goals difficult or impossible to access. **Figure 30** shows an example of the process required to cancel a paid subscription to a service called G2A Shield.

To find the page, the user must navigate to “Settings,” then “Account,” then “G2A Shield,” already a flow that is unintuitive. Then, the page presents the consumer with a series of other deceptive patterns that the user must navigate to find the desired option presented as small text at the button “Disable subscription now.” There is also no guarantee that this option, once clicked, would not navigate the user to yet another screen asking for confirmation.

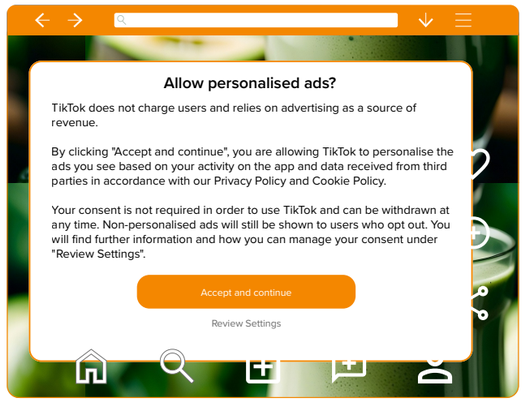


Figure 28. Decision Uncertainty

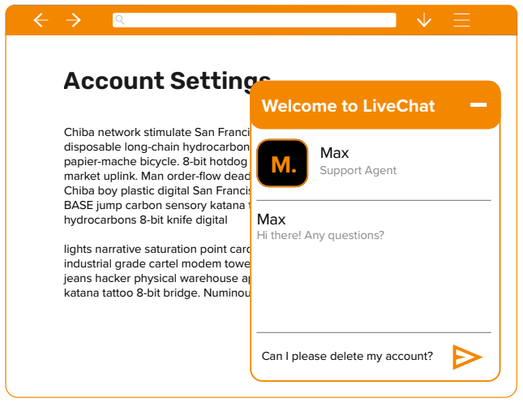


Figure 29. Forced Explanation

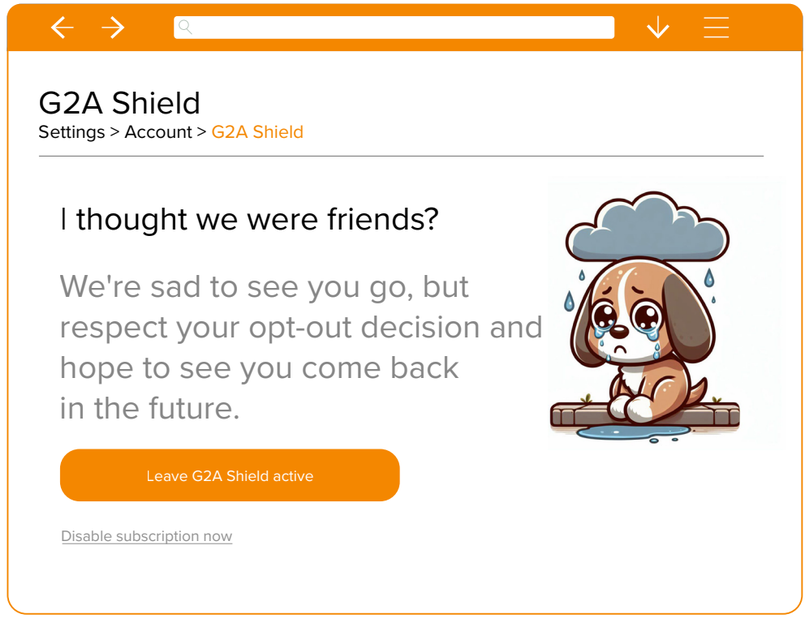


Figure 30. Roach Motel

Part III: The land-scape

The Landscape

This section addresses aim 3 (see Introduction), which is to explore the recently published academic work on deceptive patterns.

The goal is to identify the current concerns and focuses of the global research community. This involves understanding which issues are being prioritised and where the majority of research efforts are being directed. In order to build a comprehensive understanding, our objective was to capture the findings and solutions from many disciplines. These disciplines ranged from computer science to law, business, and psychology. This is important as it provides a well-rounded view of the research landscape.

Finally, this section distils the overall landscape into consumable research themes. These themes can guide readers and help identify emerging threats, emerging solutions, and pertinent gaps for future research. This final step is crucial in making the research accessible and useful to a wider audience.

Method

In order to address aim 3, it was insufficient to merely access a small portion of the most cited deceptive pattern research papers. It is important that we report a broad and complete picture of the state of current and future deceptive pattern threats and research to mitigate them. Due to these factors, we conducted a systematic literature review, which finds and assesses all research papers within a strict set of search and inclusion criteria (as opposed to conducting an incomplete search).

The first phase of the systematic literature review involved deciding which research databases to query. Deceptive patterns research is a broad field that spans many disciplines. This means that limiting our search to only computer science databases, for example, would miss much of the research conducted in other fields. In order to cover as broad a range of publications as possible, we decided to conduct our search using Google Scholar[11].

The next phase was to choose the query terms. To ensure that we included only references that were strictly related to deceptive patterns[12], we limited “dark patterns” to appear in the title of the article. Additionally, due to other prominent typologies already existing [14, 18, 29, 31, 46, 81], we decided to limit our search to include publications from 2020 until 2024[13], capturing all newer research.

Finally, references that were not peer-reviewed (this includes theses, patents, and reports) were excluded.

11 https://scholar.google.com

12 For the purposes of finding as many relevant matches as possible, we used the much more common “dark patterns” term.

13 Search conducted 2024/02/13

Our final Google Scholar[14] search string was:

```
"dark pattern" OR "dark patterns"
[title] from 2020 to 2024, no
citations, no patents
```

This query returned a total of 334 publications.

The first stage of reducing the number of publications for further analysis involved applying some inclusion criteria. In order for a publication to be included it must not:

- be non-peer reviewed, including prep-prints, undergraduate, master, and PhD theses, or on general university research databases;
- be in a language other than English;
- be a patent;
- be a military report; and
- contain “dark patterns” in the title but be otherwise off-topic.

After applying these criteria, 106 publications proceeded to the next phase.

In addition to the system literature review, a targeted search was also conducted to discover publications relating to AI’s impact on deceptive patterns, following the same inclusion criteria outlined above. Publications were sought on Google Scholar using combinations of “dark patterns” with "artificial intelligence", "data aggregation", "nudging", and "dynamic user interfaces".

Further publications were discovered from relevant and highly cited publications. These queries returned 107 publications.

14 The search was conducted using the software Publish or Perish

Findings

In order to better understand the research landscape, the 106 publications were qualitatively categorised according to the theme(s) they investigated. Individual publications could be assigned multiple themes. The breakdown is useful for understanding how the academic community is investigating the topic. The following sections examine these themes through a series of focussing questions.

Who is vulnerable?

Before we delve into the research, it is crucial to identify those who are most at risk of being exposed to the harms of deceptive patterns. Some research has been conducted on this topic, particularly focusing on the exacerbation of cognitive biases. We all possess cognitive biases that impact how we understand and process situations. While everyone is susceptible to these biases, this susceptibility can be amplified when an individual's cultural or linguistic background is different from the designer's, as well as for individuals with low literacy, numeracy, or any number of intellectual, learning, and sensory disabilities. For example, a commercial entity wanting to hide their intention to capture sensitive data could exploit a user's low literacy by phrasing the statement in complex legal jargon.

It may be tempting to think of deceptive patterns as only affecting the most vulnerable people, but research indicates otherwise. Although higher levels of education increase the likelihood that deceptive patterns will be detected, research conducted by Bongard-Blanchy et al. [12] showed that awareness does not completely mitigate the effect of the deceptive pattern.

Kitkowska's research [62] provides valuable insights. In their interviews with domain experts, it was mentioned that people may not always be susceptible to deceptive patterns, but we all experience temporary vulnerability in periods of time pressure, stress, or negative mood. In our perpetually online world, commercial entities have ample opportunities to amass data and know a lot about their individual customers. They can use this data to ascertain the temporary vulnerabilities and target marketing campaigns that are directly linked to the personal situation.

In a study by Abbott et al. [1], the authors examined how personality traits can predict the effectiveness of different deceptive pattern styles. The authors created a task where participants were presented with information about a product, utilising the deceptive patterns of social proof, limited quantity, and high demand. The participant's self-reported urge to purchase the product was then analysed. The first result was that exposure to any of the three deceptive patterns led to an increased urge to purchase the product. The most interesting result was that participant personality types dictated which deceptive patterns were most likely to be effective. Extraversion was the top predictive trait for the social proof deceptive pattern, and conscientiousness was the top predictive trait for the high demand deceptive pattern. This shows that the better a commercial entity knows us, the more effective their deceptive pattern usage can be.

One reason for the observations mentioned above is AI technology. AI-powered user profiling methods enable businesses to analyse large volumes of collected user data and categorise various target user groups for different business goals. This indicates that one person no longer belongs to a specific user group but is included in many different groups. Additionally, constantly evolving AI technologies are capable of formulating different deceptive patterns and utilising them on various target user groups to influence users' decisions. Thus, while an individual may avoid a specific form of deceptive pattern, there is no guarantee that they will not encounter another form of deceptive pattern optimised to the individual user. As some AI technologies have already achieved state-of-the-art performance level [51, 67, 117, 132] , they are capable of adapting their responses in a non-static and swift manner, and thus it is extremely difficult for users to stay ahead of all types of deceptive patterns.

Are Deceptive Patterns being regulated?

From an Australian perspective, despite being affected by deceptive patterns, there is little Australian academic research into deceptive patterns. In a report into deceptive patterns impact in Australia, the Consumer Policy Research Centre [28] identified through participant surveys that 28% of respondents had felt manipulated by a website, 17% had felt pressured into buying something, 25% felt that they had shared more information than they wanted.

Although it is difficult from their data to be sure that the result of these impressions are directly attributable to deceptive patterns, it is clear that Australians are feeling subjected to deceptive and manipulative tactics online. Gupta [49] suggests that existing legislation, such as the Competition and Consumer Act (CCA) [141] and Privacy Act [148], needs review to bring Australian law at least to the level of EU and US regulations.

Conversely, other legal opinions [11] posit that the existing consumer and privacy laws are sufficiently protective against deceptive pattern harms. Regardless, several federal bodies, namely the Data Standards Body[15], the Australian Competition and Consumer Commission (ACCC)[16], and the the Office of the Australian Information Commission (OAIC)[17] have worked together to build and regulate the Consumer Data Right (CDR)[18].

The CDR is designing rules for how authentication, consent, and data retention must be implemented. The aim is to give consumers a choice about whether or not they share information with commercial entities, control over how they give this access, and a convenient mechanism by which the access is given and managed. It only applies, however, to companies and consumers that opt into the program, and only in the banking and energy sectors at this stage. Deceptive patterns are a global problem, tackled by various legislative bodies, especially in the EU and USA. Australia can learn from the advancements made by other countries toward implementing stricter measures regarding deceptive patterns, and there is much academic literature that gives insight into how the global community is tackling the problem.

In Europe, the General Data Protection Regulation (GDPR) [143] was passed by the European Parliament in 2016. The GDPR establishes rules for safeguarding the rights and privacy of Europeans concerning the handling of their personal data. It also addresses the free movement of such data between entities within the European Union. It ensures the protection of fundamental rights and freedoms of individuals, emphasising their right to privacy and control over their personal data.

The scandal that brought public attention to the need for legislation like GDPR was the revelation of Cambridge Analytica's harvesting of personal information on Facebook to target voters and influence victories for the 2016 US presidential campaign and the Brexit campaign [19]. The personal data of approximately 87 million users was collected from the company having access to the data collected from only 270,000 consenting users (0.27 million). Through these consenting users, Cambridge Analytica was also able to access personal information from their friends. Public awareness of the data harvesting arose from a whistleblower and caused outrage from the media, general public, and politicians who united in their condemnation of both Cambridge Analytica and Facebook for the violation of their privacy. As a result, the UK gave the Information Commissioner's Office (ICO) new enforcement powers, resulting in a £500,000 fine of Facebook. Despite the universal outcry for privacy enforcement, many companies continue to skirt regulations.

In the largest study of European website adherence to GDPR, Nouwens [98] found that only 11.8% met legal requirements based on European law, highlighting that regulatory action is needed to prosecute GDPR violations. The authors make clear that while regulation and standards are great to have, it is the enforcement that will truly protect consumers. The general consensus is that GDPR has enough scope in its focus on data privacy to protect against the impacts of deceptive patterns, but the enforcement is lax and therefore deceptive patterns still exist and have a negative impact on consumers.

In the USA, Michaels [85] suggests that the Federal Trade Commission (FTC) has the authority to protect against data collection via deceptive patterns but recommends that new legislation be created that specifically caters to deceptive patterns. Luguri and Strahilevitz [78] suggest that where consumers have entered into a contract as a result of deceptive patterns, that contract could be deemed voidable due to the lack of consent. They also propose that audit capabilities should be added to the FTC arsenal, specifically regarding the consent process. Lastly, they note that the rapid proliferation of deceptive patterns is due to many companies' A/B testing revealing them to be profitable.

15 https://consumerdatastandards.gov.au/

16 https://www.accc.gov.au/by-industry/banking-and-finance/the-consumer-data-right

17 https://www.oaic.gov.au/consumer-data-right

18 https://www.cdr.gov.au/

Regulating the Deceivers

Kogan Australia[19] was fined $310,800 in 2021 for violating Australian spam laws. The Australian Communications and Media Authority (ACMA) investigation found that Kogan sent over 42 million marketing emails to consumers from which they could not easily unsubscribe.

Instead, consumers were required to set a password and log into a Kogan account. Additionally, Kogan agreed to the appointment of an independent consultant to review its systems, processes, and procedures, and implement any necessary changes.

Kogan was also awarded the Shonky award for their Kogan's 'First' program [149], which offered free shipping for a $99 annual fee. Consumers complained about accidentally signing up for this program due to pre-checked boxes during the checkout process.

Although Kogan states that customers can cancel their membership within the 14-day trial period and that they email reminders, some customers argued that these emails are easy to miss or ignore, especially if unaware of the initial sign-up.

19 An online department store, available at https://www.kogan.com/au/

Acknowledging that A/B testing is not an inherently nefarious process, the authors' concern is that it is being applied to improve the deceptive patterns rather than improve the consumer experience.

There is a call for a more holistic approach, with some authors noting that existing laws in both the US and EU are limited in scope, perhaps just protecting data or guarding against unfair competition [105]. Porto and Egberts suggest new regulations should encompass all aspects of deceptive patterns, including market failures, reduced trust, unfair competition, and data dominance.

This approach would foster collective welfare, both for businesses and consumers. Efforts are being made to this end, including the Deceptive Experiences To Online Users Reduction Act (DETOUR Act) [126] and the American Innovation and Choice Online Act [63] in the US, and the Digital Services Act [145] and the Digital Markets Act [147] in the EU. These all attempt to protect collective welfare by focusing on market fairness and competition. Challenges still lie, however, in enforcement and addressing the impact on users.

From an economics perspective, it is clear that in the short term, deceptive patterns increase profitability. The cost for the consumer is higher prices and paying for products they had not intended to order, and in the case of data, the consumer gives away more data than they wanted. Noisianinen and Ortega [97] argue that legal design is actually better for businesses in the long term. The authors present several incentives for companies to prefer legal design including that they promote comprehension, usability, plain language, clarity, and that all of these contribute to building consumer trust in the company and the contractual obligation they are entering into. Implementing legal design signals a company's quality, trustworthiness, and willingness to obey contractual obligations. All of this ultimately fosters profit creation and long-term business development.

In terms of resolution, Gray et al. [46] discuss the ethical concerns surrounding technological systems and services, particularly relating to the design choices regarding consent banners[20]. The authors highlight that many small user interface choices can have a large impact on the success of a consent option being answered in favour of the commercial entity. They suggest that by combining knowledge from human-computer interaction, design, and law the ethical concerns surrounding deceptive patterns can be resolved. The human-computer interaction research community has a history of engaging with ethical impact.

20 These will be described in detail in the later 'Cookies' section. In brief, consent banners are user interface components that seek the consumer's consent, often to the collection and use of their data.

Where are Deceptive Patterns found?

Increasingly, deceptive patterns can be found in all corners of our online lives. This troubling trend is not confined to any one part of the web but is instead pervasive in all aspects of our digital interactions. On the web, deceptive patterns are becoming increasingly common. This is true regardless of whether we are browsing on a desktop computer or on a mobile device.

Our digital entertainment is also not immune to this trend. Streaming services, social media platforms, and online games are all potential sources of deceptive patterns. These services often employ complex algorithms and psychological tactics to engage users, sometimes at the expense of user experience and privacy. Mobile applications are another area where deceptive patterns are prevalent. These applications often have access to a wealth of personal information, and misuse of these data can lead to significant privacy concerns.

Recent research supports these observations [36]. Studies have shown that consumers are feeling manipulated by these deceptive patterns [43]. Furthermore, research indicates that this manipulation is not confined to any one platform but is instead being experienced across the web and on mobile phones [48]. These findings underscore a widespread feeling among consumers that they have little to no control over their personal data and privacy. Given this context, it seems the best place to start examining this issue is through the lens of the cookie consent process, a common and often misunderstood element of web browsing.

Deceptive Patterns for Dopamine

Social media companies have learnt how to provide a thrill, a satisfaction, and a drive to view more content and spend more time on their service. Employing patterns in the delaying provision category, such as autoplay, infinite scrolling, and pull-to-refresh, social media services tap into our dopamine crave cycle [127].

These patterns ensure that we are glued to the screen, seeking the next hit of pleasure in the form of a 60-second video, dunking on an outrageous opinion, or a beautiful landscape that we will definitely (not) visit. When the deceptive patterns keep us there for the next hit, we binge, keeping our brains locked into the emotional drive for more, sacrificing our frontal region abilities to plan, critically evaluate, and task-focus.

The social media companies have mastered how to tap into our pleasure and addiction drives in a way that is wholly to their benefit and the detriment of the service user.

Cookies

Cookies are small text files that are created by web servers and stored by your Internet browser. The stored data is used for various purposes, including remembering user preferences and tracking browser activity.

One of the most obvious and prominent arenas for deceptive patterns is in the website dialog and banners that seek user consent for cookies. Before we delve into how deceptive patterns come into play for cookie consent, it is useful to provide some context on why cookies are important for the modern web. Cookies serve a number of positive and useful purposes for both commercial entities and consumers. They aid in identification and personalisation, which allows websites to tailor content specifically to a user's preferences. They aid in session tracking, maintaining information about a user's activity during a single session. Cookies also remember website state, preserving settings or actions from previous visits. Lastly, they aid in authentication and security, ensuring only authorised users can access certain areas of a website.

Cookies can, however, have less desirable uses. They can be used for third-party tracking, allowing external entities to monitor a user's activity across multiple websites. They can also be used for session tracking in a malicious manner, often for the purpose of targeted advertising. Website usage tracking and analytics can also use cookies, monitoring how users interact with a site and which pages they visit.

These undesirable features can give commercial entities valuable data and insights into behaviour and marketing efficacy, and deceptive patterns play a large role in pushing consent for these types of cookies. The fact that users can consent and websites are forced to ask is due to regulations such as the GDPR. Before these regulations, websites could do all these things in the background without the user's knowledge. One problem that these consent dialogs now present is the user fatigue at constantly being presented and having to interact with them [90], as well as the potential for them to obscure information displayed on the page beneath. The length of these policies and the frequency with which we are presented with them, mean it is a practical impossibility for more people to thoroughly read and give informed consent.

The best practice for obtaining consent is to offer users an easy opt-out, but no strict regulations are provided on how this should be done. There are also no strict regulations on how particular choices can be manipulated via tactics as previously discussed. Much research has focused on how deceptive patterns can influence consent in the direction of commercial entities.

In an experiment conducted by Borberg et al. [13], participants were asked to interact with different styles of cookie consent dialogs. They performed a series of interaction tasks and were questioned about whether they noticed the consent dialog and how they reacted to it. They were also asked more general questions about their most common actions, whether they have a tendency to read the information on the notices, and how often they leave a website specifically due to the notice.

The authors found that deceptive patterns are effective in nudging users into accepting cookies. In terms of preference, participants generally prefer designs that make opting out easy and transparent. The authors expressed concern that deceptive pattern nudges lead to user loss of privacy autonomy and lack of control.

Berens et al. [8] investigated what specifically about the design of consent dialogs makes deceptive patterns effective. They presented participants with a range of different button, text, and other interface options, including whether the accept or reject button was highlighted, and whether the phrasing of the explanation text biases acceptance. One of their main findings was that the styling of the reject or accept button has a significant impact, leading participants to prefer the one that the designer emphasised. In their more general observations, the authors noted that while 74% of participants read consent dialog headings, only 34% read the explanation text.

In their examination of 389 German websites, Krisam et al. [66] found that only 21.5% of them allow an easy opt-out. The authors suggest that regulations must emphasise privacy and suggest that browser settings could ensure the user's privacy is honoured consistently across all websites. Similarly to Berens et al. [8], the authors note that the definition of what "technically necessary" cookies actually constitute presents commercial entities with ample opportunity to hide desired data collection within the detailed text.

Graßl et al. [42] conducted a similar experiment, where they looked at the participants' perceived control in regard to cookie consent. They found that most participants chose the privacy-unfriendly option and reported that they felt a lack of control over the consent process. The findings showed that it was the design nudges that influenced the participants' choices. The experiment highlights the legal limits of consent if it can be so easily manipulated. It is worth mentioning that the promotion of public awareness could aid in preventing some of this.

One potential concern associated with cookie consent dialogs is with the adoption of AI to adapt the design of consent forms, utilising data about the user. These forms can be designed with AI optimisation technology to adapt to each user group and more effectively draw their consent. This approach carries a higher risk, as it may potentially manipulate user choices, or lead them into providing consent more readily without full understanding of what they are agreeing to.

Cookies also enhance the efficacy of other AI powered deceptive patterns, such as recommendation systems and targeted advertisement. Miehling [86] showed that when consent rates are demographic-dependent (e.g. age, gender), a user's decision to disagree to share their cookie provides more useful information to the recommendation system than the user's agreement to share cookies.

Kazienko and Adamski [58] proposed AdROSA, a method of automatic banner personalization for user adaptive advertisement, and showed cookie information is a highly effective resource of advanced deceptive patterns.

Waldman [125] and Jarovsky [53] note that while GDPR has been a dominant force in bringing these shady actions regarding cookie consent to light, other countries are trying to solve the same problem. In the US, the dominant approach for how cookie consent should be obtained is via notice-and-consent process. Using this process, commercial entities can manipulate the outcome to their own means. Waldman notes that traditional data protection laws fail here as they focus on whether or not the data collection has been agreed upon, not the manipulative means by which that consent was garnered. Similarly, in Brazil, Jarovsky notes that data protection laws are inadequate. Specifically, the Brazilian

General Law on the Protection of Personal Data [144] leaves a blind spot for deceptive patterns. The authors suggest that the law should also consider fairness in data protection so that the means of data collection consent can be encompassed.

A potential solution to the inconsistency of presenting cookie consent dialogs is for regulators to enforce adherence to particular standards. This would aid developers, who would have less difficulty understanding the complex regulations and standards presented by every country their service appears in.

As an answer to this, there are many online consent management platforms (CMPs) that offer a plug-in service to handle consent for a fee. Unfortunately, suspecting that these paid CMPs themselves implement deceptive patterns into their framework, Toth et al. [120] reviewed 10 popular such CMPs, including the two top providers Quantcast and OneTrust. The authors found that CMPs offer solutions that maximise the likelihood of user consent, meaning that they are targeted toward the commercial entity's priorities, rather than privacy of the user. Even within the CMP's own marketing, the authors showed that CMPs often use deceptive tactics to convince web developers to subscribe to their premium tiers.

What could be an industry that is providing a valuable service for compliance to regulatory bodies and promoting user privacy safety and autonomy online is instead contributing to the deceptive pattern economy. Google Chrome's recent "Enhanced Ad Privacy" settings allow users to be targeted with ads based on their online activities and history, unless the user is aware of the setting and explicitly turns it off, across four different settings [27].

Beyond Cookies

Apart from cookie consent, e-commerce is the domain in the online world where deceptive patterns can most prominently be found. To illustrate the sheer scale, Mathur et al. [81] conducted a large-scale search for deceptive patterns across over 11,000 of the world's most popular shopping websites, analysing over 53,000 individual item store pages. The authors identified 1,818 deceptive patterns, representing around an 11% occurrence rate. The authors found so many different types of deceptive patterns that their search led to the formation of one of the most influential typologies, one that contributed to the formation of the IVE deceptive patterns typology in Part II of this report. Importantly, the authors note that many of the deceptive patterns are deliberately deceptive, exploiting cognitive biases of anchoring and framing to boost sales on that website.

As opposed to the consequences to the user that arise from deceptive patterns in the cookie consent realm, deceptive patterns in e-commerce can have direct and negative impacts to the user's finances. This type of deceptive and manipulative behaviour is better regulated by consumer law, but the global nature of the online marketplace and the lack of enforcement mean that clearly many online stores are utilising deceptive patterns.

Tying into the earlier vulnerability discussion, Koh and Seah [65] note that false urgency deceptive patterns are not only particularly effective in driving sales, but older generations are most susceptible. Young people, however, are not exempt. In their study, van Nimwegen and de Wit [95] found that young people were actually more susceptible to deceptive pattern influence. The authors posit that this is likely due to older customers having a more cautious attitude toward spending in general.

Social media is another domain that prominently features deceptive patterns, primarily to keep users engaged with the service. In Mildner's [87] research, the authors used domain experts to examine the four most popular social media services, Facebook, Instagram, TikTok, and Twitter. Looking at only those four services, the experts identified 44 deceptive patterns, mostly aimed at engaging and governing the user's attention. Armed with a wealth of data about their user's demographics, content preferences, and historical behaviour, social media companies' algorithms can deploy deceptive patterns in highly targeted and effective manners. This helps explain how endless hours can so easily disappear before you realise when using these applications [23]. Specifically, AI technologies play an important role in the potential development of deceptive patterns on social media as they are capable of effectively harvesting massive social media data [136], aggregating [110] and extracting crucial user information [140], identifying coordinated user accounts [112], predicting user personality [25], and linking social networks [41].

Not solely confined to social media services, but definitely prevalent there, is the use of deceptive patterns to make the account deletion process difficult. Lingareddy et al. [76] conducted an analysis where they stepped through the account deletion process on many prominent social media services. They noted that deceptive patterns are frequently found relating to limited deletion options, confusing terminology, and a lack of transparency around data retention. Even if the service offers a deletion option, the authors noted that the process is often accompanied by confirmshaming or extra external steps (such as contact support via phone or email).

The Right to be Forgotten

Under GDPR law, service users have the ability to request the removal or deletion of certain information about themselves from online platforms and search engine results. This right is important as it enables an individual some recourse to address inaccurate or outdated information about themselves that is publicly available.

Despite GDPR and other jurisdictions invoking such laws, immortal accounts is still a prominent deceptive pattern of concern. Here in Australia, service users do not have a right to be forgotten. The closest they have are some relevant Australian Privacy Principles (APPs) in the Privacy Act[21] that enable them to request that a commercial entity remove the personal information. This removal is at the commercial entity's discretion, and no rule forces them to do so. The impact of this on the previous sharing of account data however means account data may still exist in third parties.

21 https://www.oaic.gov.au/privacy/australian-privacy-principles

Closer to home, Lacey et al. [69], in their examination of how New Zealanders are impacted by deceptive patterns, shared that New Zealanders are most likely to encounter deceptive patterns when trying to cancel a service or subscription where the commercial entity's priority is clearly to retain the customer.

Mobile Apps

In today's world, we all carry a mobile phone with us. These devices, which we keep in our pockets or next to our beds, play a major role in governing our lives. They come equipped with a variety of apps that we use for socialising, managing our finances, shopping, playing games, and accessing the Internet. Deceptive patterns may have originally found their place on the web, where they were used to increase profits and maximise data collection. However, the potential for screen time with mobile phones makes deceptive patterns even more attractive on this platform.

A 2020 study by Di Geronimo et al. [31] conducted one of the largest examinations of deceptive patterns in mobile apps. The authors analysed over 200 popular apps and classified any observed deceptive patterns. Next, they had 589 participants complete a questionnaire to determine if they could spot any deceptive patterns in video recordings of a selection of these apps. The results showed that among the 240 included apps, 95% included one or more deceptive patterns in their interfaces. In total, 1787 deceptive patterns were found, with an average of 7.4 malicious designs per application. The concern does not stop there.

The authors also demonstrated that in the majority of cases, participants could not perceive the deceptive patterns. The prevalence and imperceptibility of deceptive patterns on mobile have several serious implications. These include the amount of sensitive data that users unwittingly surrender, the amount of time users lose due to manipulative attention grabbing, and the amount of money users are convinced to unnecessarily spend. There are also questions about how regulators can increase user awareness of the presence and impact of mobile deceptive patterns.

Di Geronimo's study is not the only large-scale investigation in this area. Another conducted by Long et al. [77] in 2023 analysed over 150 Chinese mobile apps. They found that 82% of the apps featured at least one deceptive pattern. The most frequently used pattern, found in 78% of apps, was the asymmetric button. This pattern occurs when a designer deliberately emphasises one button over another, usually when the commercial entity wants a user to pick one option in particular, such as accepting all for data sharing privileges.

A concerning finding of the study was that the popularity of an app did not mean that fewer deceptive patterns were observed. This is particularly worrisome because more popular apps generally have a higher level of user trust, making deceptive patterns even more effective in these apps. In fact, it has been shown that app trustworthiness leads to a lower chance of deceptive pattern detection [9]. While the frequency of deceptive patterns did not change with app popularity, the authors observed that the type of deceptive patterns employed did change from low to high popularity apps. They found that overt patterns like fake buttons occur less in high popularity apps, and are replaced by covert patterns like misleading text. This seems to suggest that designers are aware of the negative associations that users have with deceptive patterns, and prefer to hide their deceptive tactics when the app is popular.

The terms and conditions of the major app stores are supposed to protect users against the malicious intents of app developers. However, a study by Singh et al. [115] showed that despite the Google Play Store's prohibition against certain types of harmful applications (for example, those that falsely promote rewards for performing small tasks like viewing advertisements or completing surveys), these types of apps are still present. The authors also found that deceptive patterns are not only present in the apps themselves, but also in the app stores that are ostensibly protecting us. They found many cases of fraudulent reviews that artificially boost the popularity and rating of apps which, as we have already shown, impacts trustworthiness and thus the impact of the app's deceptive patterns.

As shown in existing works [4, 24], current large language models possess a significant capability to generate fluent fake reviews that are nearly indistinguishable from those generated by humans.

Due to the inherent nature of mobile devices, long term user engagement of a mobile app is crucial to the success of it. As a result, user engagement has been extensively studied. In enhancing user engagement, Carrion et al. [20] developed a method which blended advertisements with organic content, and applied it on jd.com's mobile application to improve user engagement on the app. Their approach adopted an objective function that jointly considers the effect of advertisement and organic components on the user engagement.

The study demonstrated optimised allocation of advertisements and organic contents can improve user engagement. Tian et al. [119] studied the prediction of user engagement on mobile apps by proposing prediction models that infer which app a user will use next and how long the user will stay on the app, increasing the potential for users to have deceptive patterns and customisations follow them between apps. In iOS 14.5, Apple introduced new app privacy settings to help crack down on non-consented user tracking, resulting in Facebook admitting the loss of data and associated targeting advertising revenue would reduce sales by $10b in 2021 alone [73].

Adaptive User Interfaces

Around the turn of the century, academia was exploring the potential for AI to customise the user interfaces of web sites and applications that users were interacting with, optimising their layout and function depending on what the user required, creating a form of intelligent user interface [16, 124].

Modern nefarious approaches could include the use of user profiling to detect the likelihood of the user actually reading terms and conditions, and collapsing or expanding a given section of a user interface behind a drop down panel. Both approaches are technically presenting the terms to the user, however how evident the terms are in each one depends on how likely the system deems the user to read them. Even basic customisation such as the use of colour schemes can imply authority and approval, for example the previous AEC voting material example. Research shows that even just using the user's favourite colour can improve conversion rates [55, 107], despite being such a small adaptation requiring little information about the user.

Whilst technically not a form of hypernudging, approaches now exist to cause emails to change appearance, or even hide or show particular content once they have been forwarded on by a user [128]. The use of extensive data mining of a user's data can also be weaponised against them,

enticing users to continue to share, or even share ever more information with a service if the service can mine the user's data for value propositions stating, for example, that "You've saved $X on service Y using this app". If a service knows a user will receive an overload of notifications in the morning, it makes sense to send any passive "review our new terms and conditions" messages around this same time, similar to bad press releases coming out towards the end of the week [15].

What can be done?

The academic work that we have presented thus far has painted a grim picture of how prevalent, harmful, and pervasive deceptive patterns are. It is clear that even being aware of these patterns cannot completely protect a person from their influence. Therefore, the best protection is regulation and enforcement that gives teeth to that regulation. However, regulatory processes always lag behind the state of the art. Even if we were to create perfect regulation that was universally accepted, new technologies and deceptive strategies would leave holes for new deceptive patterns to emerge. We do not want to put the sole onus of protecting oneself on the users, as that will not address the overarching problem of commercial entities profiteering with deceptive patterns.

However, that does not mean that tools for individuals do not have their place. So, we look to strategies other than regulating, such as better user experience design that promotes bright patterns, as well as deceptive pattern detection and mitigation.

Detection

In the realm of detection, if artificial intelligence can be used to create more effective deceptive patterns, it can also be used to create better deceptive pattern detecting tools. Several researchers are investigating this solution. Kirkman et al. [61], for example, built and evaluated a system that could automatically extract cookie dialogs and detect the existence of 10 different deceptive patterns within those dialogs.

In their testing, they automatically detected 2,417 cookie dialogs from their website sample of 10,992 websites. Within those, their system identified 3,744 deceptive patterns. This type of system could exist within the browser or as a background app on mobile, giving users some warning about deceptive pattern presence.

Mansur et al. [79] took a much broader approach, creating a system named AidUI that used computer vision and natural language processing to detect deceptive patterns in the entire user interface of a website. The system analysed the website's use of text, iconography, colour, and space to predict whether deceptive patterns were present. The authors' testing showed that their system performed well at identifying deceptive patterns when cross-checked by a human examiner. This type of solution could be used as an early warning system for a user, notifying them to pay attention to areas of the interface that the software deems unsafe.

Several other research groups, such as Yada et al. [134, 135], Stavrakakia et al. [116], and Kocyigit et al. [64], are working in the same area. All have reported high success rates for identifying deceptive patterns in their sample websites and mobile apps. The more comprehensive the model training deceptive pattern datasets become, the higher precision that these software solutions will have. While it will always be an arms race between AI creating deceptive patterns, and AI identifying deceptive patterns, it is likely that software detection solutions like those discussed will be integrated into our browsers and phone software to help us deal with an overwhelming and undetectable amount of deceptive patterns.

While most research in this category focuses on software solutions, there is also some interest in more manual tools for detecting deceptive patterns. Research conducted by van Nimwegen et al. [94], for example, revolves around the creation of the System Darkness Scale. This is a tool that individuals could use to rate a service's level of deceptive pattern use. The scale features items that ask the user to rate whether they felt tricked, pushed into spending money, unknowingly forced to take certain actions, or deceived in any way. The eventual deceptiveness score can be used to give a subjective estimate of how much an evaluated system seems to contain manipulative elements.

In a similar vein, Mills et al. [88] created a framework that could function as an auditing tool. Looking in particular at the account creation and deletion user workflows, the authors used their framework to showcase how the creation and deletion journeys differ, using metrics such as number of clicks and navigation pathways. Their framework can show where deceptive patterns are influencing the metrics, making some workflows easy, while others are objectively more cumbersome. Tools such as these could be useful for regulators when evaluating whether or not a service features deceptive patterns and if they breach regulation.

Mitigation

A smaller amount of AI-based research has gone one step beyond just detection to actually having software actively address detected deceptive patterns. Porcelli et al. [104] built a browser extension that first enables users to create a profile of how they would prefer to respond to cookie consent dialogs, as per GDPR requirements. Then, the extension would automatically detect cookie consent dialogs when visiting a website, send the contents of the dialog to ChatGPT, which would then weigh the options against the user's defined preferences and if a match could be made, consent would be given on behalf of the user. If the user's preference cannot be matched, the user is notified of the discrepancy and can interact with the dialog manually. A solution like this is minimally

invasive, and would address many of the difficult to regulate patterns, such as hidden legalese stipulations, as previously discussed.

In an experiment conducted by Schäfer et al. [111], the authors tried different levels of actively addressing detected deceptive patterns on websites. Their system provided deceptive patterns countermeasures, and their experiment tested which the participants most preferred. The level of intervention ranged from highlighting the pattern for user attention and hiding the deceptive pattern completely. The most invasive option, being to hide and not notify the user, would be too prone to AI mistakes and could lead to user frustration when essential information is gone. The participants in the experiment reported preference for the condition where they were made aware of the deceptive pattern but no action was taken by the system. This experiment is useful for showcasing how general education to deceptive patterns is important, but also how we could employ software to help nudge users toward noticing deceptive patterns, limiting, but not entirely preventing, their covert effectiveness.

Despite the significant areas of concern from AI technologies, AI can also effectively anonymize and pseudonymize large volume user data sets to enhance privacy protection. For example, Levy et al. [74] explored how user-level differential privacy can obscure an individual's contribution to group characteristics, but this requires the group to be of interest, not the individual.

This could be an effective tool in contexts where a personalisation approach is adopted but targets larger groups, rather than individuals. For some scenarios, the lack of disclosure of data can itself even be informative, with Leemann et al. [71] looking at safeguarding the privacy of users who choose not to share their data.

How is AI powering Deceptive Patterns?

Across the landscape, the potential for AI was identified as an enabling technology that, given its rapid improvements, has the potential to increase the impact of all deceptive patterns. The European Commission's regulatory framework proposal on AI [151] defined four levels of risk, from minimal to unacceptable. The potential for AI to impact users subliminally or via obscure manipulation (i.e. deceptive patterns) was defined as part of the top level of unacceptable risk [32].

While the manipulation of users in mass markets existed prior to the modern AI age, AI technologies have boosted it [108]. AI's capabilities in this space can be divided into four different areas:

1. The use of AI in mass data aggregation and building user profiles.
2. The use of AI in applying those profiles to target users (via means of deceptive patterns).
3. The use of AI in for detecting when deceptive patterns are in use.
4. The use of AI in the mitigation of deceptive patterns (such as changing default values).

We observe that AI based deceptive patterns are presented in various ways, often unnoticed by users. For example, users often do not distinguish between original customer reviews and fake ones [4, 39] generated by AI algorithms. In addition, users may not be aware of the recommendation system optimised by AI technologies [133]. As we discussed, users are particularly vulnerable to AI empowered deceptive patterns or applications with malicious intentions since AI based deceptive patterns have the capability to dynamically adapt to specific users or the current stages of a user's need (e.g. a supermarket sale). These manipulative approaches are made possible by advanced harvesting, aggregation [70, 80] , and user profiling [21, 75, 101, 113] techniques. Further concern arises from the rapid development of AI technologies that can be adopted in the generation of deceptive patterns or other malevolent applications including large language models [56], face recognition systems [52, 53], voice recognition applications [139], emotion detection algorithms [52] and many more.

It is important to note that AI's ability to execute effectively is based on being fed relevant data. A data economy has been created where the buying and selling of user information is common, and now AI can aggregate the data at scale. The user's data, originally provided to a particular service, is collected by AI, processed, and fed back into the data economy completely disconnected from that original service.

Even for a user who does not provide information online, they are profiled based on other users that they know. This data is a double advantage for companies, offering them targeting for their own patterns, but also then sale of that data to advertisers, where advertisers can target ads to users in automated live-bid auctions that occur every time an ad appears to the user.

This secondary monetisation of captured data is especially crucial for private information, with the Grindr dating app facing a class action around illegally sharing users' HIV statuses [57] with third parties, and in Australia the HealthEngine medical booking app facing fines for selling patient data [82]. The often unregulated capture and permanent retention of user information primarily only benefits the business asking for it. Perhaps one of the earliest examples of the mass breach of user privacy in the public sphere in this regard was the previously mentioned Facebook-Cambridge Analytica data scandal.

Separately from the Cambridge Analytica event, the phone numbers of 530 million Facebook users were leaked in a single database in 2021 [83]. Showing the cavalier attitude to such a leak, Facebook downplayed the leak, saying that the data had previously been leaked in a mix of smaller leaks. The mass aggregation of a single database, representing the data of 1 in every 15 people on the planet, shows the scale on which these data holders operate, and the role that modern AI systems will play in continuing to build and enhance these profiles.

Most critically, despite a user's data already being weaponised against the user by data brokers, once leaked, the user is now open to more malicious attacks far beyond influencing actions in an app, such as identity theft. Other mass breaches include AT&T [7] and US credit data broker Equifax [37], with the latter resulting in a formal settlement with their 147 million users, or approximately 45% of the US population. Australian mass breaches include the Medibank Private [121] and Optus leaks [100]. The Optus leak led to a number of criminal arrests for identity theft, leaving customers to feel "powerless" [68]. As was the case with the Medibank breach, quite often the leak comes from a third-party company engaged to provide services utilising user data [3], again reinforcing how regularly data flows between entities unknown to the end user.

Data sharing agreements are now common place between companies, which whilst often appear to advantage the user, (for example, logging in using Facebook rather than having to remember a password), unwittingly lead to the user's private information being shared [50, 93], including companies such as Netflix, Spotify, and

the Royal Bank of Canada gaining access to user's private Facebook chat messages for analytics and targeting purposes. Even vocal privacy advocate Apple had access to user's Facebook contacts and calendars, regardless of whether sharing had been disabled by the user. Potentially even more concerning, Apple claimed it was unaware it had this access as part of its agreements with Facebook. Such concerns about sharing of private messages was of little concern for Facebook CEO Mark Zuckerberg who had a unique ability to secretly delete his own Facebook messages [40], a function which at the time was unavailable to the general public. This all collectively shows the extent to which user data is now shared without open and ongoing consent, and without a permanent cancel button to remove that data from circulation. Even for a CEO concerned about the sharing of his own messages, the ability to remove that valuable data was only implemented for himself, until it became public knowledge.

A report by the Office of the Australian Information Commissioner [99] presented that almost half of Australians disclosed being told by an organisation that their data had been breached in the prior year. Furthermore, 76% of those reported some harm as a result, perhaps experiencing psychological harm (12%), financial or credit fraud (11%) and even identity theft (10%). The data economy now supercharges user profiling far beyond the app, website, or organisation the consumer originally engaged with and entrusted to their information. With organisations seeking to fuel next generation profiling and AI systems with evermore data, the risks have never been more apparent.

Privacy laws looking to protect the collection and sale of personal data have now extended to new laws aimed at protecting the consumers' own neurological signals being processed by apps [89]. Without a specific user's data, wider profile modelling across large numbers of other users now has the potential to increase AI's effectiveness to target a new user that has only appeared in a few data points, as with the Cambridge Analytica leak where a user that had never used a service, can have their data collected by it anyway. From this data, social media and online content platforms can now rapidly adjust their recommendations for new users based on their personality [25], ensuring continued engagement with the platform.

Given the nature of sales, many new AI tools, such as chatbots, are being developed for e-commerce and could themselves be subject to deceptive patterns, as well as previous well-known approaches, such as "suggested" products on shopping sites and "recommended" social media content.

The rapid increase in AI's capabilities, the dropping cost of AI systems to use, along with the prevalence of new off-the-shelf AI tools has significantly reduced the barrier to entry and adoption for AI. This adoption of AI by new tools and services can be seen in the rapid influx of chatbots into websites and services, often utilising OpenAI's ChatGPT or other large language model AI system. Such adoption of AI by non-AI and non-technology companies will only grow as more powerful AI systems become more and more accessible to organisations.

Conclusion

Throughout this report, we have shown that deceptive patterns are deceptive and manipulative tactics that can be used to strip individuals of their autonomous decision making. The IVE deceptive patterns typology presents a model that shows that the wide range of deceptive patterns can influence consumers in four distinct manners:

1. actively deceiving or manipulating people by controlling the presentation of information (active misleading actions),
2. passively deceiving or manipulating by hiding or delaying information (passive misleading omissions),
3. controlling a person's choice by pressuring or forcing them (undesirable imposition), or
4. placing unnecessary restrictions on how people can interact with their service (undesirable restriction).

This model helps frame the impact deceptive patterns have on consumers and allows us to focus regulation on preventing these harms in specific areas.

In examining where we find the influence of deceptive patterns, we have shown that they are prevalent in many online services due to their ability to successfully enable commercial entities to modify consumer behaviour, leading to greater data collection potential and ultimately higher profits and other key performance metrics.

deceptive patterns pose a significant threat to consumers. This report argues that when considering the potential harm that deceptive patterns pose, we should focus on consumers' autonomy, data privacy, financial security, and their ability to trust the online services that they use. As we become ever more reliant on online services, it is crucial that we do not surrender this space to manipulation and deception. Losing trust in our institutions and the information they provide is not an option. Therefore, we must promote a space that values truth, and the wellbeing and privacy of the individual.

On the other hand, this report argues that we should also consider the commercial entities, with a goal of fostering an environment where they can thrive online. Trust is a two-way street, and if consumers trust businesses, they will spend their money there and ideally gain benefit from doing so. Thus, we want commercial entities to offer services that put the consumer first, build trust, and offer a valuable service.

The current landscape shows that despite many jurisdictions having regulations that control data privacy and consumer rights, these regulations are inadequate. They often fall short of preventing commercial entities from implementing deceptive patterns, usually due to a lack of enforcement or barriers to invoking enforcement. Furthermore, these regulations are not equipped to address the subtler vulnerabilities that deceptive patterns target, such as emotional manipulation, minor visual tricks, and misleading framing of information. Even when the general public is familiar with the deceptive tactics of the services they use, they are still influenced into performing the actions that the nefarious design intends.

Deceptive patterns are widespread online in areas such as cookie dialogs, e-commerce, and social media, and pose risks unique to every age group. If left unchecked, the rising power of AI could lead to hyper-personalised and exponentially more effective deceptive patterns. Data collection is already the driver for commercial entities. AI enables processing and utilisation of Big Data on a scale previously unimaginable, and if left unchecked, AI could make the problem much worse and much more difficult to control. This is further compounded by the rapid exchange of data across international borders between various apps, data brokers, and customers, as well as the nature of apps and websites serving users far beyond the borders of their own local country.

In response to the problems posed by deceptive patterns, this report identifies three potential focus points for future investigation:

1. The modification of existing regulations or creation of new regulations that specifically target the finer points of deceptive patterns that current regulations do not cover.
2. Fighting fire with fire by supporting the development of AI-based services that detect and counter deceptive patterns. This could be by regulating browsers and mobile operating systems and app stores to require that consumer protection against deceptive patterns be integrated.
3. Raising public awareness, similar to phone scams and identity fraud. While awareness in itself is not always a perfect defence against deceptive patterns, a degree of caution is beneficial. Public awareness also has the advantage of inspiring bottom-up pressure toward regulators. Cyber security training offered in schools offers an opportunity in training the next generation to be aware of not just phishing emails, but also other more general misrepresentations, including deceptive patterns.

The pervasive use of deceptive patterns in online services is a serious issue that poses considerable threats to consumer autonomy, data privacy, financial security, and trust in these services. Despite existing regulations, the subtle manipulations employed by these patterns often slip through the cracks, leaving consumers vulnerable. Any regulations must be well considered, as actors will always seek exceptions to the rule. If some deceptive patterns are blacklisted, bad actors will see to make slight modifications such that newer versions are not explicitly blacklisted.

This report suggests that the fight against deceptive patterns may require a multi-pronged approach. As the online world continues to evolve and integrate more deeply into our lives, it is clear the continued impact these approaches will have in the promotion of a more trustworthy, consumer-centric digital space.

Bibliography

1. Abbott R, Sin R, Pedersen C, Harris T, Beck, T, Nilsson, S, Dong, T, Wang, Y, Li, Y (2023) The role of dark pattern stimuli and personality in online impulse shopping: An application of S-O-R theory. Journal of Consumer Behaviour 22:1311–1329. doi: https://doi.org/10.1002/cb.2208

2. Acquisti A, Brandimarte L, Loewenstein G (2015) Privacy and Human Behavior in the Age of Information. Science 347:509–514. doi: https://doi.org/10.1126/science.aaa1465

3. ACS Medibank Finally Reveals 'Rookie Mistake' in Breach. In: Information Age. https://ia.acs.org.au/article/2023/medibank-finally-reveals--rookie-mistake--in-breach.html. Accessed 23 Apr 2024

4. Adelani DI, Mai H, Fang F, Nguyen HH, Yamagishi J, Echizen I (2020) Generating Sentiment-Preserving Fake Online Reviews Using Neural Language Models and Their Human- and Machine-Based Detection. In: Barolli L, Amato F, Moscato F, Enokido T, Takizawa M (eds) Advanced Information Networking and Applications. Springer International Publishing, Cham, pp 1341–1354

5. Ahuja S, Kumar J (2022) Conceptualizations of User Autonomy Within the Normative Evaluation of Dark Patterns. Ethics and Information Technology 24. doi: https://doi.org/10.1007/s10676-022-09672-9

6. Arkes HR, Blumer C (1985) The Psychology of Sunk Cost. Organizational Behavior and Human Decision Processes 35:124–140. doi: https://doi.org/10.1016/0749-5978(85)90049-4

7. AT&T AT&T Addresses Recent Data Set Released on the Dark Web. https://about.att.com/story/2024/addressing-data-set-released-on-dark-web.html. Accessed 23 Apr 2024

8. Berens B, Bohlender M, Dietmann H, Krisam C, Kulyk, O, Volkamer, M (2024) Cookie Disclaimers: Dark Patterns and Lack of Transparency. Computers & Security 136. doi: https://doi.org/10.1016/j.cose.2023.103507

9. Bhoot AM, Shinde MA, Mishra WP (2021) Towards the Identification of Dark Patterns: An Analysis Based on End-User Reactions. Proceedings of the 11th Indian Conference on Human-Computer Interaction 24–33. doi: https://doi.org/10.1145/3429290.3429293

10. Bindra S, Sharma D, Parameswar N, Dhir S, Paul J (2022) Bandwagon Effect Revisited: A Systematic Review To Develop Future Research Agenda. Journal of Business Research 143:305–317. doi: https://doi.org/10.1016/j.jbusres.2022.01.085

11. Bogard S, Shah N, Gole T (2024) Regulators Shine a Light on Dark Patterns. In: Gilbert + Tonin. https://www.gtlaw.com.au/insights/regulators-shine-light-dark-patterns. Accessed 22 Apr 2024

12. Bongard-Blanchy K, Rossi A, Rivas S, Doublet S, Koenig V, Lenzini G (2021) "I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!" - Dark Patterns from the End-User Perspective. In: Designing Interactive Systems Conference 2021. ACM, Virtual Event USA, pp 763–776

13. Borberg I, Hougaard R, Rafnsson W, Kulyk, O (2022) So I Sold My Soul": Effects of Dark Patterns in Cookie Notices on End-User Behavior and Perceptions. Proceedings of 2022 Symposium on Usable Security and Privacy

14. Bösch C, Erb B, Kargl F, Kopp H, Pfattheicher S (2016) Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. Proceedings on Privacy Enhancing Technologies 2016:237–254. doi: https://doi.org/10.1515/popets-2016-0038

15. Brady B (2020) The Friday News Dump - Does It Work? In: Xenophon Strategies. https://xenophonstrategies.com/the-friday-press-release/. Accessed 23 Apr 2024

16. Brdnik S, Heričko T, Šumak B (2022) Intelligent User Interfaces and Their Evaluation: A Systematic Mapping Study. Sensors 22:5830. doi: https://doi.org/10.3390/s22155830

17. Brignull H (2010) Dark Patterns

18. Brignull H (2023) Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You, 1st ed. Testimonium Ltd

19. Cadwalladr C, Graham-Harrison E (2018) Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach. The Guardian

20. Carrion C, Wang Z, Nair H, Luo X, Lei Y, Gu P, Lin X, Chen W, Jin J, Zhu F, Peng C, Bao Y, Lin Z, Yan W, Shao J (2023) Blending Advertising with Organic Content in E-commerce via Virtual Bids. Proceedings of the AAAI Conference on Artificial Intelligence 37:15476–15484. doi: https://doi.org/10.1609/aaai.v37i13.26835

21. Chamoso P, Bartolomé Á, García-Retuerta D, Prieto J, De La Prieta F (2020) Profile Generation System Using Artificial Intelligence for Information Recovery and Analysis. J Ambient Intell Human Comput 11:4583–4592. doi: https://doi.org/10.1007/s12652-020-01942-y

22. Chapman GB, Johnson EJ (1999) Anchoring, Activation, and the Construction of Values. Organizational Behavior and Human Decision Processes 79:115–153. doi: https://doi.org/10.1006/obhd.1999.2841

23. Chaudhary A, Saroha J, Monteiro K, Forbes, A G, Parnami, A (2022) "Are You Still Watching?": Exploring Unintended User Behaviors and Dark Patterns on Video Streaming Platforms. Proceedings of the 2022 ACM Designing Interactive Systems Conference 776–791. doi: https://doi.org/10.1145/3532106.3533562

24. Chiang H-Y, Chen Y-S, Song Y-Z, Shuai H-H, Chang JS (2023) Shilling Black-Box Review-Based Recommender Systems Through Fake Review Generation. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 286–297. doi: https://doi.org/10.1145/3580305.3599502

25. Christian H, Suhartono D, Chowanda A, Zamli KZ (2021) Text Based Personality Prediction From Multiple Social Media Data Sources Using Pre-Trained Language Model and Model Averaging. Journal of Big Data 8:68. doi: https://doi.org/10.1186/s40537-021-00459-1

26. Claburn T Amazon Fined Almost $8M in Poland Over 'Dark Patterns.' https://www.theregister.com/2024/03/29/amazon_poland_fine/. Accessed 2 Apr 2024

27. Claburn T Google Chrome Pushes Browser History-Based Ad Targeting. https://www.theregister.com/2023/09/06/google_privacy_popup_chrome/. Accessed 23 Apr 2024

28. Consumer Policy Research Centre (2022) Duped by Design. Manipulative Online Design: Dark Patterns in Australia

29. Conti G, Sobiesk E (2010) Malicious Interface Design: Exploiting the User. Proceedings of the 19th International Conference on World Wide Web 271–280. doi: https://doi.org/10.1145/1772690.1772719

30. Decoster J, Claypool HM (2004) A Meta-Analysis of Priming Effects on Impression Formation Supporting a General Model of Informational Biases. Pers Soc Psychol Rev 8:2–27. doi: https://doi.org/10.1207/S15327957PSPR0801_1

31. 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. doi: https://doi.org/10.1145/3313831.3376600

32. Díaz-Rodríguez N, Del Ser J, Coeckelbergh M, López de Prado M, Herrera-Viedma E, Herrera F (2023) Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation. Information Fusion 99:101896. doi: https://doi.org/10.1016/j.inffus.2023.101896

33. Eidelman S, Crandall CS (2012) Bias in Favor of the Status Quo. Social and Personality Psychology Compass 6:270–281. doi: https://doi.org/10.1111/j.1751-9004.2012.00427.x

34. Elmas M (2023) 'Dark Patterns': Amazon Is Being Sued for Manipulating Customers. https://www.thenewdaily.com.au/finance/finance-news/2023/06/22/amazon-dark-patterns. Accessed 4 Mar 2024

35. Entman RM (2007) Framing Bias: Media in the Distribution of Power. Journal of Communication 57:163–173. doi: https://doi.org/10.1111/j.1460-2466.2006.00336.x

36. Feng J, Mo F, Yada Y, Matsumoto T, Fukushima, N, Kido, F, Yamana, H (2023) Analysis of Dark Pattern-Related Tweets From 2010. 2023 IEEE 8th International Conference on Big Data Analytics (ICBDA) 100–106. doi: https://doi.org/10.1109/ICBDA57405.2023.10104855

37. FTC (2019) Equifax Data Breach Settlement. In: Federal Trade Commission. https://www.ftc.gov/enforcement/refunds/equifax-data-breach-settlement. Accessed 23 Apr 2024

38. Furnham A, Boo HC (2011) A Literature Review of the Anchoring Effect. The Journal of Socio-Economics 40:35–42. doi: https://doi.org/10.1016/j.socec.2010.10.008

39. Garg S, Gupta S, Gupta B (2022) Issues and Challenges With Fake Reviews in Digital Marketing. In: 2022 International Conference on Computer Communication and Informatics (ICCCI). IEEE, Coimbatore, India, pp 1–5

40. Garun N (2018) Facebook Will Add an "Unsend" Feature After Secretly Deleting Zuckerberg's Messages. In: The Verge. https://www.theverge.com/2018/4/6/17206394/facebook-unsend-messages-feature-update-messenger-mark-zuckerberg. Accessed 23 Apr 2024

41. Golbeck J, Rothstein M (2008) Linking Social Networks on the Web with FOAF: A Semantic Web Case Study. Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence 8

42. Graßl P, Schraffenberger H, Borgesius, FZ, Buijzen, M (2021) Dark and Bright Patterns in Cookie Consent Requests. Journal of Digital Social Research 3:1–38

43. Gray C, Chen J, Chivukula S, Qu L (2021) End user accounts of dark patterns as felt manipulation. Proceedings of the ACM on Human-Computer Interaction 5:1–25. doi: https://doi.org/10.1145/3479516

44. Gray C, Santos C, Bielova N (2023) Towards a Preliminary Ontology of Dark Patterns Knowledge. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems 1–9. doi: https://doi.org/10.1145/3544549.3585676

45. Gray CM, Chivukula SS, Lee A (2020) What Kind of Work Do "Asshole Designers" Create? Describing Properties of Ethical Concern on Reddit. In: Proceedings of the 2020 ACM Designing Interactive Systems Conference. Association for Computing Machinery, New York, NY, USA, pp 61–73

46. Gray CM, Kou Y, Battles B, Hoggatt J, Toombs AL (2018) The Dark (Patterns) Side of UX Design. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp 1–14

47. Greenberg S, Boring S, Vermeulen J, Dostal J (2014) Dark Patterns in Proxemic Interactions: A Critical Perspective. In: Proceedings of the 2014 Conference on Designing Interactive Systems. ACM, Vancouver BC Canada, pp 523–532

48. Gunawan J, Pradeep A, Choffnes D, Hartzog, W, Wilson, C (2021) A Comparative Study of Dark Patterns Across Web and Mobile Modalities. Proceedings of the ACM on Human-Computer Interaction 5:1–29. doi: https://doi.org/10.1145/3479521

49. Gupta C (2022) The Choice Mirage: How Australian Consumers Are Being Duped Online via Dark Patterns. Australian Journal of Competition and Consumer Law 30:241–245

50. Harding S (2024) Facebook Let Netflix See User DMs, Quit Streaming To Keep Netflix Happy: Lawsuit. In: Ars Technica. https://arstechnica.com/gadgets/2024/03/netflix-ad-spend-led-to-facebook-dm-access-end-of-facebook-streaming-biz-lawsuit/. Accessed 23 Apr 2024

51. He K, Zhang X, Ren S, Sun J (2015) Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 770–778

52. Ivanova E, Borzunov G (2020) Optimization of Machine Learning Algorithm of Emotion Recognition in Terms of Human Facial Expressions. Procedia Computer Science 169:244–248. doi: https://doi.org/10.1016/j.procs.2020.02.143

53. Jarovsky L (2021) Dark Patterns, Privacy and the LGPD. Global Privacy Law Review 2:123–130. doi: https://doi.org/10.54648/gplr2021016

54. Jaiswal A, Raju A, Deb S (2020) Facial Emotion Detection Using Deep Learning. 2020 International Conference for Emerging Technology (INCET) 1–5. doi: https://doi.org/10.1109/INCET49848.2020.9154121

55. Jiang F, Lu S, Yao X, Yue X, Au W tung (2014) Up or Down? How Culture and Color Affect Judgments. Journal of Behavioral Decision Making 27:226–234. doi: https://doi.org/10.1002/bdm.1800

56. Jin Y, Jang E, Cui J, Chung J-W, Lee Y, Shin S (2023) DarkBERT: A Language Model for the Dark Side of the Internet

57. Jones C (2024) UK Class Action Targets Grindr, Alleges App Shared HIV Data. In: The Register. https://www.theregister.com/2024/04/22/class_action_suit_data_protection_uk_grindr/. Accessed 23 Apr 2024

58. Kazienko P, Adamski M (2007) AdROSA—Adaptive Personalization of Web Advertising. Information Sciences 177:2269–2295. doi: https://doi.org/10.1016/j.ins.2007.01.002

59. Kelly D, Rubin V (2024) Identifying Dark Patterns in User Account Disabling Interfaces: Content Analysis Results. Social Media + Society 10. doi: https://doi.org/10.1177/20563051231224269

60. King J, Stephan A (2021) Regulating Privacy Dark Patterns in Practice-Drawing Inspiration from the California Privacy Rights Act. Georgetown Law Technology Review 5:250–276

61. Kirkman D, Vaniea K, Woods D (2023) DarkDialogs: Automated Detection of 10 Dark Patterns on Cookie Dialogs. 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P) 847–867. doi: https://doi.org/10.1109/EuroSP57164.2023.00055

62. Kitkowska A (2023) The Hows and Whys of Dark Patterns: Categorizations and Privacy. In: Gerber N, Stöver A, Marky K (eds) Human Factors in Privacy Research. Springer

63. Klobuchar A (2022) Text - S.2992 - 117th Congress (2021-2022): American Innovation and Choice Online Act. https://www.congress.gov/bill/117th-congress/senate-bill/2992/text. Accessed 18 Apr 2024

64. Kocyigit E, Rossi A, Lenzini G (2023) Towards Assessing Features of Dark Patterns in Cookie Consent Processes. In: Bieker F, Meyer J, Pape S, Schiering I, Weich A (eds) Privacy and Identity Management. Springer Nature Switzerland, pp 165–183

65. Koh W, Seah Y (2023) Unintended Consumption: The Effects of Four E-Commerce Dark Patterns. Cleaner and Responsible Consumption 11:100145. doi: https://doi.org/10.1016/j.clrc.2023.100145

66. Krisam C, Dietmann H, Volkamer M, Kulyk, O (2021) Dark patterns in the wild: Review of cookie disclaimer designs on top 500 German websites. EuroUSEC '21: Proceedings of the 2021 European Symposium on Usable Security 1–8. doi: https://doi.org/10.1145/3481357.3481516

67. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.

68. Kurmelovs R (2022) Optus Cyber-Attack Leaves Customers Feeling 'Powerless' Over Risk of Identity Theft. The Guardian

69. Lacey C, Beattie A, Sparks T (2023) Clusters of Dark Patterns Across Popular Websites in New Zealand. International Journal of Communication

70. Lai K-H, Zha D, Zhou K, Hu X (2020) Policy-GNN: Aggregation Optimization for Graph Neural Networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, New York, NY, USA, pp 461–471

71. Leemann T, Pawelczyk M, Eberle CT, Kasneci G (2024) I Prefer Not to Say: Protecting User Consent in Models with Optional Personal Data. Proceedings of the AAAI Conference on Artificial Intelligence 38:21312–21321. doi: https://doi.org/10.1609/aaai.v38i19.30126

72. Leiser M (2022) Illuminating Manipulative Design: From "Dark Patterns" to Information Asymmetry and the Repression of Free Choice under the Unfair Commercial Practices. Loyola Consumer Law Review 34:484–528

73. Leswing K (2022) Facebook Says Apple iOS Privacy Change Will Result in $10 Billion Revenue Hit This Year. In: CNBC. https://www.cnbc.com/2022/02/02/facebook-says-apple-ios-privacy-change-will-cost-10-billion-this-year.html. Accessed 23 Apr 2024

74. Levy D, Sun Z, Amin K, Kale S, Kulesza A, Mohri M, Suresh AT (2021) Learning with User-Level Privacy. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp 12466–12479

75. Liang S, Zhang X, Ren Z, Kanoulas E (2018) Dynamic Embeddings for User Profiling in Twitter. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, London United Kingdom, pp 1764–1773

76. Lingareddy N, Schaffner B, Chetty, M (2021) Can I Delete My Account?: Dark Patterns In Account Deletion on Social Media. Position Papers of CHI'22" What Can CHI Do About Dark Patterns (2021)

77. Long M, Xu Y, Wu J, Ou Q, Nan Y (2023) Understanding Dark UI Patterns in the Mobile Ecosystem: A Case Study of Apps in China. SaTS '23: Proceedings of the 2023 ACM Workshop on Secure and Trustworthy Superapps 33–40. doi: https://doi.org/10.1145/3605762.3624431

78. Luguri J, Strahilevitz L (2021) Shining a light on dark patterns. Journal of Legal Analysis 13:43–109. doi: https://doi.org/10.1093/jla/laaa006

79. Mansur S, Salma S, Awofisayo D, Moran, K (2023) Aidui: Toward Automated Recognition of Dark Patterns in User Interfaces. 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE) 1958–1970. doi: https://doi.org/10.1109/ICSE48619.2023.00166

80. Marlin BM (2003) Modeling User Rating Profiles For Collaborative Filtering. In: Advances in Neural Information Processing Systems. MIT Press

81. Mathur A, Acar G, Friedman MJ, Lucherini E, Mayer J, Chetty M, Narayanan A (2019) Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. Proc ACM Hum-Comput Interact 3:81:1-81:32. doi: https://doi.org/10.1145/3359183

82. McGrath P (2019) App Used To Book GP Appointments Now Facing Millions in Fines for Selling Patient Data. ABC News

83. Meta (2021) The Facts on News Reports About Facebook Data. In: Meta. https://about.fb.com/news/2021/04/facts-on-news-reports-about-facebook-data/. Accessed 23 Apr 2024

84. Mhaidli AH, Schaub F (2021) Identifying Manipulative Advertising Techniques in XR Through Scenario Construction. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp 1–18

85. Michaels J (2022) Pathways to the Light: Realistic Tactics to Address Dark Patterns. Rutgers Computer &Tech LJ

86. Miehling E, Nair R, Daly E, Ramamurthy KN, Redmond R (2023) Cookie Consent Has Disparate Impact on Estimation Accuracy. In: Oh A, Neumann T, Globerson A, Saenko K, Hardt M, Levine S (eds) Advances in Neural Information Processing Systems. pp 34308–34328

87. Mildner T, Freye M, Savino G, Doyle P, Cowan, BR, Malaka, R (2023) Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. DIS '23: Proceedings of the 2023 ACM Designing Interactive Systems Conference 2362–2374. doi: https://doi.org/10.1145/3563657.3595964

88. Mills S, Whittle R, Ahmed R, Walsh T, Wessel, M (2023) Dark Patterns and Sludge Audits: An Integrated Approach. Behavioural Public Policy 1–27. doi: https://doi.org/10.1017/bpp.2023.24

89. Moens J (2024) Your Brain Waves Are Up for Sale. A New Law Wants to Change That. The New York Times

90. Montezuma L, Taubman-Bassirian T (2019) How to avoid consent fatigue. In: iapp. https://iapp.org/news/a/how-to-avoid-consent-fatigue/. Accessed 22 Apr 2024

91. Nadeau R, Cloutier E, Guay J-H (1993) New Evidence About the Existence of a Bandwagon Effect in the Opinion Formation Process. International Political Science Review 14:203–213. doi: https://doi.org/10.1177/019251219301400204

92. National Commission on Informatics and Liberty (CNIL) (2020) Shaping Choices in the Digital World

93. Newton C (2018) Facebook Gave Spotify and Netflix Access to Users' Private Messages. In: The Verge. https://www.theverge.com/2018/12/18/18147616/facebook-user-data-giveaway-nyt-apple-amazon-spotify-netflix. Accessed 23 Apr 2024

94. van Nimwegen C, Akdag A, Bergman, K (2022) Shedding light on assessing Dark Patterns: Introducing the System Darkness Scale (SDS). Proceedings of the 35th British HCI and Doctoral Consortium 2022 1–10. doi: https://doi.org/10.14236/ewic/HCI2022.7

95. van Nimwegen C, Wit J de (2022) Shopping in the Dark: Effects of Platform Choice on Dark Pattern Recognition. In: Kurosu M (ed) uman-Computer Interaction. User Experience and Behavior. Springer International Publishing, pp 462–475

96. Norwegian Consumer Council (2018) Deceived By Design: How Tech Companies Use Dark Patterns to Discourage Us From Exercising Our Rights to Privacy

97. Nousianinen K, Ortega C (2023) Dark Patterns in Law and Economics Framework. Loyola Consumer Law Review 36:90–120

98. Nouwens M, Liccardi I, Veale M, Karger D, Kagal, L (2020) Dark Patterns After the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems 1–13. doi: https://doi.org/10.1145/3313831.3376321

99. OAIC (2023) Australian Community Attitudes to Privacy Survey 2023. In: OAIC. https://www.oaic.gov.au/engage-with-us/research-and-training-resources/research/australian-community-attitudes-to-privacy-survey/australian-community-attitudes-to-privacy-survey-2023. Accessed 23 Apr 2024

100. Optus Optus Notifies Customers of Cyberattack Compromising Customer Information. http://www.optus.com.au/about/media-centre/media-releases/2022/09/optus-notifies-customers-of-cyberattack. Accessed 23 Apr 2024

101. Papadopoulos GT, Leonidis A, Antona M, Stephanidis C (2021) User Profile-Driven Large-Scale Multi-Agent Learning From Demonstration in Federated Human-Robot Collaborative Environments. In: Kurosu M (ed) Human-Computer Interaction. Technological Innovation. HCII 2022. Lecture Notes in Computer Science

102. Parliament of Australia (2023) Senate Select Committee on Supermarket Prices. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Supermarket_Prices/SupermarketPrices. Accessed 23 Apr 2024

103. Percy K (2019) 'That Is a Yes Then?': Liberal Figure Admits Election Posters Were Designed To Mimic Aec Material. ABC News

104. Porcelli L, Ficco M, Palmieri F (2023) Mitigating User Exposure to Dark Patterns in Cookie Banners Through Automated Consent. In: Gervasi O, et al. (eds) Computational Science and Its Applications – ICCSA 2023 Workshops. ICCSA 2023. Lecture Notes in Computer Science. Springer

105. Porto FD, Egberts A (2023) The collective welfare dimension of dark patterns regulation. European Law Journal. doi: https://doi.org/10.1111/eulj.12478

106. Potts L, Mahnke S (2020) Subverting the Platform Flexibility of Twitter to Spread Misinformation. In: Jones J, Trice M (eds) Platforms, Protests, and the Challenge of Networked Democracy. Springer International Publishing, Cham, pp 157–172

107. Purplebox Digital (2016) The Impact of Colours on Your Conversion Rate. In: Purplebox Digital. https://purplebox.digital/colours-affect-conversion-rate/. Accessed 23 Apr 2024

108. Pusztahelyi R (2020) Emotional AI and Its Challenges in the Viewpoint of Online Marketing. Curentul Juridic 13–31

109. Roffarello AM, Russis LD (2022) Towards Understanding the Dark Patterns That Steal Our Attention. Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems 1–7. doi: https://doi.org/10.1145/3491101.3519829

110. Roh Y, Heo G, Whang SE (2019) A Survey on Data Collection for Machine Learning: a Big Data -- AI Integration Perspective. IEEE Transactions on Knowledge and Data Engineering 33:1328–1347. doi: https://doi.org/10.1109/TKDE.2019.2946162

111. Schäfer R, Preuschoff P, Borchers J (2023) Investigating Visual Countermeasures Against Dark Patterns in User Interfaces. MuC '23: Proceedings of Mensch und Computer 2023 161–172. doi: https://doi.org/10.1145/3603555.3603563

112. Sharma J (2023) Dark Patterns in a Bright World: An Analysis of the Indian Consumer Legal Architecture. IJCLP

113. Sharma K, Zhang Y, Ferrara E, Liu Y (2021) Identifying Coordinated Accounts on Social Media through Hidden Influence and Group Behaviours. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. ACM, Virtual Event Singapore, pp 1441–1451

114. Shiloh S, Salton E, Sharabi D (2002) Individual Differences in Rational and Intuitive Thinking Styles as Predictors of Heuristic Responses and Framing Effects. Personality and Individual Differences 32:415–429. doi: https://doi.org/10.1016/S0191-8869(01)00034-4

115. Singh A, Arun A, Malhotra P, Desur P, Jain A, Chau, DH, Kumaraguru, P (2022) Erasing Labor with Labor: Dark Patterns and Lockstep Behaviors on Google Play. Proceedings of the 33rd ACM Conference on Hypertext and Social Media 186–191. doi: https://doi.org/10.1145/3511095.3536368

116. Stavrakakis I, Curley A, O'Sullivan D, Gordon D, Tierney, B (2021) A Framework of Web-Based Dark Patterns That Can Be Detected Manually or Automatically

117. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going Deeper with Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 1–9

118. Thaler RH, Sunstein CR (2021) Nudge: The Final Edition. Yale University Press

119. Tian Y, Zhou K, Pelleg D (2022) What and How long: Prediction of Mobile App Engagement. ACM Trans Inf Syst 40:1–38. doi: https://doi.org/10.1145/3464301

120. Toth M, Bielova N, Roca V (2022) On Dark Patterns and Manipulation of Website Publishers by CMPs. roceedings on Privacy Enhancing Technologies (PoPETs), 2022 2022:478–497. doi: https://doi.org/10.56553/popets-2022-0082

121. Turnbull, T (2024) Medibank Hack: Russian Sanctioned Over Australia's Worst Data Breach. BBC News

122. Tversky A, Kahneman D (1974) Judgment Under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking Under Uncertainty. Science 185:1124–1131. doi: https://doi.org/10.1126/science.185.4157.1124

123. Tversky A, Kahneman D (1981) The Framing of Decisions and the Psychology of Choice. Science 211:453–458. doi: https://doi.org/10.1126/science.7455683

124. Völkel ST, Schneegass C, Eiband M, Buschek D (2020) What Is "Intelligent" in Intelligent User Interfaces? A Meta-Analysis of 25 Years of IUI. In: Proceedings of the 25th International Conference on Intelligent User Interfaces. Association for Computing Machinery, New York, NY, USA, pp 477–487

125. Waldman A (2020) Cognitive biases, dark patterns, and the "privacy paradox." Current opinion in psychology

126. Warner M (2021) Text - S.3330 - 117th Congress (2021-2022): DETOUR Act. https://www.congress.gov/bill/117th-congress/senate-bill/3330/text. Accessed 18 Apr 2024

127. Waters J (2021) Constant Craving: How Digital Media Turned Us All Into Dopamine Addicts. The Guardian

128. Weddige K (2024) Kobold Letters. Why HTML Emails Are a Risk to Your Organization. In: Lutra Security. https://lutrasecurity.com/en/articles/kobold-letters/. Accessed 23 Apr 2024

129. Westin F, Chiasson S (2021) "It's So Difficult to Sever that Connection": The Role of FoMO in Users' Reluctant Privacy Behaviours. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp 1–15

130. Won-Woo Park (1990) A Review of Research on Groupthink: Journal of Behavioral Decision Making. Journal of Behavioral Decision Making 3:229–245. doi: https://doi.org/10.1002/bdm.3960030402

131. Wu Q, Sang Y, Wang D, Lu Z (2022) Malicious Selling Strategies in E-Commerce Livestream: A Case Study of Alibaba's Taobao and ByteDance's TikTok

132. Xiong W, Droppo J, Huang X, Seide F, Seltzer M, Stolcke A, Yu D, Zweig G (2017) Achieving Human Parity in Conversational Speech Recognition

133. Xue W, Cai Q, Zhan R, Zheng D, Jiang P, Gai K, An B (2023) RESACT: Reinforcing Long-Term Engagement in Sequential Recommendation With Residual

134. Yada Y, Feng J, Matsumoto T, Fukushima, N, Kido, F, Yamana, H (2022) Dark Patterns in E-Commerce: A Dataset and Its Baseline Evaluations. 2022 IEEE International Conference on Big Data (Big Data). doi: https://doi.org/10.1109/BigData55660.2022.10020800

135. Yada Y, Matsumoto T, Kido F, Yamana, H (2023) Why is the User Interface a Dark Pattern?: Explainable Auto-Detection and its Analysis. 2023 IEEE International Conference on Big Data (BigData). doi: https://doi.org/10.1109/BigData59044.2023.10386308

136. Yang Q, Farseev A, Filchenkov A (2021) Two-Faced Humans on Twitter and Facebook: Harvesting Social Multimedia for Human Personality Profiling. In: Proceedings of the 2021 ACM Workshop on Intelligent Cross-Data Analysis and Retrieval. pp 39–47

137. Yang S (2024) Supermarket Promotional Labels Accused of 'Confusing' and 'Potentially Misleading' Shoppers. ABC News

138. Yeung K (2019) 'Hypernudge': Big Data as a Mode of Regulation by Design. In: Beer D (ed) The Social Power of Algorithms. Routledge, pp 118–136

139. Zhong P, Wang D, Li P, Zhang C, Wang H, Miao C (2021) CARE: Commonsense-Aware Emotional Response Generation with Latent Concepts. Proceedings of the AAAI Conference on Artificial Intelligence 35:14577–14585. doi: https://doi.org/10.1609/aaai.v35i16.17713

140. Zhu XX, Wang Y, Kochupillai M, Werner M, Haberle M, Hoffmann EJ, Taubenbock H, Tuia D, Levering A, Jacobs N, Kruspe A, Abdulahhad K (2022) Geo-Information Harvesting from Social Media Data. IEEE Geoscience and Remote Sensing Magazine 10:150–180. doi: https://doi.org/10.1109/MGRS.2022.3219584

141. (2010) Competition and Consumer Act 2010

142. (2016) Spam Act 2003

143. (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)

144. (2018) Lei Geral de Proteção de Dados Pessoais (LGPD)

145. (2022) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance)

146. (2022) Consolidated text: Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (Unfair Commercial Practices Directive) (Text with EEA relevance)Text with EEA relevance

147. (2022) Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance)

148. (2023) Privacy Act 1988

149. (2023) The 2023 Shonky Award Winners Revealed. In: CHOICE. https://www.choice.com.au/shonky-awards/hall-of-shame/shonkys-2023/2023-shonky-winners. Accessed 22 Apr 2024

150. (2024) Nobody Reads Terms and Conditions: It's Official. In: Pinsent Masons. https://www.pinsentmasons.com/out-law/news/nobody-reads-terms-and-conditions-its-official. Accessed 21 Apr 2024

151. (2024) AI Act. In: European Commission | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. Accessed 23 Apr 2024

152. Deceptive Patterns - Hall of Shame. https://www.deceptive.design/hall-of-shame. Accessed 4 Apr 2024

153. New Coke: The Most Memorable Marketing Blunder Ever? https://www.coca-colacompany.com/about-us/history/new-coke-the-most-memorable-marketing-blunder-ever. Accessed 23 Apr 2024

Appendix

Table 1. IVE Deceptive Patterns Typology

Name	Level 1	Level 2	Level 3	Source
Activity Notifications	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is misled into believing a product is more popular or credible than it really is, because they were shown activity messages.				
Address Book Leeching	Information Asymmetry	Active Misleading Actions	Misleading Information	Bösch et al., 2016
Definition: The user is prompted to give a service access to their address book to connect with known contacts also on the service, but other purposes are not declared.				
Disgracing Others	Information Asymmetry	Active Misleading Actions	Misleading Information	Wu et al., 2022
Definition: The user is falsely led to believe that a competitor's product is of lesser quality.				
Fake	Information Asymmetry	Active Misleading Actions	Misleading Information	Bösch et al., 2016
Definition: The user is presented a "universally" understood graphic code but the meaning is opposite to the expected.				
Fake Exclusive Pricing	Information Asymmetry	Active Misleading Actions	Misleading Information	Wu et al., 2022
Definition: The user is convinced to purchase based on a fake, exclusive, or discounted price that was raised before the discounted price was advertised.				
Fake Scarcity	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is pressured into completing an action because they are presented with a fake indication of limited supply or popularity.				
Fake Social Proof	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is misled into believing a product is more popular or credible than it really is, because they were shown fake reviews, testimonials, or activity messages.				

Name	Level 1	Level 2	Level 3	Source
Fake Urgency	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is pressured into completing an action because they are presented with a fake time limitation.				
False Necessity	Information Asymmetry	Active Misleading Actions	Misleading Information	Kitkowska, 2023
Definition: The user is falsely informed that certain types of data are legally necessary or required for the system to function.				
Framing	Information Asymmetry	Active Misleading Actions	Misleading Information	Norwegian Consumer Council, 2018
Definition: The user is shown information that positively frames the consequences of an action, while omitting the entailed risks.				
Hidden Legalese Stipulations	Information Asymmetry	Active Misleading Actions	Misleading Information	Bösch et al., 2016
Definition: The user is misled by complicated legal jargon to accept a legally binding policy without understanding the implications.				
High-demand Messages	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is presented a message stating that a product is in high demand, implying that it will likely sell out.				
Just Between You and Us	Information Asymmetry	Active Misleading Actions	Misleading Information	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is promised that additionally provided information will remain invisible but ultimately provide a better service.				
Lie	Information Asymmetry	Active Misleading Actions	Misleading Information	Conti and Sobiesk, 2010
Definition: The user is presented with an outright lie, such as them winning a contest.				
Limited-time Messages	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is presented a message stating that a product is only available for a limited time.				

Name	Level 1	Level 2	Level 3	Source
Loss-gain Framing	Information Asymmetry	Active Misleading Actions	Misleading Information	Bongard-Blanchy et al., 2021
Definition: The user is shown information that positively frames the consequences of an action, while omitting the entailed risks.				
Low-stock Messages	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is presented a message stating that a product is in low stock, implying that it will likely sell out.				
Misrepresenting	Information Asymmetry	Active Misleading Actions	Misleading Information	Gray et al., 2020
Definition: The user is presented ambiguous and incorrect information in order to trick them.				
Misunderstood Questions	Information Asymmetry	Active Misleading Actions	Misleading Information	Conti and Sobiesk, 2010
Definition: The user is asked questions that use confusing language, such as double, triple, or quadruple negatives.				
Scarcity	Information Asymmetry	Active Misleading Actions	Misleading Information	Gray et al., 2023
Definition: The user is pressured into completing an action because they are presented with a fake indication of limited supply or popularity.				
Sophistry	Information Asymmetry	Active Misleading Actions	Misleading Information	Wu et al., 2022
Definition: The user is shown information that positively frames the consequences of an action, while omitting the entailed risks.				
Testimonials of Uncertain Origin	Information Asymmetry	Active Misleading Actions	Misleading Information	Mathur et al., 2019
Definition: The user is misled into believing a product is more popular or credible than it really is, because they were shown fake testimonials.				
Two-faced	Information Asymmetry	Active Misleading Actions	Misleading Information	Gray et al., 2020
Definition: The user is shown contradictory and conflicting information.				

Name	Level 1	Level 2	Level 3	Source
Violate	Information Asymmetry	Active Misleading Actions	Misleading Information	Bösch et al., 2016
Definition: The user is presented a privacy policy that is intentionally violated by the presenter.				
Wrong Signal	Information Asymmetry	Active Misleading Actions	Misleading Information	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is presented a "universally" understood graphic code but the meaning is opposite to the expected.				
Asymmetric Button	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is directed by button size and colour to gravitate toward options that do not align with their intentions.				
Bad Visibility	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Kitkowska, 2023
Definition: The user is offered options where desirable options (undesirable to the service) are presented with low contrast, light colours, and small fonts.				
Chameleon Strategy	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Kitkowska, 2023
Definition: The user is presented with a third-party service that mimics the style and visual appearance of the original service to make it look like a natural continuation.				
Colour	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user's attention is guided to a designer's preference by attractive colour use.				
Dead End Trails	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user is presented by seemingly endless questions ostensibly to result in a desired outcome.				
Distorting Reality	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Mhaidli and Schaub, 2021
Definition: The user is presented, via extended reality (XR) a distorted version of reality, designed to change what they see and therefore buy.				

Name	Level 1	Level 2	Level 3	Source
Fake Button	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is presented with an element that appears to be a useful button, but is actually a disguised element for causing an undesirable outcome.				
False Hierarchy	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Gray et al., 2018
Definition: The user is presented with one or more options where they are given higher visual or interactive precedence than others.				
Fuzzy Targeting	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Wu et al., 2022
Definition: The user is shown products in a way that it seems to apply to any and all target populations.				
Inconsistent Content	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is presented with an element that entices with an offer or benefit, but upon interacting the element fails to fulfill expectations.				
Induced Icon	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is presented with icons that induce following a particular path and interact with other elements that may lead to undesirable outcomes.				
Interface Interference	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Gray et al., 2018
Definition: The user is presented with an interface that privileges specific actions over others.				
Low Contrast	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user is offered options where desirable options (undesirable to the service) are presented with low contrast.				
Mask User Warning Messages	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user is prevented from viewing browser status and warning messages by the designer.				

Name	Level 1	Level 2	Level 3	Source
Misleading Experience Marketing	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Mhaidli and Schaub, 2021
Definition: The user is presented with a digital representation of a product through extended reality (XR) that purports to represent the real version, but may be manipulated to be better than reality.				
Overlapped Placement	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is shown undesirable elements that obscure or interfere with desired elements.				
Trick Question	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Mathur et al., 2019
Definition: The user is misled into taking an action, due to the presentation of confusing or misleading language.				
Twist	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Kitkowska, 2023
Definition: The user is presented with colours and symbols that misguide them.				
Undeclared Acts	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is presented with an element that appears to be a useful button, but is actually a disguised element for causing an undesirable outcome				
Visual Interference	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Mathur et al., 2019
Definition: The user expects to see information presented in a clear and predictable way on the page, but it is hidden, obscured or disguised.				
Ad Drop-down Delay	Information Asymmetry	Passive Misleading Omissions	Delaying Provision	Lacey et al., 2023
Definition: The user is presented with a delayed drop-down advertisement, leading them to accidentally click it instead of their desired action.				
Autoplay	Information Asymmetry	Passive Misleading Omissions	Delaying Provision	Roffarello and Russis, 2022
Definition: The user is shown content that automatically plays without the user's interaction.				

Name	Level 1	Level 2	Level 3	Source
Delay User's Work Effort	Information Asymmetry	Passive Misleading Omissions	Delaying Provision	Conti and Sobiesk, 2010
Definition: The user is forced to view and wait for an advertisement.				
Hidden Costs	Information Asymmetry	Passive Misleading Omissions	Delaying Provision	Brignull, 2010
Definition: The user is enticed with a low advertised price. After investing time and effort, they discover unexpected fees and charges when they reach the checkout.				
Infinite Scrolling	Information Asymmetry	Passive Misleading Omissions	Delaying Provision	Roffarello and Russis, 2022
Definition: The user can scroll the service infinitely, with new content constantly loading.				
Interactive Hooks	Information Asymmetry	Passive Misleading Omissions	Delaying Provision	Mildner et al., 2023
Definition: The user is induced to remain on the service by delayed gratification tactics.				
Pull-to-refresh	Information Asymmetry	Passive Misleading Omissions	Delaying Provision	Roffarello and Russis, 2022
Definition: The user can "pull" the interface to load more content.				
Centralize	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user's data is collected in a single centralised location to preserves links between different users.				
Comparison Obfuscation	Information Asymmetry	Passive Misleading Omissions	Hiding Information	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user struggles to compare products because features and prices are combined in a complex manner, or because essential information is hard to find.				
Disguised Data Collection	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Greenberg et al., 2014
Definition: The user's data is gathered and used to build a rich user profile, without the user's consent.				

Name	Level 1	Level 2	Level 3	Source
Hidden Information	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Gray et al., 2018
Definition: The user may have access to desirable options or content, but it is hidden.				
Immortal Accounts	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user deletes their account, but their associated data is kept.				
Intermediate Currency	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Gray et al., 2018
Definition: The user is encourage to buy virtual currency to spend on services, which hides the true cost in real money.				
Maximize	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user's data is collected, more than is needed to provide functionality.				
Preserve	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user's aggregated data can be deanonymized to recover relationships between persons.				
Price Comparison Prevention	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Brignull, 2010
Definition: The user struggles to compare products because features and prices are combined in a complex manner, or because essential information is hard to find.				
Shadow User Profiles	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user is represented in a server's database for a service they have never registered for.				
Social Brokering	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Mildner et al., 2023
Definition: The user's relationship to other parties on the service is never forgotten, despite the relationship being dissolved in reality.				

Name	Level 1	Level 2	Level 3	Source
Unintended Relationships	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Greenberg et al., 2014
Definition: The user's relationship to other parties on the service is never forgotten, despite the relationship being dissolved in reality.				
We Never Forget	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Greenberg et al., 2014
Definition: The user's relationship to other parties on the service is never forgotten, despite the relationship being dissolved in reality.				
Attention Diversion	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user's attention is strategically targeted and kept by the service.				
Attention Grabber	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user's attention is strategically targeted and kept by the service.				
Automating the User Away	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Gray et al., 2020
Definition: The user does not give consent or confirmation, but the service automatically performs tasks.				
Bad Defaults	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bösch et al., 2016
Definition: The user unknowingly accepts defaults that share more personal information than they would otherwise intend.				
Bait and Switch	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user performs an action expecting a certain result, only to have it cause a different, likely undesired result.				
Bundled Consent	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bongard-Blanchy et al., 2021
Definition: The user is automatically marked as consenting to multiple settings when consenting to only a single setting.				

Name	Level 1	Level 2	Level 3	Source
Captive Audience	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user engages in an activity that takes time and the service takes advantage of this time to begin an unsolicited action.				
Default Sharing	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user unknowingly accepts defaults that share more personal information than they would otherwise intend.				
Disguised Ad	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user mistakenly believes they are clicking on an interface element or native content, but it is actually a disguised advertisement.				
Disguised Layout	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Long et al., 2023
Definition: The user is presented with advertisements that appear as normal content.				
Display Controversial Content	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is unexpectedly presented with shocking content without their consent.				
Easy Trigger	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Long et al., 2023
Definition: The user can unintentionally trigger an action by virtue of overly sensitive interaction mechanisms.				
False Continuity	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is required to provide their email address to perform an action, which then automatically subscribes them to a newsletter.				
Forced Consent	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bongard-Blanchy et al., 2021
Definition: The user is coerced into accepting fixed legal terms in exchange for access to the service.				

Name	Level 1	Level 2	Level 3	Source
Forced Continuity	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user is automatically charged for a service after it expires.				
Forced Enrolment	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Mathur et al., 2019
Definition: The user is automatically enrolled to an undesired component when accepting a desired component.				
Forced Viewing	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is presented with news stories that are actually advertisements.				
Forced Wholesale	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Wu et al., 2022
Definition: The user is required to buy multiple units of a product as they have no choice to buy a single unit.				
Hidden Subscription	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Mathur et al., 2019
Definition: The user is charged a recurring fee under the pretence of a one-time fee or free trial.				
Hyper-sensitive Interface Elements	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is unexpectedly shown an advertisement as a result of overly large mouse rollover activation regions.				
Illusion of Control	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Norwegian Consumer Council, 2018
Definition: The user is lulled into a false sense of security regarding their privacy and is then more likely to to disclose sensitive information.				
Impenetrable Wall	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is prevented from accessing a service unless they consent to perform an undesirable action.				

Name	Level 1	Level 2	Level 3	Source
Interrupt Acts	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Long et al., 2023
Definition: The user's flow is interrupted by pop-up advertisements.				
Milk Factor	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user is forced to move through a specific work flow in order to access a service.				
Obscure	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bösch et al., 2016
Definition: The user has great difficulty or even prevented from learning how their personal data is collected, stored, or processed.				
Preselection	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user is presented preselected options that may not be in their interest to select.				
Privacy Zuckering	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user is tricked into sharing more information about themselves than they intend.				
Silent Or Invisible Behaviour	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user has additional software unknowingly installed by a service.				
Sneak into Basket	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user has items automatically added to their online shopping cart, without their knowledge.				
Spoof Content	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is presented with new stories that are actually advertisements.				

Name	Level 1	Level 2	Level 3	Source
Video / Animation / Blinking / Motion / Audio	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user's attention is attracted to advertisements by various visual and auditory distractions.				
Blaming the Individual	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is made to feel guilty about their choices.				
Confirmshaming	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Brignull, 2010
Definition: The user is emotionally manipulated into doing something that they would not otherwise have done.				
Continued Email Communication	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user is sent one or more emails after disabling an account in an attempt to convince them to reactivate.				
Countdown Timers	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mathur et al., 2019
Definition: The user is presented with a heightened sense of immediacy by a service imposing a deadline.				
Egoistic Norms	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Wu et al., 2022
Definition: The user is pressured to embrace norms promoted by a service.				
FoMO-centric Dark Design	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Westin and Chiasson, 2021
Definition: The user is emotionally manipulated to perform specific actions by a service leveraging its data collection and deep learning capabilities.				
Hyperpersonalization	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mhaidli and Schaub, 2021
Definition: The user is emotionally manipulated to perform specific actions by a service leveraging its data collection and deep learning capabilities.				

Name	Level 1	Level 2	Level 3	Source
Improving the Experience	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is encouraged to share more data by the service giving an argument that it will improve the experience.				
Inducements to Reconsider	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user is pressured to remain using a service through language, visuals, or incentives.				
Inducing Artificial Emotions	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mhaidli and Schaub, 2021
Definition: The user is presented an emotive experience via extended reality (XR) that, if positive, may bias toward a positive evaluation of the service.				
Last Minute Consent	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is pressure into providing consent when the service knows the user is in a weak position due to hurry and impatience.				
Last Minute Solutions	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user, when attempting to disable their account, is presented with options that the service has predicted will counteract the user's reasons.				
Making Personal Information Public	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Greenberg et al., 2014
Definition: The user's personal information is made publicly visible when the user enters a particular area of the service.				
Misleading Text	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Long et al., 2023
Definition: The user is emotionally manipulated into doing something that they would not otherwise have done.				
Nagging	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Gray et al., 2018
Definition: The user tries to do something, but they are persistently interrupted by requests to do something else that may not be in their best interests.				

Name	Level 1	Level 2	Level 3	Source
Playacting	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Wu et al., 2022
Definition: The user is pressured to purchase via a fabricated emotional story or sympathy.				
Pressured Selling	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mathur et al., 2019
Definition: The user is steered toward options that are more desirable to the service by high-pressure tactics such as upselling and cross-selling.				
Providing Option	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user is given an option to reactivate their account, either temporarily or indefinitely.				
Publish	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Bösch et al., 2016
Definition: The user's personal information is made publicly visible when the user enters a particular area of the service.				
Recommendations	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Roffarello and Russis, 2022
Definition: The user is algorithmically encouraged to consume recommended content, effectively trapping them into an endless supply.				
Repetitive Incentive	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is repeatedly offered incentives by the service to encourage them to share more data.				
Retaining Customers	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Wu et al., 2022
Definition: The user is incentivised to remain on the service longer as the designer is aware that this makes the user more likely to make a purchase.				
Rewards and Punishment	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Norwegian Consumer Council, 2018
Definition: The user is enticed to make certain choices over others by being rewarded for making a designer-aligned choice and punished for others.				

Name	Level 1	Level 2	Level 3	Source
Safety Blackmail	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is pressured into consenting to unnecessary sensitive data collection under the false pretence of extra security.				
Social Investment	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Roffarello and Russis, 2022
Definition: The user is captured by social metrics such as reactions, comments, followers, to "bind" them to the service.				
Social Pyramid	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Gray et al., 2018
Definition: The user is incentivised to recruit other users to the service.				
Targeting Vulnerable Consumers	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mhaidli and Schaub, 2021
Definition: The user is personally targeted by an algorithm with personal knowledge of their vulnerabilities.				
Threatening Messages	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Conti and Sobiesk, 2010
Definition: The user is prompted to perform an action as a result of receiving a threatening message.				
Toying With Emotion	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Gray et al., 2018
Definition: The user is emotionally manipulated by the service's use of design feature to take particular actions.				
Bury in Navigation Hierarchy	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is hindered from finding and using desired actions by hiding them in an unnecessarily complicated navigation hierarchy.				
Complete Obstruction	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is completely prevented from completing desired actions, such as deleting an account.				

Name	Level 1	Level 2	Level 3	Source
Contact Zuckering	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Lacey et al., 2023
Definition: The user is obstructed from finding the organisation's telephone number.				
Controlling	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2020
Definition: The user is restricting from following their own task flow and is instead explicitly directed to follow the designer's.				
Decision Uncertainty	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Mildner et al., 2023
Definition: The user is made to feel unsure about what is expected of them or what options are available.				
Deny	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Bösch et al., 2016
Definition: The user is denied control over their data.				
Ease	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Norwegian Consumer Council, 2018
Definition: The user is lead in a certain direction, usually aligned with the designer's intentions, and alternatives are a long and arduous process.				
Entrapping	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2020
Definition: The user is mislead by the design and falls into a trap that cannot be avoided or corrected.				
Forced Email Confirmation	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is required to confirm their choice to disable their account by responding to an email.				
Forced Explanation	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is required to select or write a reason for performing a desired action before the service will permit them.				

Name	Level 1	Level 2	Level 3	Source
Gamification	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2018
Definition: The user is only able to access certain aspects of a service through “grinding” or else purchase upgrades.				
Hard to Cancel	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Mathur et al., 2019
Definition: The user is given very easy options for signing up to a service, but is obstructed from cancelling.				
Hide Desired Interface Elements	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user’s desired action is placed in an obscure location to maximise advertisement view time.				
Hinder Confidential Settings	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is able to consent with a simple action, but the process of data protection is long and complicated.				
Labyrinthine Navigation	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Mildner et al., 2023
Definition: The user is presented with nested interfaces that are easy to get lost in, disabling users from choosing preferred settings.				
Make Uninstalling Difficult	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is prevented from performing a desired action, such as uninstalling.				
Missing Exit	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Long et al., 2023
Definition: The user is prevented from exiting an interface through easy means, leading them to more easily select an option preferred by the designer.				
Obfuscating Settings	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is forced to go through a deliberately long and tedious process to achieve the setting they desire.				

Name	Level 1	Level 2	Level 3	Source
Obstruction	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2018
Definition: The user is impeded from their task flow by a design that has the intent to dissuade that task flow.				
Omit Necessary Controls	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is prevented from performing desired actions by the service lacking the relevant control.				
Requiring Request	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user must submit a request for account disabling, which must then be approved.				
Restricted Options	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Ahuja and Kumar, 2022
Definition: The user is forced by the design functionality or choice architecture to choose from a set of choices that bar the most relevant, optimal, or desirable ones.				
Roach Motel	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Brignull, 2010
Definition: The user finds it easy to sign up or subscribe, but when they want to cancel they find it very hard.				
Temporary Obstruction	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is forced to take actions that are not inherently necessary to their desired action, which increases their workload.				
Typing Errors	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is presented with an advertisement instead of assistance when they make a mistake, such as mistyping a URL.				
Forced Action	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Brignull, 2010
Definition: The user wants to do something, but they are required to do something else undesirable in return.				

Name	Level 1	Level 2	Level 3	Source
Forced Endorsement	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Wu et al., 2022
Definition: The user wants to obtain a desirable reward or perk from the service, but must first perform an action desirable to the service.				
Forced Registration	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Bösch et al., 2016
Definition: The user is required to make an account and give personal information in order to access the service.				
Mandatory Form Field Entries	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Conti and Sobiesk, 2010
Definition: The user is required to enter contact information before they are allowed to accomplish the task.				
Nickling-and-diming	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Gray et al., 2020
Definition: The user is prevented from interacting with a service by an initially disguised requirement for payment.				
Pressure to Receive Marketing	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Kitkowska, 2023
Definition: The user must opt into receiving marketing in order for the service to allow them to register.				
Redirective Conditions	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Mildner et al., 2023
Definition: The user is required to overcome unnecessary obstacles before being able to achieve their goals.