Behavioural Public Policy (2026), **10**, 171–197
doi:10.1017/bpp.2023.24

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Dark patterns and sludge audits: an integrated approach

Stuart Mills[1], Richard Whittle[2], Rafi Ahmed[3], Tom Walsh[4] and Martin Wessel[5]

[1]Department of Economics, University of Leeds, Leeds, UK, [2]Y-PERN and Department of Economics, University of Leeds, Leeds, UK, [3]University of Oxford, Oxford, UK, [4]Teifi Digital, Vancouver, Canada and [5]Behavioural Insights Team, London, UK
Corresponding author: Stuart Mills, Email: s.mills1@leeds.ac.uk

(Received 30 January 2023; revised 9 June 2023; accepted 18 July 2023)

Abstract
Dark patterns are user interface design elements which harm users but benefit vendors. These harms have led to growing interest from several stakeholders, including policymakers. We develop a high-level analytical framework – the dark patterns auditing framework (DPAF) – to support policymaker efforts concerning dark patterns. There are growing links between dark patterns and the behavioural science concept of sludge. We examine both literatures, noting several worthwhile similarities and important conceptual differences. Using two 'sludge audits,' and the DPAF, we examine 14 large online services to provide a high-level review of the user experience of these services. Our approach allows policymakers to identify areas of the user 'journey' (*dark paths*) where sludge/dark patterns persist. For regulators with constrained resources, such an approach more be advantageous when planning more granular analyses. Our approach also reveals several important limitations, notably, within some of the tools for sludge auditing which we develop, such as the 'equal clicks principle.' We discuss these limitations and directions for future research.

Keywords: dark patterns; sludge; sludge audit; user experience design; behavioural audit

Introduction

Dark patterns are of growing interest to several stakeholders, from user interface designers (Brignull, 2011), to lawyers (Luguri and Strahilevitz, 2021), to behavioural scientists (Kozyreva *et al.*, 2020; Sin *et al.*, 2022), and ultimately, policymakers (Mathur *et al.*, 2019). In 2021, the European Union expressed concern at the proliferation of dark patterns in online services (Chee, 2021), as did the OECD the following year (OECD, 2022).

What are dark patterns? Brignull (2011) – a user interface and experience designer (UI and UX, respectively) who coined the term – defines dark patterns as, 'tricks used in websites and apps that make you do things you didn't mean to, like buying or

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

signing up for something.' Mathur *et al.* (2019, p. 1) offer an authoritative review of dark patterns, assessing their prevalence in online user interfaces. They define dark patterns as, 'user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions.' Kozyreva *et al.* (2020) link dark patterns with the behavioural science concept of choice architecture, and in doing so provide *two* definitions. *Firstly*, dark patterns are defined as '[U]ser interfaces employed to steer people's choices toward unintended decisions in the service of commercial interests,' (p.107). *Secondly*, as, 'manipulative and ethically questionable use[s] of persuasive online architectures' (p. 112–113).

This article builds upon attempts to integrate behavioural science into the study of dark patterns by contributing a novel dark patterns framework which incorporates the behavioural science notion of sludge and sludge auditing (Sunstein, 2022). The above definitions serve as a useful starting point for this objective. Broadly, we define dark patterns as design strategies to *influence* decision-makers, within *online spaces*, to choose options which they will find *undesirable* or otherwise *sub-optimal*, for the *benefit of the influencing party*.

We retain generality in this definition because the literature does not offer an especially specific description of *how* dark patterns ultimately influence decision-makers. For instance, Brignull (2011) describes dark patterns as, 'tricks' but does not elaborate on *how* decision-makers are tricked, or indeed, what degree of influence or coercion constitutes a trick. Similarly, Mathur *et al.* (2019) regard dark patterns as potentially coercive and deceitful; but both Mathur *et al.* (2019) and Kozyreva *et al.* (2020) describe them as *steering* decision-makers – a concept which has previously been linked to nudging (e.g., Sunstein, 2014). Further to this, Kozyreva *et al.* (2020) explicitly link dark patterns to the behavioural science concepts of choice architecture and nudging, as have several others to a lesser extent (e.g., Newall, 2022a; Sin *et al.*, 2022; Sunstein, 2019, 2022). These behavioural concepts generally reject the labels of coercion or manipulation, and instead advocate as a core ethical principle one's freedom of choice (e.g., Thaler and Sunstein, 2008; Sunstein, 2016; Schmidt and Engelen, 2020; Lades and Delaney, 2022). We initially discuss dark patterns in a broad language because while different disciplines seem to agree on what a dark pattern *is*, some perspectives on dark patterns clearly allow for more coercive techniques (e.g., lying, deceit, tricks, coercion) while others allow for less coercive – though still potentially harmful – approaches (e.g., steering, influencing; Maier and Harr, 2020).

Such an issue is not merely a problem of *language* but is reflected in common examples. Kozyreva *et al.* (2020) offer some 'categories and types' of dark patterns, based on the work of Mathur *et al.* (2019). Here, they argue *urgency, social proof, scarcity,* and *forced action* (amongst others) all represent variations on dark patterns. *Urgency* involves emphasising time constraints to a decision-maker; *social proof* informs a decision-maker of the actions of others; *scarcity* emphasises the limited availability of a particular option; and *forced action* demands a decision-maker (now in name only) perform some action.

From the perspective of behavioural science, *urgency*, *social proof*, and *scarcity* do not *force* decision-makers to choose a particular option in the sense that decision-makers retain some freedom of choice (Sunstein, 2019), though these interventions may still be ethically questionable insofar as lying or deceit (e.g., lying about scarcity)

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

are used to influence people (Sunstein, 2019; Lades and Delaney, 2022). Urgency prompts may influence decision-makers by playing on a decision-maker's present bias (O'Donoghue and Rabin, 1999, 2015; Delaney and Lades, 2017), while social proof and scarcity prompts may achieve the same result by exploiting a propensity to care about the opinions of others (Cialdini *et al.*, 1990; Sunstein, 1996; Schultz *et al.*, 2007), and loss aversion (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Ruggeri *et al.*, 2020), respectively. *Forced action* – in terms of freedom of choice – stands out as a wholly unacceptable technique in a free society and is certainly a distinctly different means of 'influencing' an individual, compared to, say, nudging (Thaler and Sunstein, 2008).

In developing our framework, we immediately draw a distinction between 'forced action' dark patterns, and dark patterns which maintain decision-maker autonomy. This is not to say that the latter promote or respect human autonomy, only that they do not explicitly *eliminate* autonomy. For instance, one of the categories of dark pattern which Gray *et al.* (2018) propose is *nagging* – designs which continuously prompt (or *nag*) a user to select a particular option. Insofar as such a technique disrespects a decision-maker's first (or *nth*) rejection of the option, this technique does not respect that decision-maker's autonomy. Nevertheless, insofar as nagging does not *force* the decision-maker to choose whatever option is being pushed, this dark pattern still allows the decision-maker to be autonomous (also see Mills, 2022). Figure 1 shows an initial matrix of dark pattern features, as well as our initial distinction between 'forceful' dark patterns and 'steering' dark patterns, which serves as a launching pad for the remainder of this article.

The framework we offer in this article (the dark patterns auditing framework, or DPAF) attempts to simplify a vast array of dark pattern techniques and examples into a set of descriptive components. These components can then be used to describe individual dark pattern *processes*, as a means of auditing how easy, or how difficult, it is for decision-makers to achieve their desired outcomes. We offer two main components – the *detour*, and the *roundabout*. Broadly, detours are a classification of techniques designed to delay or distract a decision-maker. For instance, a long series of 'are-you-sure' checks when trying to unsubscribe from an online mailing list. By contrast, roundabouts are a classification of techniques designed to tire, bore, or otherwise redirect a decision-maker when they try to achieve an outcome, for instance, poor online infrastructure which obscures the link to *begin* unsubscribing, continuously taking a user back to the homepage.

Detours delay a person from doing what they want – which could in itself be considered harmful – because the person's actions are typically unhelpful to the influencing party. Adding many easy, yet unnecessary steps, makes the process *longer* (in terms of steps), not harder. Roundabouts make it harder for a person to do what they want, again for the benefit of the influencing party. Designing online interfaces so they are unnecessarily difficult makes the process *harder* (in terms of steps), not longer. A potentially worthwhile rule-of-thumb is that detours try to change unhelpful behaviours (from the influencing party's perspective), while roundabouts try to maintain helpful behaviours (again, from the influencing party's perspective).

Detours and roundabouts are the core components of our framework. From these components, we can also derive a third component – *shortcuts*. Shortcuts allow one to

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

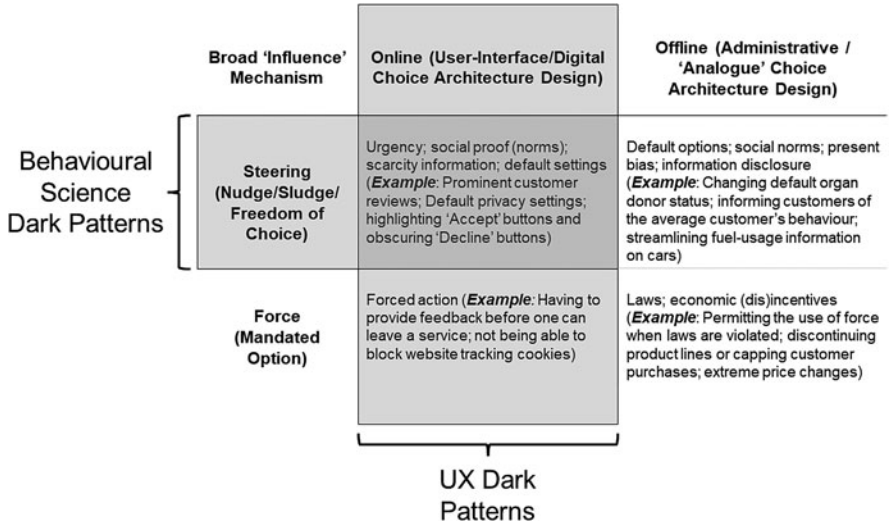


Figure 1. An initial visualisation of dark patterns, distinguishing between 'steering' patterns, and those coercive, 'forceful' patterns.

avoid other dark patterns, but often at a cost. For instance, prompts to 'select cookie preferences' (a detour) may also come with the option to 'Accept All' which skips various steps of choosing preferences, but also opts a person into excessive or otherwise unwelcome tracking of their online activities (a shortcut). Taking a shortcut is not a forced action, because a decision-maker always retains the option to not take the shortcut (e.g., manually set cookie preferences). Yet, a shortcut is a type of dark pattern insofar as it influences a user to do something they would otherwise not (e.g., pay the small fee). Finally, our framework also incorporates *forced action* for completeness. Forced action represents an additional cost a decision-maker *must* incur in order to achieve their main objective. For instance, being unable to sign up for an online service if one does not accept all terms and conditions demanded of them.

While we do not focus significantly on forced action in this article, it is important to emphasise what we do and do not consider forced action. For instance, news articles which are hidden behind paywalls would typically not constitute forced action insofar as this design is for the purpose of generating business revenue, rather than influencing decision-makers *per se*. However, mandatory terms and conditions on social media websites such as Facebook are here considered forced action because Facebook *could* tailor the availability of its service to align with the obligations that each user has individually agreed to abide by – but they do not. Likewise, while users may often receive prompts to choose their cookie settings on websites (cookies being tracking software), with users often being able to 'opt out' of 'non-essential' cookies, users rarely have the option to 'opt out' of 'essential' cookies, with essentiality even less frequently defined and explained (Mathur *et al.*, 2019). In sum, forced action within this article will be understood as actions forced upon users which go beyond the necessities for online services to function (Zuboff, 2019).

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

We combine the DPAF with a 'sludge audit' of several large, online services. Sludge audits have been proposed by Sunstein (2019, 2022) as a means of assessing and removing behavioural impediments to decision-makers – what has been dubbed *sludge* (Thaler, 2018). To our knowledge, ours is one of the first attempts to conduct and report on a sludge audit which integrates dark patterns (also see Xiao *et al.*, 2021; Behavioural Insights Team, 2022; Newall *et al.*, 2022). Sunstein (2019, 2022) suggests that sludge audits could share some elements of cost-benefit analysis (also see Sunstein and Gosset, 2020). This is because some impediments may be positive – for instance, making it harder for someone to falsely claim a government benefit. However, such impediments should not be so significant as to cause harm – for instance, making it too hard for genuine claimants to access their entitlements. We recognise the importance of this argument, though also note the challenge of such an analysis at a high level. Often, regulators and policymakers have limited resources and must utilise tools to best identify where these resources should be used. Our approach to sludge auditing, and the DPAF itself, is designed as a high-level tool for determining areas of UI design which are likely to contain dark patterns and/or sludge. As such, our sludge audit approach does not incorporate explicit cost–benefit analysis, but certainly does direct policymaker attention to potentially problematic areas where cost–benefit analysis may be worthwhile.

We draw inspiration from several notions expressed in Thaler's (2018, 2021) discussion of sludge. Specifically, we develop what we call the 'equal clicks principle' (ECP). It holds that services should be as easy to leave as they are to join. We offer this principle as a heuristic for sludge auditing online services. The ECP may not be appropriate in some circumstances – as we discuss – but is a useful rule-of-thumb for the high-level analysis we provide.

The structure of this article is as follows. We begin by reviewing the literature on dark patterns, discussing the multitude of dark pattern taxonomies which have been developed, before linking these taxonomies to the behavioural science concept of sludge. We then elaborate more on the elements contained in Table 1 in discussing our dark patterns audit framework (DPAF – or *dark paths*). Following this, we offer a proof-of-concept of the framework by undertaking a 'sludge audit' of 14 online services, including Facebook, Twitter, Amazon, and Amazon's premium service Prime. We show that many services do not accord with the 'equal clicks principle' and require more effort on the part of the user to leave than to join. Furthermore, we present 'pathway plots,' for three services – Facebook, Spotify, and eToro – which further visualises the user journey. We annotate these plots using the DPAF to demonstrate how sludge auditing can contribute to dark pattern identification. We conclude with a discussion of our approach, including the limitations of our approach, and offer some suggestions for future research.

Dark patterns and sludge: a review

Brignull (2011) coined the term 'dark patterns' to describe what he saw as a proliferation of deceptive practices in UI design as Web 2.0 services began to mature in the late 2000s and early 2010s. Brignull's (2011) 'taxonomy' is rather simple, consisting of four components, three of which are described in loose, behavioural language. These

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Table 1. Dark patterns audit framework (DPAF)

Component	Description	Example
Detour	A dark pattern designed to delay or distract a decision-maker.	Up-selling at an online retailer checkout.
Roundabout	A dark pattern designed to tire or bore a decision-maker, or otherwise redirect a decision-maker when they are trying to achieve an outcome.	A long series of 'are-you-sure' checks.
Shortcut	A dark pattern designed to exploit feelings created by detours and roundabouts to encourage decision-makers to take 'easier,' but ultimately more costly, decisions.	Salient 'Accept All' online tracking (cookie) prompts.
Forced Action	A dark pattern which forces a decision-maker to incur an additional, unexpected or undesired cost, in order to achieve their objective.	Mandatory terms and conditions requirements for new users of an online service.

are (1) the human tendency to filter out information; (2) the human tendency to stick to defaults; and (3) the human tendency to follow the example of others. In all three, Brignull (2011) notes that these strategies can be 'applied honestly' for the benefit of users, but also 'applied deceptively' for the benefit of a service, and not the users. The fourth component, which Brignull (2011) distinguishes as separate from the previous three, is *forced continuity*. Forced continuity techniques include automatically rolling-over subscriptions without telling users, exploiting a person's limited ability to remember when subscriptions auto-renew, and act.

Brignull's (2011) initial discussion of dark patterns has seen substantial elaboration in recent years. Gray *et al.* (2018) offer a five-component taxonomy, including *nagging* (persistent interactions to change user behaviour), *obstruction* (purposely making actions more difficult to encourage easier actions to be taken), *sneaking* (hiding information which if revealed would benefit the user), *interface interference* (manipulation which makes specific user actions seem more appealing than others), and *forced action* (requiring a user to take an action to receive desired functionality). Elements of this taxonomy overlap, particularly *obstruction* and *sneaking*, though Gray *et al.* (2018) elaborate on Brignull's (2011) original taxonomy with a plethora of examples (many of which come from Brignull's subsequent work). One worthwhile evolution concerns Brignull's (2011) notion of *forced continuity*. Gray *et al.* (2018) define forced continuity as a kind of *sneaking* – as auto-renewal information is purposely hidden – and discuss *forced action* as a distinctly different category.

Another 'taxonomy' comes from Bösch *et al.* (2016), who explicitly link common techniques within noted dark patterns to psychological mechanisms in their discussion of 'dark strategies.' They argue that dark patterns overwhelmingly exploit one of two psychological mechanisms. *Firstly*, dark patterns typically try to encourage users to only engage so-called System 1 thinking – the fast, intuitive mode of cognition described under dual-processing theory (e.g., Kahneman, 2003, 2011). Bösch *et al.*

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

(2016) argue that people are more likely to be susceptible to techniques such as *hidden information* and *purposeful confusion* when operating in System 1. *Secondly,* dark patterns exploit 'humans' fundamental need to belong' (Bösch *et al.*, 2016, p. 246) to influence what actions people think are acceptable and valuable, irrespective of the mode of thinking they are engaged in. In addition, Bösch *et al.* (2016) note that other psychological mechanisms, including a suite of cognitive biases such as present bias, and intentional manipulation of cognitive dissonance, can factor into dark pattern design. This leads them to elaborate on ideas such as *bad defaults* – the malign use of default options – but also more coercive strategies such as *forced registration* – the requirement to register for services which could otherwise be provided without registration.

Prior to Brignull's (2011) use of the term 'dark patterns,' Conti and Sobiesk (2010) had assembled a taxonomy of 'malicious interface design' (p. 271) techniques, consisting of 11 strategies for influencing user behaviour. These include techniques to confuse users (e.g., *Confusion, Distraction, Obfuscation*), techniques to trick users (e.g., *Trick, Shock, Manipulating Navigation, Exploiting Errors, Interruption*), and techniques to force or coerce users (e.g., *Coercion, Forced Work, Restricting Functionality*). Each technique, or category, is further developed with subcategories (for instance, *Trick* contains the subcategories *silent behaviour*, *lying*, and *spoofing*, which cover strategies such as promoting fake news articles). Conti and Sobiesk (2010) used Likert scales to measure user frustration with various subcategories, with 1 being 'No frustration,' and 7 being 'Extreme frustration.' Coercive strategies such as *coercive payment* and *forced waiting* were found to score high (5.15, and 5.89, out of 7, respectively), though not as high as manipulative – though technically not autonomy eliminating – strategies such as *unnecessary interruptions* (6.22) and *installation without permission* (6.96).

In another contribution which distinguishes this work from others, Conti and Sobiesk (2010) also measured user tolerance for dark patterns by different website categories using a sample of undergraduate students. Websites where users are typically seeking quick information, such as *weather* and *search engine* websites, saw users demonstrate the lowest tolerance (2.77, and 2.85, out of 7, respectively), while content-orientated websites, such as *shopping* (4.20), *social networking* (4.30), and *pornography* (4.39) websites produced the highest tolerance scores.

In their authoritative review of dark patterns on shopping websites, Mathur *et al.* (2019) identify seven categories of dark pattern strategies, including *sneaking*, *urgency*, *misdirection*, *social proof*, *scarcity*, *obstruction*, and *forced action*. While several of these categories have already been seen, explicitly or thematically, Mathur *et al.* (2019, p. 5–6) cross-reference these categories with five more behavioural-orientated features to produce a much richer taxonomy of dark patterns. These five features are: *asymmetric* ('Does the user interface design impose unequal weights or burdens on the available choices presented to the user in the interface?'), *covert* ('Is the effect of the user interface design choice hidden from users?'), *deceptive* ('Does the user interface design induce false beliefs either through affirmative misstatements, misleading statements, or omissions?'), *hides information* ('Does the user interface obscure or delay presentation of necessary information to the user?'), and *restrictive* ('Does the user interface restrict the set of choices available to users?'). These features

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

are said to be driven by various behavioural biases, including the anchoring effect (Tversky and Kahneman, 1974; Chapman and Johnson, 1999; Yasseri and Reher, 2022), the default effect (Madrian and Shea, 2001; Johnson and Goldstein, 2003; Jachimowicz *et al.*, 2019), and the sunk cost fallacy (Arkes and Blumer, 1985; Tversky and Kahneman, 1986). In making this formulation, Mathur *et al.* (2019) are able to talk in more detail about each of their categories. For instance, *obstruction* can sometimes merely involve *hiding information*, which is not necessarily a restrictive act. However, *obstruction* can sometimes involve explicit restriction by not making options available. For a behavioural scientist, this distinction should be considered substantial.

These taxonomies simultaneously demonstrate an array of ideas within the dark patterns literature, and consistent themes. There are also frequent links to behavioural science, though these remain to be further explored. For instance, several scholars (Brignull, 2011; Bösch *et al.*, 2016; Kozyreva *et al.*, 2020) describe dark patterns as 'nudges' or 'nudge-like,' with *darkpatterns.org*, a website established by Brignull (2011) to document instances of dark patterns in UI design, frequently using the term 'nudge' to describe dark pattern strategies. Others, such as Mathur *et al.* (2019) and Bösch *et al.* (2016) link dark patterns to dual-processing theory and System 1 thinking in particular (also see Mirsch, Lehrer and Jung, 2017; Caraban *et al.*, 2019). The OECD (2022) explicitly defines some dark patterns as based on cognitive and behavioural biases, as do Mathur *et al.* (2019) and Waldman (2020).

In recent years, behavioural scientists have likewise drawn links between dark patterns and 'nudge-like' techniques. Sunstein (2022, p. 661) describes how people can be 'nudged into harmful choices' which is later referred to as, 'a kind of dark pattern,' while Newall (2022a, p. 5) writes, 'that dark patterns can use many of the same techniques as nudges.' Both Sunstein (2022) and Newall (2022a) make these parallels within discussions of *sludge*, an emerging concept within behavioural science, and one with substantial relevance to a behavioural discussion of dark patterns (Kozyreva *et al.*, 2020; OECD, 2022; Sin *et al.*, 2022).

Sludge, like dark patterns, is a term surrounded by continuous debate, though built from established themes. Some (Soman *et al.*, 2019; Mrkva *et al.*, 2021; Sunstein, 2022) treat sludge in a way comparable to a 'bad nudge,' or the use of behavioural science and choice architectural design in a way which leaves a decision-maker worse off. Broadly, Newall (2022a, p. 6) adopts this perspective in suggesting sludge should be defined in a way, 'that mirrors nudge.' Others (Shahab and Lades, 2021) have related sludge to the transaction costs literature, suggesting sludge is a choice architecture which impedes information search and in turn leaves people worse off. Others still (Mills, 2020) have focused more so on what sludge *does*, rather than if the effect is 'good' or 'bad,' and have in turn suggested sludge makes decisions slower and harder, while nudging makes decisions faster and easier. Thaler (2018) – who is widely credited with popularising the term 'sludge' – broadly holds that sludge makes decisions harder to take and leads decision-makers to worse outcomes than they would otherwise reach.

In some respects, sludge can be understood as a behavioural science approach to dark patterns (Sin *et al.*, 2022). For instance, Mills (2020, p. 6) explicitly discusses how sludge could use 'obscurant friction' or 'social scorn and stigma' (p. 4) to

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

influence decision-makers, much in the same way the dark patterns literature frequently discusses obscurant techniques and social influence. Brignull (2011) is also compelled to argue that many of the techniques labelled as 'dark patterns,' could be used to the benefit of users, and that it is the explicit decision to use various UI designs to cause harm which leads these designs to be dark patterns. Such a perspective is also mirrored in the sludge debate, especially concerning the role of nudging. So-called 'bad nudges' or 'dark nudges' (e.g., Newall, 2019) are offered as forerunners of sludge, with Newall (2022a) explicitly letting go of the term 'dark nudge,' given the emergence of the more widely used phrase 'sludge.' Further to this debate about 'good' and 'bad,' Newall (2022a, p. 3) is critical of Mills' (2020) interpretation of sludge insofar as Mills' (2020) notion of 'all nudges produce sludges, and *vice versa*' 'effectively (mean nudge and sludge) become very similar.' Newall (2022a) is more supportive, however, of Mills' (2020) use of the terms 'Pareto' and 'rent-seeking' interventions to describe behavioural interventions which increase a decision-maker's welfare, and interventions which only benefit the intervener, respectively. Summarising their critique, Newall (2022a, p. 3) notes, 'these two new terms … may be more important than nudge or sludge.' In contrast to the above definition of a dark pattern, where the technique is used for the benefit of the UI designer at the expense of the user, these terms may also be useful in a discussion of dark patterns.

Sludge is also a useful, additional appendage to the behavioural science language concerning dark patterns insofar as it accounts for dark pattern techniques which are less 'nudge-like,' such as *obscuring* information, or *purposely confusing* or *distracting* users. It does not replace the use of nudge within the dark patterns literature, however, with this term (nudge) likely superior in describing techniques such as *social proof* and *default effects* (normative debates aside). Furthermore, the addition of sludge into this discussion does not necessitate a re-interpretation of dark patterns through a different psychological framework, as both nudge and sludge are developed within the broad heuristics and biases tradition which dark patterns scholars have already drawn upon (e.g., Mathur *et al.*, 2019).

However, it is important to engage with an emerging discussion within behavioural science, namely, the notion that dark patterns and sludge are – more or less – equivalent. Newall (2022a) suggests dark patterns may just be a UI take on sludge, as do Hallsworth and Kirkman (2020). As above, Sunstein (2022) has entertained the idea that some sludge can be described as a dark pattern, which does not go so far as to draw an equivalency, but does emphasise the overlap. Based on our interpretation of both literatures, there is clearly an overlap between the concepts. This is shown visually in Figure 1. Yet, it is likely a mistake to utilise the terms interchangeably. *Firstly*, the dark patterns literature clearly considers *force* and *forced action* to be a kind of dark pattern, whereas many behavioural science discussions would distinguish between burdens induced by choice architecture (i.e., *sludge*; Shahab and Lades, 2021; Sunstein, 2022) and explicit costs, commands, or rules. *Secondly*, dark patterns are explicitly online, UI designs, whereas sludge has been discussed as both an online and offline phenomenon. For instance, early discussions of sludge concerned voter registration burdens, excessive paperwork, and various other physical burdens (Thaler, 2018; Sunstein, 2019, 2022).

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Yet, utilising the term 'sludge,' both develops the language dark patterns and can enhance understanding. Some dark patterns are clearly designed to *change* user behaviour (e.g., *nagging*), while others are designed to *maintain* behaviours (e.g., *obscurantism*). For instance, an array of advertisements coupled with a link to a 'pay-to-skip,' option is designed change the user behaviour from watching, to paying (Lewis, 2014). By contrast, hiding a 'cancel subscription' button on a website is designed to tire or bore the user, and encourage them to *stay subscribed*. The first example can be interpreted, as many dark patterns scholars have, as a kind of nudge – the advertisements are annoying, and the pay-to-skip link is an easy option which becomes increasingly attractive as the advertisements persist. Equally, the second example can be interpreted, as many behavioural scientists would, as a kind of sludge – the decision-maker knows what they want to do, but faces barriers which erode their psychological will to do so.

Further to informing policy, while Brignull (2011) asserts that various dark pattern techniques could be used for the benefit of individuals, an essentially identical question – how can choice architecture be used to help people? – has been debated within the behavioural literature for many years (e.g., Sunstein, 2013). These two fields thus share much, both in intellectual outlook and in policy challenges.

Dark patterns auditing framework

We develop the DPAF following the various frameworks already found in the dark patterns literature. We combine this framework with a sludge audit to develop a process which can provide regulators and policymakers the means of undertaking a high-level review of online services. The purpose of our approach is not to identify specific behavioural mechanisms or to produce a comprehensive account of a service or user experience. Rather, it is to identify potentially harmful, manipulative, or otherwise welfare-reducing components of a user experience, either for internal improvement by the service itself or for use by regulators and policymakers to target action. For regulators and policymakers operating under constrained resources, the process we outline below may prove advantageous in the optimal use of these resources. We acknowledge that adopting a high-level perspective may leave some elements of the audited services to be minimised, and that our analysis will often benefit from further and more detailed data to develop coherent policy recommendations. For instance, the use of timing prompts and other time-based mechanisms (e.g., urgency, fear of missing out, present bias) is common in the dark patterns literature (Gray *et al.*, 2018; Mathur *et al.*, 2019), yet our approach does not explicitly interrogate the effects of time. In our Discussion section, we elaborate on the various limitations of the DPAF and offer some thoughts on further research.

The DPAF consists of four simple components. The first two, *detours* and *roundabouts*, are developed as a means of integrating sludge into the dark patterns lexicon (and *vice versa*). Dark pattern taxonomies often discuss instances where actions and decisions are purposely made more difficult (e.g., obscuring information; Bösch *et al.*, 2016; Mathur *et al.*, 2019) or generally slower (e.g., are-you-sure-checks, nagging; Gray *et al.*, 2018; OECD, 2022). These dark patterns would represent sludge in the behavioural literature (Thaler, 2018; Mills, 2020; Shahab and Lades, 2021;

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Newall, 2022a; Sunstein, 2022). Yet, the dark patterns literature suggests that some 'sludge-like' techniques are designed to simply make decisions harder or push users towards slightly different outcomes (e.g., suspending a subscription, rather than cancelling it), while others are more concerned with maintaining a status quo behaviour.

We distinguish between dark patterns which *change* behaviour, and those which *maintain* behaviour, in our terms *detour* and *roundabout*. A detour may delay a user from reaching their final preferred outcome by changing their behaviour. For instance, a user may still eventually cancel their subscription even if it is suspended, but *while* it is suspended, there remains the option that they re-activate the subscription, say by simply logging in (as is the case with various social media services audited below). Equally, the use of defaults may not prevent a user from eventually unsubscribing from a mailing list but may temporarily change their behaviour into being subscribed (as is frequently the case with terms of service and cookies in various services audited below). A roundabout may place so many barriers to success in the way of a user that they fail to make substantial progress, giving up through boredom or exhaustion. For instance, some services surveyed do not appear to have 'delete account' functions on their landing pages, instead only offering 'temporarily deactivate account,' functions. To find the latter, one must use functions outside of the service (e.g., a search engine). Without such action, a user will be stuck going back and forth between the same web pages. In the offline setting, Sunstein (2022) and Thaler (2018) have identified several examples of what we would call 'roundabouts' with strategies such as postal returns and unnecessary bureaucratic criteria which produce obfuscation in processes.

Both detours and roundabouts create the conditions for the third component of the DPAF, these being *shortcuts*. Shortcuts are dark patterns that make actions and decisions easier, but only typically by bypassing sludge which was not necessary to begin with. In this sense, shortcuts resemble *nudges* – a notion which, as above, the dark patterns literature has previously integrated into various taxonomies. For instance, many services audited below used detours and roundabouts to make leaving the service more difficult, but all continuously offered options to easily 'go back' or 'cancel' the decision to delete the account. Shortcuts often also occur as default options surrounding online features such as tracking cookies, mailing lists, and personalised advertisements, with an easy default given (e.g., 'Yes, allow cookies') in contrast to a more onerous 'Options' section where a user can select their preferences. As with deleting one's account, these 'Options' sections will almost always feature easy and salient 'Accept All' shortcuts (Mathur *et al.*, 2019; OECD, 2022). A final common example identified in our audit, below, is 'alternative sign-up' mechanisms, where one can use a pre-existing account (typically a Facebook, Twitter, or Google account) to quickly sign up for a new service. This shortcut – which avoids having to manually complete various details which the other services already have – is an important feature of the current internet landscape (Andersson-Schwarz, 2017).

Finally, as in the introduction, various dark pattern taxonomies discuss strategies such as 'forced action,' which cannot easily be aligned to a behavioural framework. These more coercive techniques – such as requiring users to verify their accounts, to give feedback when leaving, and to accept terms and conditions – are nevertheless

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

features of modern, online services, and belong in any taxonomy of dark patterns. Likewise, forced action strategies add to the overall burden a person faces when using a service, and therefore might be said to contribute to the cumulative 'sludginess' of the process, but do not speak to any specific psychological technique quite as, say, a default option nudge would. We distinguish between forced action and roundabouts insofar as a user will always have some way of avoiding an action when on a roundabout, be it via a deceptive shortcut, or via external assistance (such as using a search engine to find information which the service is currently hiding). By definition, forced action cannot be avoided by the user.

Method

We investigate the ease of activating and deleting user accounts for 14 large, online services. In doing so, we undertake a sludge audit, collecting both quantitative and qualitative data (see Supplementary Material), and apply the DPAF to demonstrate the utility of the approach. We focus on services noted as being particularly problematic on Brignull's dark pattern forum, *darkpatterns.org*, and that are readily available in the UK.

All data were collected by an author via email sign-up from 28 October 2022 to 7 November 2022 using a throwaway email account. A second audit of the same services was undertaken by a different researcher from 29 May 2023 to 2 June 2023. The Mozilla Firefox browser was used, with cookie-tracking turned off, as well as AdBlocker Ultimate, a Firefox recommended advertisements blocker. All web histories, cookies, and caches were cleared prior to beginning the audit. These measures were taken to ensure previous user activities (e.g., previously accepted cookies) did not interfere with the audits in any way.

There are numerous benefits of undertaking two audits. *Firstly*, the two audits broadly align, both quantitatively in terms of 'clicks' (see Table 2) and qualitatively in terms of details recorded by the auditors (see Supplementary Materials), adding to our confidence about the data. *Secondly*, the findings suggest limited changes in website functionality over time, which is worthwhile investigating given how commonly the websites of large technology companies can change.

While some previous studies have utilised pre-audit checklist approaches to collect data (e.g., Behavioural Insights Team, 2022), we adopted an auditor-directed approach. The benefits of this approach were two-fold. *Firstly*, given the variety of services being audited, we deemed it unlikely that a standardised checklist would be adequate for all services being examined. Previous audits of gambling services have been able to use more standardised approaches because of the homogeneity of the service, and because of the strict regulation which may serve as a pre-audit guide. *Secondly*, by giving auditors the freedom to choose what they documented, our data better approximate 'naïve' user experiences. While this is unlikely to be a perfect approximation – as researchers already have prior knowledge of behavioural theory and are already aware of these services – it enables an experiential dimension within our data collection, shown in the collection of qualitative data, which may better capture typical user experiences (e.g., exasperations) compared to standardised, checklist approaches.

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Table 2. Online services and the 'clicks ratio'

		Audit (1) (October/November 2022)			Audit (2) (May/June 2023)		
Service[a]	Sector	Activation 'Clicks' (A)	Deletion 'Clicks' (B)	Clicks Ratio (A/B)	Activation 'Clicks' (A)	Deletion 'Clicks' (B)	Clicks Ratio (A/B)
Facebook[b]	Social Media	5	9	0.56	5	24	0.21
Instagram[b]	Social Media	5	14	0.36	5	10	0.50
Twitter	Social Media	17	8	2.13	16	8	2.00
Reddit	Social Media	8	5	1.60	10	6	1.67
TikTok	Social Media	7	7	1.00	9	7	1.29
LinkedIn	Social Media	15	7	2.14	15	6	2.50
Amazon[c]	e-Commerce	5	10	0.50	6	7	0.86
Amazon Prime[c,d]	e-Commerce	4	6	0.67	n/a	n/a	n/a
eBay	e-Commerce	3	11	0.27	3	11	0.27
Netflix	Streaming	6	4	1.50	5	5	1.00
Twitch[c]	Streaming	5	10	0.50	5	7	0.71
Spotify	Streaming	4	19	0.21	3	9	0.33
Trading212	Finance	5	4	1.25	5	4	1.25
eToro	Finance	5	9	0.56	5	9	0.56

[a]Several services have premium versions (e.g., Spotify, LinkedIn) or other associated costs (e.g., Amazon, eToro). All services audited use free versions of the service, or temporary free trial periods.
[b]Meta Inc.
[c]Amazon Inc.
[d]Auditor 2 unable to examine due to difficulties in verifying a throwaway account.

All data collected within the audit reports are available in the Supplementary Material.

We focus on activating and deleting user accounts for several reasons. *Firstly*, this is the most basic user interaction with all of the services examined. All services examined require a user account to gain full-service functionality. *Secondly*, joining and leaving a service is a behavioural interaction which is comparable across sectors, and is frequently used as an example in the sludge literature (e.g., unsubscribing from subscriptions). *Thirdly*, it does not require any researchers to have sector-specific knowledge of the service. For instance, without knowledge of various trading platforms, it may be difficult to identify the prevalence of dark patterns being used by a specific financial trading service. Our approach to the sludge audit and application of the DPAF comes in two stages.

Firstly, having audited the process of joining and leaving the online services, we note the number of steps, or 'clicks,' required to join and leave, respectively. This allows us to calculate what we call the 'equal clicks principle' (ECP). The ECP is offered as a normative rule-of-thumb that a 'good' service should be as easy to leave as it is to join. The ECP is advantageous for a high-level analysis such as ours because it is an easy to calculate headline metric which can be used to identify particularly anomalous or notable services. As we show, the ECP allows us to compare the 14 selected services in relation to one another in terms of ease of joining and leaving, and allows us to perform an initial review of the accessibility of these services. This is done through calculating a 'clicks ratio' for the services – the number of 'clicks' required to join the service, divided by the number of 'clicks' required to leave. A 'clicks ratio' of less than one suggests a service is harder to leave than it is to join, and is thus potentially utilising harmful or manipulative UI design.

However, the ECP is not a perfect metric. We recognise that 'equal clicks' is a normative standard which might not be appropriate for specific sectors. For instance, because of the potential costs and risks involved, a better standard for, say, the finance industry may be a 'clicks ratio' which is greater than one, rather than one itself. This would mean the service is harder to join than it is to leave, deterring irresponsible or ill-conceived financial trading. The same is likely applicable for industries such as the gambling industry (Newall, 2022b). The ECP may also hide some legitimate criticism of online services. For instance, just because a service is harder to join than it is to leave does not mean that the processes involved in leaving are legitimate – they may still be manipulative and harmful to consumers. The ECP is thus limited but remains useful for directing policymaker and regulatory attention at a high-level. We speculate on potential alternative standards and develop the 'Capped Sigmoid Principle' (CSP) as a possible alternative (see Results).

Secondly, having calculated and discussed the ECP scores for our 14 audited services, we produce what we call 'pathway plots' for three of these services – Facebook, Spotify, and eToro. Pathway plots visualise the user 'journey' throughout the audited service; specifically, their journey in first joining, and then leaving the service. This approach draws inspiration from some work within the domain of ethical web design, where individual webpages or online behaviours have been 'audited' and annotated to identify, say, the presence of unnecessary cognitive biases (Harris, 2009; de Oliveira and Carrascal, 2014; Duane, 2021). The pathway plots we present allow us to overlay

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

the DPAF to highlight specific elements of the user 'journey' where a user is likely facing substantial difficulty. While our pathway plots do not focus on identifying specific cognitive biases, these visualisations do allow us to highlight some prominent biases and behavioural techniques where appropriate, such as the use of strategic default options and loss aversion messaging. As such, our development and use of pathway plots allow us to visualise the qualitative data gathered in the sludge audit, and build upon the initial conclusions developed using the ECP to arrive at a more comprehensive picture of role of dark patterns and sludge within a given online service. Again, for policymakers and regulators operating under constrained resources, once the ECP has been used to identify noteworthy services to examine, the pathway plot can be used to identify specific elements of that service's user 'journey' to further attend to.

Results

We provide summaries and some initial results in Table 2.

'Clicks' ratio and the equal clicks principle

Figure 2 plots the number of 'clicks' required to sign up for an account on these various services against the number of 'clicks' required to delete the account. The line $y = x$ shows the 'equal clicks principle' (ECP), with all points below the line requiring more clicks to activate than delete an account, and *vice versa*.

Our results show that of the 14 services audited, most (57%) were harder to leave than they were to join (i.e., 'Click' Ratio < 1). Figure 2 shows a prominent cluster of services where it takes approximately twice as many 'clicks' to leave as it does to join. Particular offenders in this regard are Spotify and Facebook, with Meta and Amazon services generally performing poorly.

Outliers on the other extreme are the social media platforms LinkedIn and Twitter. Both provide users with relatively easy account deletion compared to the ease of joining, as both services front-load mandatory onboarding features. In the case of LinkedIn, details about employment *must* be given, while with Twitter, a user *must* select a minimum number of accounts and topics to follow. The decision of *when* to collect details also skews some of the results in Figure 2. eBay, for instance, appears much more onerous to leave than to join. However, this is because eBay allows a user to create an account with minimal details. It is only when a user attempts to engage with *any* of the service's functionality – such as deleting the account – that a user is forced to provide additional details, such as one's address. Accounting for this, a service like eBay may be closer to the ECP than it initially appears.

For financial services, it is also relatively easy to make an account with these services. However, being financial services, an initial account lacks substantial functionality owing to the absence of funds. As both services acknowledge, deleting a *funded* account may be much more onerous and costly than deleting an *unfunded* account. Yet, adding funds to these accounts also incurs additional 'clicks.' The finance sector may therefore warrant special attention in sludge and dark pattern audits, as it is debatable as to how close or how far from the 'equal clicks' line they truly lie

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

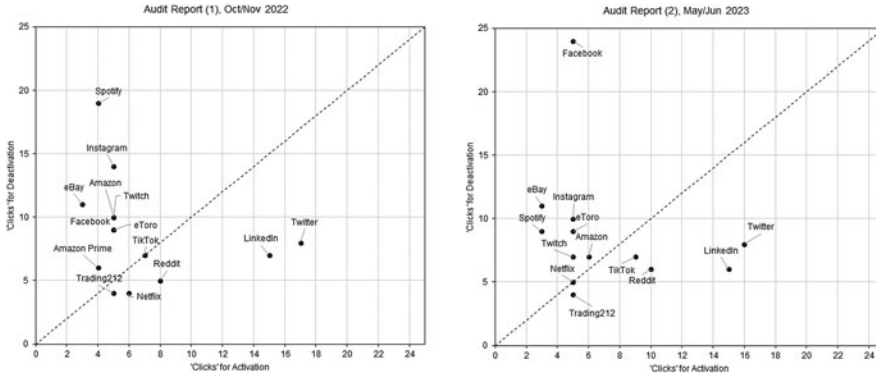


Figure 2. Plots of audited services in Terms of 'Clicks' required to join and leave said services.

(Newall and Weiss-Cohen, 2022). Further research may seek to investigate the ECP of these services and other paid services, when payment is actually incurred. Nevertheless, it is interesting to note the relative ease of setting up an unfunded account with these services. eToro enables one to sign up using their Facebook account, while Trading212 required users to select which type of financial products they wanted to trade, without providing any information within the sign-up screen about these products. Previous research into trading applications align with this result, with Newall and Weiss-Cohen (2022) reporting that eToro frequently uses deceptive disclosure practices to hide the average trading performance of their users.

Audit report (1)

In Audit Report (1), TikTok and Amazon Prime were particularly notable. TikTok was as easy to join as it was to leave, being one of the few services to prominently feature a 'Delete Account' button. However, like various other services (e.g., Instagram, Twitter), the account would not be deleted until after a cooling-off period had passed. Again, sector-specific research may be warranted to determine if such a cooling-off period promotes or undermines consumer welfare. For instance, gambling services have been found to exploit cooling-off mechanisms by processing user deposits into the platform faster than they process payments out of the platform to users (the 'delay ratio;' Newall, 2022b).

Amazon Prime is an interesting example insofar as it is quite close to the ECP. However, despite requiring relatively few 'clicks' to leave, the pages and details a user is shown as they leave are especially behaviourally laden, including invocation of loss aversion, and a plethora of opportunities to 'go back,' and not cancel. Furthermore, Amazon Prime relies on a user having an Amazon account, which is particularly onerous to delete. While one has an Amazon account, it is perilously easy to sign up for Prime, as Prime is interjected into the functionality of the Amazon service in several ways, including prominently on the home screen. As with financial services, a service such as Prime is perhaps better understood through a more in-depth audit process, rather than the high-level approach taken here.

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Audit report (2)

In terms of broad patterns, Audit Report (2) is similar to Audit Report (1). For instance, Twitter and LinkedIn remain outliers in terms of taking many clicks to join, but relatively few to leave. Likewise, TikTok, Reddit, Netflix, and Trading212 all lie reasonably close to the ECP, while Meta services, eBay, and eToro cluster above the ECP.

However, Audit Report (2) finds Facebook to be a major outlier, requiring many more clicks to leave than to join, while Spotify is reasonably similar to several other services. This is a reverse of Audit Report (1). An advantage of auditor-directed reviews is one may closely examine the causes of this discrepancy. Both auditors found themselves stuck in roundabouts which required either external help (e.g., web searching) or experimentation to escape. This adds credibility to the DPAF as a high-level framework, though also reveals that different people experience these services differently, and thus a simple 'clicks' audit is not sufficient to cast judgement on the behavioural composition of these services.

Capped sigmoid principle

Beyond sector-specific criticisms of the ECP, one might also draw a broad criticism that *generally* services should be easier to leave than they are to join, and *generally* services should not be hard to leave just because they are hard to join. With this in mind, we postulate an alternative normative standard by which to consider our data, which we dub the CSP. Figure 3 shows this projection, alongside the ECP, over our data.

The CSP follows sigmoid function $y = \frac{1}{1+e^{-x}}$ until the CSP meets the ECP. This will be at the halfway point of the sigmoid curve. Then, the CSP 'caps' at this value. As such, the CSP addresses these general criticisms, as below the 'cap,' a service is always easier to leave than it is to join, and at the cap, is always *as easy to leave* regardless of the burden of joining. Note that the sigmoid function may imply that some 'good' services require near-zero clicks to leave when a few clicks are needed to join, which may not be sensible. As such, we assume all services that require at least one click to join and at least one to leave.

From the CSP perspective, one might argue that 'good' services should find themselves on or within the area bounded by the CSP, and 'bad' services will be outside this bounded area. As Figure 3 shows, this different appraisal standard suggests Twitter and LinkedIn are actually easier to leave than they perhaps need to be, while Reddit and Netflix are only slightly more onerous than they should be. In fact, Audit Report (2) suggests Reddit is normatively optimal, in terms of ease of joining and leaving, when evaluated against the CSP. Still, the majority of services audited find themselves beyond the normative 'cliff' which the CSP creates.

Pathway plots

We now present pathway plots of the 'user experience' recorded in the process of conducting a sludge audit of these services. We plot a selection of the 14 services examined in our sludge audit to demonstrate the principle of a pathway plot, its use, and

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

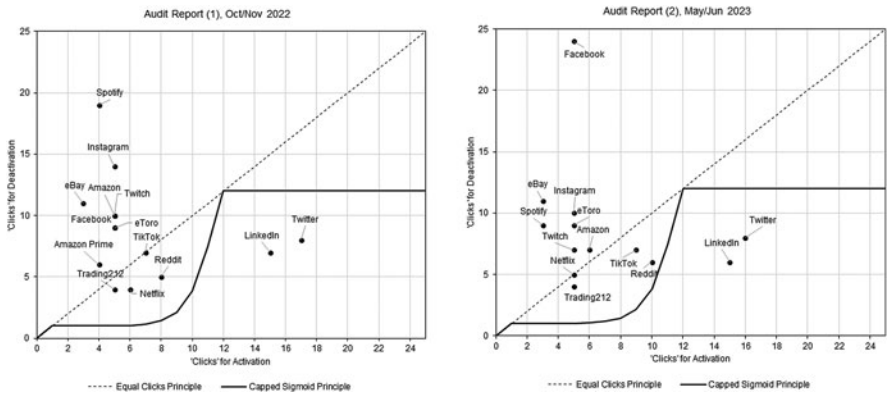


Figure 3. Capped sigmoid principle.

the integration of the DPAF within it. All plots come from data collected in Audit Report (1).

Figure 4 shows the pathway plot for Facebook. At a glance, it can be seen that it is significantly easier to activate a Facebook account than it is to delete. Yet, areas with potential dark patterns can also be seen. For instance, in the sign-up procedure, users are defaulted into accepting all cookies; users must endure a detour to avoid these cookies. In deleting the account, an example of a roundabout can be seen. Users are defaulted into only deactivating their accounts – an action which always allows reactivation – rather than deleting their accounts. Deactivation will continue to be pushed on the user unless the user can navigate successfully. In both instances, shortcuts are created. For cookie acceptance, it becomes easier to accept all cookies, while for deleting the account, it becomes easier to simply deactivate. These shortcuts may be impediments to the user's desires, but may also benefit Facebook – the broad definition of a dark pattern.

Figure 5 shows the pathway plot for Spotify. Again, it is immediately clear at a glance that this service is substantially easier to join than leave. Spotify is particularly worthwhile to examine because of their use of roundabouts. In Audit Report (1), various roundabouts were identified when a user is searching for a way to delete their account. This is likely to cause the user to frequently have to 'retrace their steps.' In our analysis, at least four roundabouts were identified. The cumulative effect of these components is likely to be that a user gives up, as shown in the outcome one receives if one follows the shortcut implicit in the roundabout.

Figure 6 is the pathway plot for the online stock trading service eToro. As previously, at a glance, it is clear that it is easier to join eToro than it is to leave, in terms of the clicks required. eToro is interesting to analyse, however, because of the concentration of forced action involved in the user experience. For instance, while cookie acceptance is a typical forced action observed within our audit (e.g., Facebook), eToro also features the sector-specific forced action of having a user consider various information from the Financial Conduct Authority (FCA). We do not highlight this to suggest such action is unreasonable, but to instead highlight that there may be

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

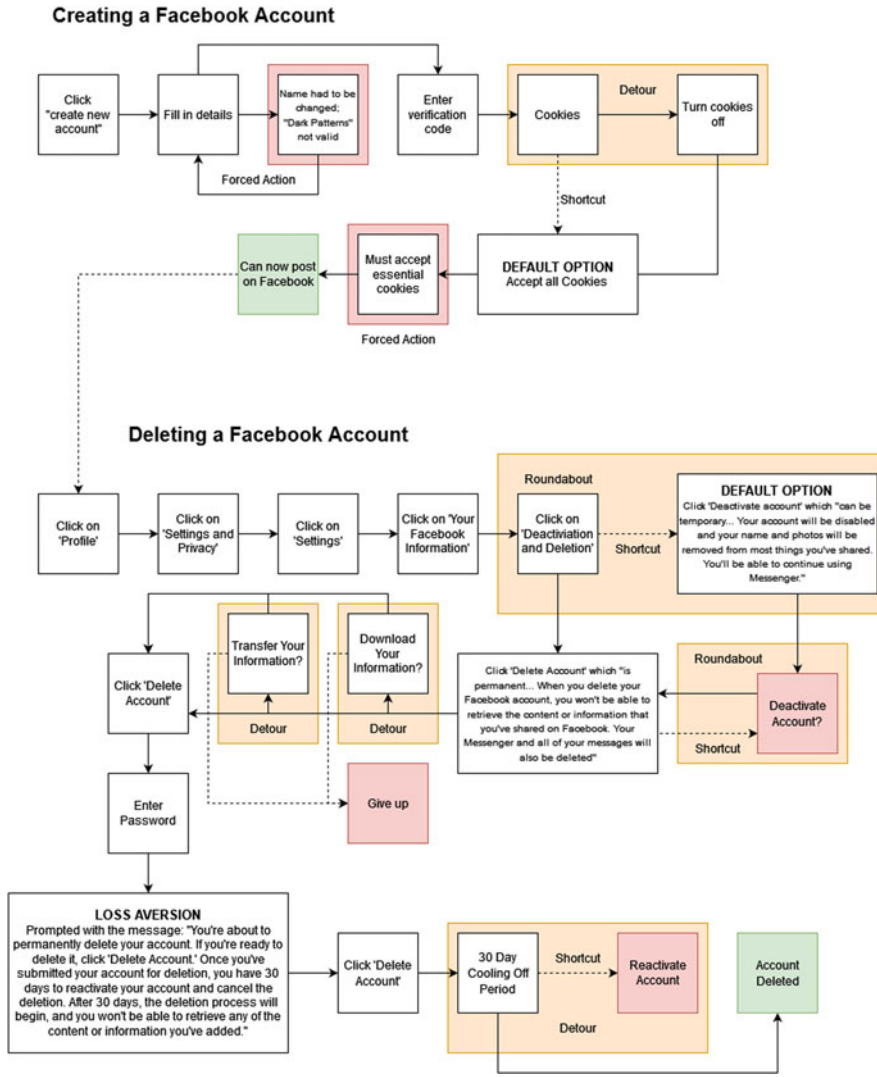


Figure 4. Facebook pathway plot.

sector-specific adjustments which are relevant when undertaking a sludge audit and using the DPAF (also see Newall and Weiss-Cohen, 2022).

While such forced action may be acceptable, or required by a regulator, the eToro pathway plot shows further forced action which seems unnecessary and may be interpreted as trying to encourage a user to abandon their efforts to delete their account. Specifically, eToro places various demands on a user to provide feedback for why they are deleting their account. Again, this is perhaps for regulatory compliance, though may also serve as a means of making the deletion process more onerous, and thus off-putting, through what we call 'cumulative sludge.'

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

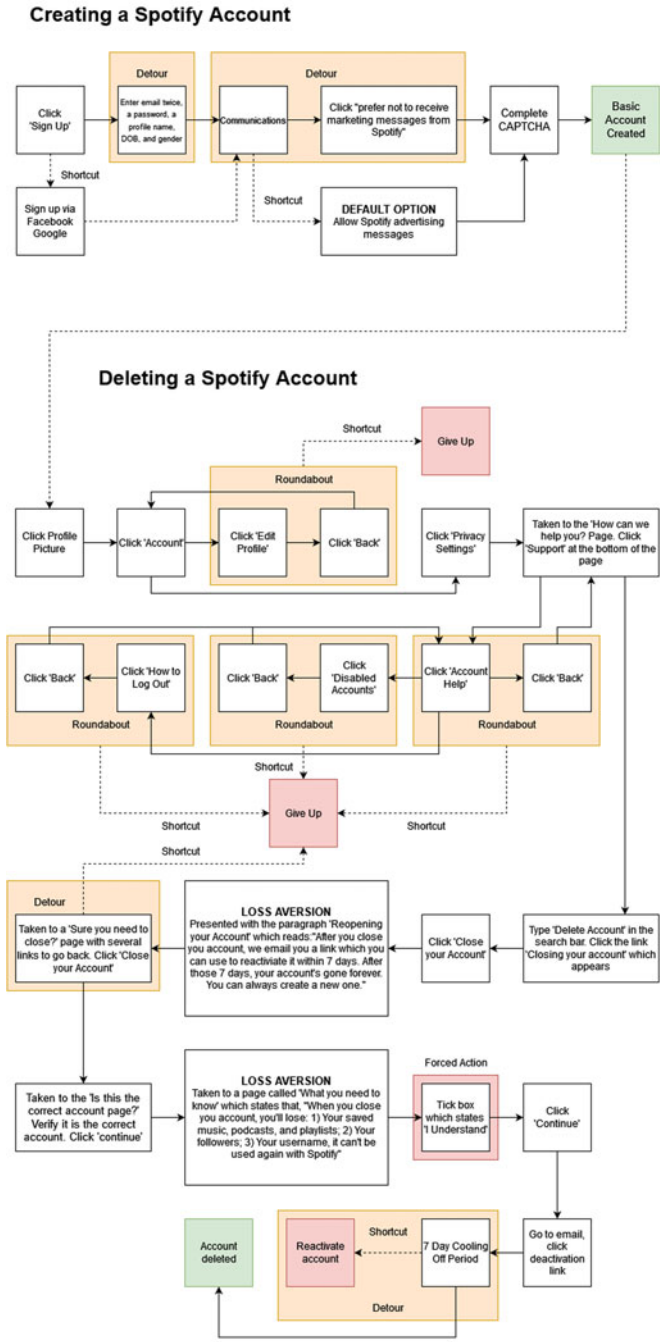


Figure 5. Spotify pathway plot.

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

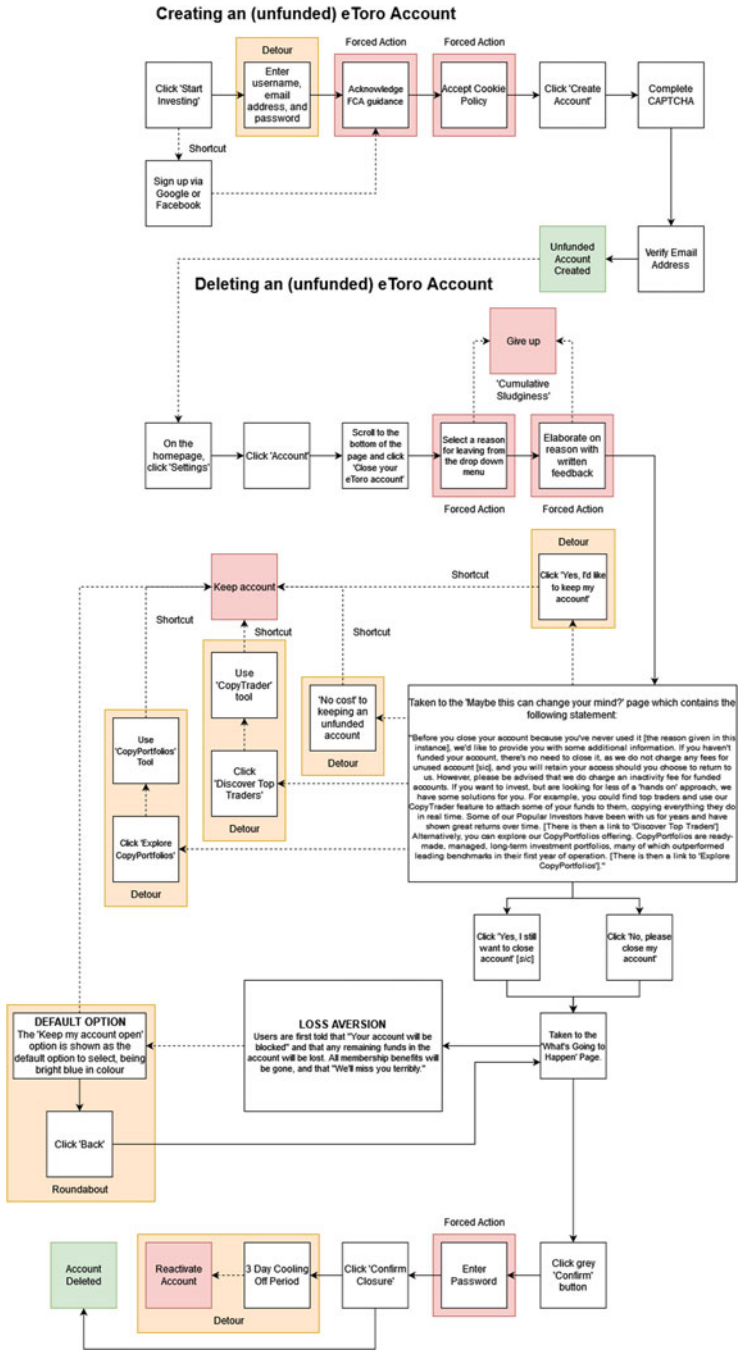


Figure 6. eToro pathway plot.

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Discussion

Our high-level audit of sludge and dark patterns in several prominent, online services, reveals two important conclusions.

Firstly, from a 'clicks,' perspective, our sludge audit suggests that popular online services are indeed easier to join than they are to leave. This aligns with earlier work (Behavioural Insights Team, 2022) for gambling services. The prominent clustering which can be seen in Figure 2 may just be result of poor website design. Yet, given many of these services benefit from retaining users, and linking individual activity to a user account, it is not unreasonable to entertain that these difficulties may in fact be by design, as various regulators have begun to discuss (Citizens Advice Bureau, 2022; FTC, 2022; OECD, 2022). Our approach to sludge auditing is consistent with growing regulatory concerns and provides a promising direction for further research. For instance, both our use of the ECP and the evident weaknesses of the ECP reveal opportunities for further development of these auditing procedures, as we have tried to do with our proposed CSP alternative.

Secondly, from a dark patterns perspective, our sludge audit process is able to integrate our novel DPAF approach to begin to identify potential areas of user vulnerability so as to direct regulatory attention. This may be particularly valuable for regulators with limited resources. For instance, Figure 5 shows that Spotify has significant navigation challenges in its UI, with deletion options being extremely difficult to find. Our process would allow a regulator to now focus attention on this aspect of the service.

Our approach is not without limitations. It is hindered both by being an early contribution to the behavioural auditing literature, and by practical limitations determining the scope of our application. While much has been said of sludge audits (Sunstein, 2022) and sludge generally (Thaler, 2018; Mills, 2020; Shahab and Lades, 2021; Newall, 2022a; Sin *et al.*, 2022; Sunstein, 2022), the process of sludge auditing would benefit from further application, criticism, and development. For instance, we note the sector-specific difficulties in sludge auditing in the financial sector in our analysis. We also note the potential discrepancies *between* auditors, though our two audit reports broadly align.

Related to these difficulties, we have been impeded by various practical limitations. The data collected in our sludge audit have largely taken the form qualitative data, and so is labour-intensive to collect. Furthermore, data collection itself is limited by geography, in two ways. *Firstly*, some services are not available in some places – for instance, the trading app Robinhood is not available in the UK. *Secondly*, various UX features are not the same across jurisdictions – for instance, General Data Protection Regulation (GDPR) causes European 'cookie' policy to be quite different from that of the US. These factors are likely to influence data collection. While some have been able to overcome some of these geographical challenges (Andrade *et al.*, 2022), different regulatory regimes will necessarily mean that audits of global services must be tempered by geographical constraints.

Another relevant aspect to consider is *how* users interact with these services. In our audit, we chose to only examine computer (desktop) experiences, but many access these services via smartphones or tablets. For instance, the video sharing service TikTok focuses on vertical video optimised for viewing on a smartphone. In the finance sector,

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

services which can be accessed via a smartphone, such as eToro and Trading212, have led to a new phrase – 'thumb traders' – to be coined. It is likely that the experience of desktop users will be very different to that of smartphone users. This may be particularly relevant in the context of, say, developing economies, where the internet is largely accessed via smartphone applications (Öhman and Aggarwal, 2020).

Related to this, many services – particularly social media services – utilise sophisticated algorithms to tailor user experiences, and thus psychologically influence user usage of the service, which likely extends to the account deletion decision. This is to say, modern online services are rarely static, but can and do change in response to user behaviour (Yeung, 2017; Morozovaite, 2021; Mills, 2022). This perspective on usage is generally absent from our analysis, but more importantly, brings into question the notion of consistent auditing of online services *at all.* Deeply embedded users may experience services very differently to naïve users, as, say, addiction behaviour within the gambling sector attests to (Newall, 2019, 2022b; Andrade *et al.*, 2022; Newall *et al.*, 2022; Newall and Weiss-Cohen, 2022; Xiao *et al.*, 2021). Thus, our analysis – generally adopting the role of a naïve user – may not be representative of the average user, or average user experience. This challenge is likely to persist in further audits (e.g., Behavioural Insights Team, 2022).

Finally, our audit was an *external* audit, meaning we are only able to collect data from publicly accessible online infrastructure. An internal audit will likely yield different, and more valuable, results. This is not to say that our sludge audit with the DPAF is without value. Often, regulators will need to begin with external audits before building a case for further, internal investigation. Furthermore, an initial external audit may create incentives for online services to be more cooperative with regulators through an internal audit, less regulation and punishment be determined via an external audit alone.

Despite these limitations, as above, we regard our contribution – both of an example sludge audit and of an integrated dark patterns framework in the form of the DPAF – as a worthwhile initial activity as the field of behavioural auditing develops through incorporating ideas such as dark patterns and sludge.

Conclusion

This article presents a novel dark patterns framework, the DPAF, which is designed to be used in conjunction with a sludge audit, a developing idea in behavioural science. We undertake two audits of several large, online services to show how sludge audits and the DPAF can combine to produce a valuable, high-level analysis approach for regulators and policymakers.

The contributions of this article are two-fold. *Firstly*, we offer a conceptual contribution in the form of the DPAF, which summarises many dark pattern taxonomies into a simple – if, perhaps, *overly simple* – four-component framework for high-level usage. *Secondly*, we present one of the first high-level sludge audits in the literature, developing two possible tools – the 'equal clicks principle' and pathway plotting – which may spur further intellectual development of sludge auditing. Our audits illustrate how the DPAF can be used.

This article also reveals various limitations of the tools we have here developed, and of the field in its current state. While our efforts develop a package of tools

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

for high-level analysis, regulators and policymakers will frequently want more detailed data and understanding. While the DPAF supports decisions about where to target limited resources, it does not itself provide details about, say, the specific behavioural biases involved in various online user experiences. Only more granular auditing procedures – which we hope to see developed – can facilitate such insight. Furthermore, the nascency of this field reveals limitations in the tools utilised. For instance, we note that while the 'equal clicks principle' may be a worthwhile heuristic and normative standard for 'good' website design *broadly*, it is not necessarily applicable across different sectors (or other stratifications, such as devices, or cultures). As above, we have little previous work to draw upon to adjust our approach, but hope that our initial efforts spur further development in this space in the future.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/bpp.2023.24.

Acknowledgements. The authors are grateful to Jens Mads Koen for supporting the undergraduate student participation in this project, and to Liam Delaney, for supporting postgraduate student participation in this project. The authors are also grateful to Cass Sunstein, Michael Hallsworth, Maximilian Kroner Dale, and Artur Mascarenhas for their helpful comments and interest.

Competing interest. The authors declare no conflict(s) of interest arising from the respective author associations and the content of this submission.

References

Andersson-Schwarz, J. (2017), 'Platform logic: an interdisciplinary approach to the platform-based economy', *Policy and Internet*, **9**(4): 374–394.

Andrade, M., S. Sharman, L. Y. Xiao and P. W. S. Newall (2022), 'Safer gambling and consumer protection failings among 40 frequently visited cryptocurrency-based online gambling operators', *Psychology of Addictive Behaviors*. doi:10.1037/adb0000885.

Arkes, H. R. and C. Blumer (1985), 'The psychology of sunk cost', *Organizational Behavior and Human Decision Processes*, **35**(1): 124–140.

Behavioural Insights Team (2022), '*Behavioural Risk Audit of Gambling Operator Platforms*', https://www.bi.team/wp-content/uploads/2022/07/Behavioural-Risk-Audit-of-Gambling-Operator-Platforms-findings-report-July-2022.pdf [4 April 2023].

Bösch, C., B. Erb, F. Kargl, H. Kopp and S. Pfatteicher (2016), 'Tales from the dark side: privacy dark strategies and privacy dark patterns', *Proceedings on Privacy Enhancing Technologies*, **4**: 237–254.

Brignull, H. (2011), '*Dark Patterns: Deception vs. Honesty in UI Design*', A List Apart. https://alistapart.com/article/dark-patterns-deception-vs.-honesty-in-ui-design/ [24 October 2022].

Caraban, A., E. Karapanos, D. Gonçalves and P. Campos (2019), '23 Ways to nudge: a review of technology-mediated nudging in human-computer interaction', *CHI, 2019*. doi: 10.1145.3290605.3300733

Chapman, G. B. and E. J. Johnson (1999), 'Anchoring, activation, and the construction of values', *Organizational Behavior and Human Decision Processes*, **79**(2): 115–153.

Chee, F. Y. (2021), '*Key EU Parliament Committee Agrees Though Position on DSA Tech Rules*', Reuters. https://www.reuters.com/markets/deals/key-eu-parliament-committee-agrees-tough-position-dsa-tech-rules-2021-12-14/#:~:text=BRUSSELS%2C%20Dec%2014%20(Reuters),forthcoming%20negotiations%20with%20EU%20countries [19 December 2022].

Cialdini, R. B., R. R. Reno and C. A. Kallgren (1990), 'A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places', *Journal of Personality and Social Psychology*, **58**(6): 1015–1026.

Citizens Advice Bureau (2022), '*Tricks of the Trade: How Online Customer Journeys Create Consumer Harm and What To Do About It*', Citizens Advice Bureau. https://www.citizensadvice.org.uk/about-us/our-work/policy/policy-research-topics/consumer-policy-research/consumer-policy-research/tricks-of-the-

trade-how-online-customer-journeys-create-consumer-harm-and-what-to-do-about-it/ [19 December 2020].
Conti, G. and E. Sobiesk (2010), 'Malicious interface design: exploiting the user', *Proceedings of the 19th International Conference on World Wide Web*, pp. 271–280. doi:10.1145/1772690.1772719
Delaney, L. and L. K. Lades (2017), 'Present bias and everyday self-control failures: a day reconstruction study', *Journal of Behavioral Decision Making*, **30**(5): 1157–1167.
De Oliveira, R. and J. P. Carrascal (2014), 'Towards effective ethical behavior design', *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, pp. 2149–2154. doi: 10.1145/2559206.2581182
Duane, J. N. (2021), 'Designing ethical nudge for E-commerce sites: a pilot study', *Academy of Management Proceedings*, **2021**(1): e.15485.
FTC (2022), '*FTC Report Shows Rise in Sophisticated Dark Patterns Designed to Trick and Trap Consumers*', Federal Trade Commission. https://www.ftc.gov/news-events/news/press-releases/2022/09/ftc-report-shows-rise-sophisticated-dark-patterns-designed-trick-trap-consumers [19 December 2022].
Gray, C. M., Y. Kou, B. Battles, J. Hoggatt and A. Toombs (2018), 'The dark (patterns) side of UX design', *CHI, 2018*. pp. 1–14.
Hallsworth, M. and E. Kirkman (2020), *Behavioral Insights*. UK: MIT Press.
Harris, J. K. (2009), 'Ethical issues in web design', *Journal of Computing Sciences*, **25**(2): 214–220.
Jachimowicz, J. M., S. Duncan, E. U. Weber and E. J. Johnson (2019), 'When and why defaults influence decisions: a meta-analysis of default effects', *Behavioural Public Policy*, **3**(2): 158–186.
Johnson, E. J. and D. Goldstein (2003), 'Do defaults save lives?', *Science*, **302**: 1338–1339.
Kahneman, D. (2003), 'Maps of bounded rationality: psychology for behavioral economics', *The American Economic Review*, **93**(5): 1449–1475.
Kahneman, D. (2011), *Thinking, Fast and Slow*, UK: Penguin Books.
Kahneman, D. and A. Tversky (1979), 'Prospect theory: an analysis of decision under risk', *Econometrica*, **47**(2): 263–291.
Kozyreva, A., S. Lewandowsky and R. Hertwig (2020), 'Citizen versus the internet: confronting digital challenges with cognitive tools', *Psychological Science in the Public Interest*, **21**(3): 103–156.
Lades, L. K. and L. Delaney (2022), 'Nudge FORGOOD', *Behavioural Public Policy*, **6**(1): 75–94.
Lewis, C. (2014), *Irresistible Apps: Motivational Design Patterns for Apps, Games, and Web-based Communities*. USA: Apress.
Luguri, J. and L. J. Strahilevitz (2021), 'Shining a light on dark patterns', *Journal of Legal Analysis*, **13**(1): 43–109.
Madrian, B. C. and D. F. Shea (2001), 'The power of suggestion: inertia in 401(k) participation and savings behavior', *The Quarterly Journal of Economics*, **116**(4): 1149–1187.
Maier, M. and R. Harr (2020), 'Dark design patterns: an end-user perspective', *Human Technology*, **16**(2): 170–199.
Mathur, A., G. Acar, M. J. Friedman, E. Lucherini, J. Mayer, M. Chetty and A. Narayanan (2019), 'Dark patterns at scale: findings from a crawl of 11 K shopping websites', *Proceeding on the ACM Human-Computer Interactions*, **2**(81): 1–32.
Mills, S. (2020), 'Nudge/sludge symmetry: on the relationship between nudge and sludge and the resulting ontological, normative and transparency implications', *Behavioural Public Policy*. doi:10.1017/bpp.2020.61.
Mills, S. (2022), 'Finding the 'nudge' in hypernudge', *Technology in Society*, **71**: e. 102117.
Mirsch, T., C. Lehrer and R. Jung (2017), '*Digital Nudging: Altering User Behavior in Digital Environments*', https://www.alexandria.unisg.ch/250315/1/Mirsch%20Lehrer%20Jung%20(2017)_Digital%20Nudging%20-%20Altering%20User%20Behavior%20in%20Digital%20Environments.pdf. [9 January 2023]
Morozovaite, V. (2021), 'Two sides of the digital advertising coin: putting hypernudging into perspective', *Market and Competition Law Review*, **5**(2): 105–145.
Mrkva, K., N. A. Posner, C. Reeck and E. J. Johnson (2021), 'Do nudges reduce disparities? Choice architecture compensates for low consumer knowledge', *Journal of Marketing*, **85**(4): 67–84.
Newall, P. W. S. (2019), 'Dark nudges in gambling', *Addiction Research and Theory*, **27**(2): 65–67.
Newall, P. W. S. (2022a), 'What is sludge? Comparing Sunstein's definition to others'', *Behavioural Public Policy*, doi:10.1017/bpp.2022.12.
Newall, P. W. S. (2022b), 'Reduce the speed and ease of online gambling in order to prevent harm', *Addiction*. doi:10.1111/add.16028.

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Newall, P. W. S. and L. Weiss-Cohen (2022), 'The gamblification of investing: how a new generation of investors is being born to lose', *International Journal of Environmental Research and Public Health*, **19**(9): 53–91.

Newall, P. W. S., L. Walasek, E. A. Ludvig and M. J. Rockloff (2022), 'Nudge versus sludge in gambling warning labels: how the effectiveness of a consumer protection measure can be undermined', *Behavioral Science and Policy*, **8**(1): 17–23.

O'Donoghue, T. and M. Rabin (1999), 'Doing it now or later', *American Economic Review*, **89**(1): 103–124.

O'Donoghue, T. and M. Rabin (2015), 'Present bias: lessons learned and to be learned', *American Economic Review*, **105**(5): 273–279.

OECD (2022), '*Dark Commercial Patterns*', OECD Digital Economy Papers no. 336. https://www.oecd-ilibrary.org/docserver/44f5e846-en.pdf?expires=1666811516&id=id&accname=guest&checksum=141DA43B1CA0CFA24217B28948A1B004 [26 October 2022].

Öhman, C. and N. Aggarwal (2020), 'What if Facebook goes down? Ethical and legal considerations for the demise of big tech', *Internet Policy Review*, **9**(3). doi:10.14763/2020.3.1488.

Ruggeri, K., S. Alí, M. L. Berge, G. Bertoldo, L. D. Bjørndal, A. Cortijos-Bernabeu, C. Davison, E. Demić, C. Esteban-Serna, M. Friedemann, S. P. Gibson, H. Jarke, R. Karakasheva, P. R. Khorrami, J. Kveder, T. L. Andersen, I. S. Lofthus, L. McGill, A. E. Nieto, J. Pérez, S. K. Quail, C. Rutherford, F. L. Tavera, N. Tomat, C. van Reyn, B. Većkalov, K. Wang, A. Yosifova, F. Papa, E. Rubaltelli, S. van der Linden and T. Folke (2020), 'Replicating patterns of prospect theory for decision under risk', *Nature Human Behaviour*, **4**: 622–633.

Schmidt, A. T. and B. Engelen (2020), 'The ethics of nudging: An overview', *Philosophy Compass*, **15**(4): e. 12658.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein and V. Griskevicius (2007), 'The constructive, destructive, and reconstructive power of social norms', *Psychological Science*, **18**(5): 429–434.

Shahab, S. and L. K. Lades (2021), 'Sludge and transaction costs', *Behavioural Public Policy*. doi:10.1017/bpp.2021.12.

Sin, R., T. Harris, S. Nilsson and T. Beck (2022), 'Dark patterns in online shopping: do they work and can nudges help mitigate impulse buying?', *Behavioural Public Policy*. doi:10.1017/bpp.2022.11.

Soman, D., D. Cowen, N. Kannan and B. Feng (2019), '*Seeing Sludge: Towards a Dashboard to Help Organizations Recognize Impedance to End-User Decisions and Action*', Behavioural Economics in Action at Rotman (BEAR) Report Series. https://www.rotman.utoronto.ca/-/media/Files/Programs-and-Areas/BEAR/White-Papers/BEARxBIOrg-Seeing-Sludge-1.pdf?la=en&hash=5CAB338A32025E08D366F4297AF4F59EABC8781D. [24 October 2022].

Sunstein, C. R. (1996), 'Social norms and social roles', *Columbia Law Review*, **96**(4): 903–968.

Sunstein, C. R. (2013), *Why Nudge? The Politics of Libertarian Paternalism*. USA: Yale University Press.

Sunstein, C. R. (2014), 'Nudging: a very short guide', *Journal of Consumer Policy*, **37**: 583–588.

Sunstein, C. R. (2016), *The Ethics of Influence: Government in the Age of Behavioral Science*. USA: Cambridge University Press.

Sunstein, C. R. (2019), 'Sludge and ordeals', *Duke Law Journal*, **68**: 1843–1883.

Sunstein, C. R. (2022), 'Sludge audits', *Behavioural Public Policy*, **6**(4): 654–673.

Sunstein, C. R. and J. L. Gosset (2020), 'Optimal sludge? The price of program integrity', *Duke Law Journal Online*, **70**: 74–90.

Thaler, R. H. (2018), 'Nudge, not sludge', *Science*, **361**(6401): 431.

Thaler, R. H. (2021), '*Nudge: the Final Edition. LSE Online Event*', London School of Economics and Political Science. https://www.youtube.com/watch?v=FEkfqQAp6wk [24 January 2022].

Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*. UK: Penguin Books.

Tversky, A. and D. Kahneman (1974), 'Judgment under uncertainty: heuristics and biases', *Science*, **185** (4157): 1124–1131.

Tversky, A. and D. Kahneman (1986), 'Rational choice and the framing of decisions', *The Journal of Business*, **59**(4): 251–278.

Tversky, A. and D. Kahneman (1992), 'Advances in prospect theory: cumulative representation of uncertainty', *Journal of Risk and Uncertainty*, **5**: 297–323.

Waldman, A. E. (2020), 'Cognitive biases, dark patterns, and the 'privacy paradox'', *Current Opinions in Psychology*, **31**: 105–109.

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press

Xiao, L. Y., L. L. Henderson, Y. Yang and P. W. S. Newall (2021), 'Gaming the system: Sub-optimal compliance with loot box probability disclosure regulations in China', *Behavioural Public Policy*. doi: 10.1017/bpp.2021.23.

Yasseri, T. and J. Reher (2022), 'Fooled by facts: quantifying anchoring bias through a large-scale experiment', *Journal of Computational Social Science*, **5**: 1001–1021.

Yeung, K. (2017), ''Hypernudge': big data as a mode of regulation by design', *Information, Communication and Society*, **20**(1): 118–136.

Zuboff, S. (2019), *The Age of Surveillance Capitalism*. London, UK: Portfolio.

Cite this article: Mills S, Whittle R, Ahmed R, Walsh T, Wessel M (2026). Dark patterns and sludge audits: an integrated approach. *Behavioural Public Policy* **10**(1): 171–197. https://doi.org/10.1017/bpp.2023.24

https://doi.org/10.1017/bpp.2023.24 Published online by Cambridge University Press