Behavioural Public Policy (2025), 1–50
doi:10.1017/bpp.2024.49

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Dark patterns and consumer vulnerability

Amit Zac[1], Yu-Chun Huang[2], Amédée von Moltke[2], Christopher Decker[2] and Ariel Ezrachi[2]

[1]University of Amsterdam, ETH Zurich, Zurich, Switzerland and [2]Centre for Competition Law and Policy, The University of Oxford, Oxford, UK
Corresponding author: Amit Zac; Email: a.zac@uva.nl

We would like to thank Ma'ayan Zac-Cohen and Hila Zac for their assistance with the design of the website. This paper forms part of a research project funded by a grant from the Leverhulme Trust.

(Received 19 December 2023; revised 28 July 2024; accepted 7 October 2024)

Abstract
Dark patterns that manipulate consumer behaviour are now a pervasive feature of digital markets. Depending on the choice architecture utilised, they can affect the perception, behaviour and purchasing patterns of online consumers. Using a novel empirical design, we find strong evidence that individuals across all groups are susceptible to dark patterns, and only weak evidence that user susceptibility is materially affected by commonly used general proxies for consumer vulnerability (such as income, educational attainment or age). Our conclusions provide empirical support for broad restrictions on the use of dark patterns, such as those contained in the EU's Digital Services Act, that protect all consumer groups. Our study also finds that added friction, in the form of required payment action following successful deployment of dark patterns, reduces their effectiveness. This insight highlights the instances in which dark patterns would be most effective – when no further action is required by the user. Consumer vulnerability is therefore more pronounced when dealing with online providers who store users' payment details and can rely on a 'single click' to complete the purchase.

Keywords: competition law; dark patterns; online behavioural manipulation; online platforms

1. Introduction

Dark patterns are online user interfaces which seek to subvert, manipulate or impair user autonomy, decision-making or choice (Narayanan *et al.*, 2020; OECD, 2022).[1]

[1]In Europe, dark patterns have been defined as interfaces that persuade users to engage in unwanted behaviours or into undesired decisions and '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'. See DSA (2022).

They are used in e-commerce, mobile apps (Geronimo et al 2020), shopping websites (Mathur *et al.*, 2019), social media networks (Mathur *et al.*, 2018) and for privacy notices (Utz *et al.*, 2019; Nouwens *et al.*, 2020). Common examples of dark patterns include misleading statements such as 'Only 1 left!' (Exploding offers), the use of questions that trick people into giving an answer they do not intend (Trick questions), interfaces which make it difficult to exit a screen, reject an offer or cancel a subscription (Roach motel), and manipulations that make users feel ashamed or guilty for not accepting or opting into a service (Confirm shaming). Table A1 in Appendix A presents other examples of dark patterns.

Dark patterns are used to exploit widely held consumer decision-making biases and heuristics. One recent study estimated that around 97% of the most popular websites and applications in Europe apply practices perceived by users as a dark pattern (Nouwens *et al.*, 2020; European Commission, 2022). This is perhaps unsurprising considering that a growing body of studies have found that dark patterns can be highly effective in achieving their aim of manipulating consumer decision-making (Gray *et al.*, 2018; Mathur *et al.*, 2019; Bongard-Blanchy *et al.*, 2021).

The impact of dark patterns extends beyond the mere manipulation of online purchasing choices. Individual autonomy can also be undermined through reductions in privacy and a decrease in decision-making agency (Zarsky, 2019; Spencer, 2020). More generally, economic and social inefficiencies can arise when citizens and consumers fail to properly exercise their preferences or are required to take costly and wasteful self-protection measures that decreases welfare.

In broad terms, 'dark patterns' can be distinguished from other benign or beneficial online nudges based on two criteria: the purpose for which the pattern is being used and the principal beneficiary. The principal purpose of online nudging (or benign 'patterns') is arguably to *assist* consumers in overcoming or mitigating decision-making biases in ways that are 'welfare enhancing' and are in their long-term interests. In contrast, the principal purpose of dark patterns is to *complicate* or *obscure* consumer decision-making in ways which directly benefits a seller or online service provider and may not benefit the consumer.[2]

Admittedly, the qualitative judgement applied in defining dark patterns and distinguishing them from more positive nudges, has led to somewhat inconsistent classifications used by scholars and regulators (Bielova *et al.*, 2023; Gray *et al.*, 2023). For this study, however, we focus on a small number of behavioural manipulations that are clearly labelled as dark patterns and used by online sellers to affect users' choices.[3]

The increased popularity of dark patterns and their distorting effects have led to growing international consensus about the need to control their use, and protect those who are most susceptible to these tactics. However, the optimal means to achieve such protection are still debated. A key unresolved policy question concerns the identity of those who warrant protection. Should new regulations aim to protect all online

[2]Assuming a counterfactual where the consumer would not have made a purchase in the absence of the dark pattern.

[3]The term 'manipulation' is used here to refer to conduct that intentionally attempts to influence a subject's behaviour by exploiting a cognitive bias or heuristic in a decision-making process.

consumers, or should they focus only on key groups deemed to be most susceptible to harmful manipulation owing to the specific and observable traits they display (e.g., age, gender, income or other detectable characteristics commonly associated with consumer vulnerability)?

Existing research does not offer a conclusive answer to this question. While some studies have found that less educated consumers are significantly more susceptible to dark patterns (Luguri and Strahilevitz, 2021), other studies have found no evidence that the ability to detect dark patterns was correlated with age, employment status or level of education (Di Geronimo *et al.*, 2020). Emerging policies in different parts of the world also appear to be adopting different positions on those in need of protection. For example, the US Federal Trade Commission has noted that dark patterns may generate greater impact on lower-income consumers or other vulnerable populations (FTC, 2022). By contrast, the recently introduced Digital Services Act (DSA) in the European Union assumes that *all* users are susceptible to dark patterns and thus prohibits them altogether.

Using a novel experimental design, we empirically test the vulnerability of different user groups to dark patterns in a real-world online setting. First, we investigate the effectiveness of different dark patterns in influencing and manipulating user decision-making. To do so, we examine whether some of the most used dark patterns have a material impact on the autonomy and decision-making of the 'average' online user. Second, we test whether dark patterns generate significant heterogeneous effects among different groups of users. We explore whether they have more material effects on specific user groups, such as those on lower incomes, with lower education attainment or the elderly. Third, we consider whether the effect of dark patterns tapers off when friction is introduced – for instance, by requiring users to insert payment details after successful deployment of the dark pattern.

To carry out the experiment, we designed a fictitious online trading platform which offered users the opportunity to purchase a financial product. We used this platform to observe how different types of commonly deployed dark patterns affected users' decisions. A critical aim of our experiment was to replicate a 'real-life' manipulation of user decision-making as far as possible, including by requesting that users who wanted to purchase the financial product make an immediate payment. As such, our experiment adds an important methodological element of realism that has been missing in previous studies of dark patterns which have either been conducted in lab settings or have not required the participant to commit to payment following the manipulation (relying on an implied commitment to pay, rather than a real and immediate financial commitment).

We find strong evidence that commonly deployed dark patterns can materially affect user decision-making. However, in contrast to some previous studies which have found that dark patterns are relatively more effective on specific user groups, we find that user susceptibility to dark patterns is largely uniform among all users regardless of income, education and to some degree age. Our empirical results reveal that all user groups are potentially susceptible to being manipulated by dark patterns.

We further find a distinct reduction in uptake when a requirement for payment is introduced, following the successful deployment of dark patterns. This suggests that dark patterns are of greatest effect when the online interface requires a 'single click' to complete the purchase or engage in action. Accordingly, we note that users' vulnerability is more pronounced when they engage with online platforms that already hold users' payment details, or when the manipulation relies on immediate action. Our results are of direct relevance to current policy debates about the scope of future prohibitions of dark patterns and the optimal means to protect consumers.

The paper comprises six sections. Section 2 surveys the existing research on online consumer vulnerability and dark patterns. Section 3 elaborates on our experiment design. Section 4 describes our data sample, including the underlying correlation structure within the personal questionnaire. In Section 5, we present the results of the main experiment, separating the findings for 'average' effects of dark patterns across all users, and the 'heterogenous' effects among specific user groups. Section 6 explores the policy implications of our findings.

2. Literature review

2.1 Dark patterns and decision-making biases

A diverse body of literature examines how consumers can be manipulated and influenced through commercial practices that seek to exploit decision-making biases or take advantage of common decision-making heuristics (Hanson and Kysar, 1999; Mathur *et al.*, 2019). Behavioural economic research in particular has shown that such biases are systematic and predictable[4] (Ariely, 2009; Dijksterhuis et al., 2005; Bar-Gill, 2012; Hanson and Kysar, 1999). In some cases, the purported aim of such practices may be to improve consumer welfare such as where a commercial entity (or a government agency) seeks to 'nudge' a consumer towards making a choice that is in their own long-term interests (Camerer *et al.*, 2003; Thaler and Sunstein, 2008). However, in the case of dark patterns, the principal objective of the manipulations is generally to further business and commercial interests by subverting consumer choice in ways that can undermine consumer welfare and wider societal interests. It is this type of harmful manipulation in the online environment that we focus on in this paper.

Some studies have sought to map the connections between dark patterns and notable cognitive biases, namely the anchoring effect, the default effect and the framing effect (Mathur *et al.*, 2019). Together with other relevant decision-making biases such as salience and hyperbolic discounting, these biases describe the psychology underpinnings of the three dark patterns that are the focus of our study: Confirm shaming, False hierarchy and Roach motel. Let us elaborate on these three patterns:

Confirm shaming affects consumer decision-making through the use of insulting or shaming language to guilt users into a particular choice, or to coerce them to accept an offer (Özdemir, 2020; Barros *et al.*, 2022). It exploits the framing effect, which is a cognitive bias where individuals tend to make different decisions based on the presentation of the same information (Tversky and Kahneman, 1981; Hanson and Kysar, 1999). Specifically, framing influences whether an opportunity is perceived positively or negatively (Waldman, 2020), thereby affecting consumer behaviour.

[4]While the term 'bias' is not settled across research fields, it is often used to refer to our cognitive abilities as human beings invariably being 'bounded' or 'limited' or suboptimal relative to some benchmark of rationality (Kahneman and Tversky, 1996).

False hierarchy affects consumer decision-making through the promotion of certain options over others. It can be viewed as a graphical version of confirm shaming, insofar as it frames choice alternatives through visual design (Gray *et al.*, 2021), with inferior/less profitable options being assigned colours that are subtler and thus blend into the interface, while preferable/more profitable options assigned bolder colours that make them stand out. False hierarchy can also be viewed as exploiting the anchoring effect (Mathur *et al.*, 2019). This bias suggests that individuals often place too much emphasis on an initial piece of information (i.e., the option they see at the first sight), which significantly influences their judgements and choices (Tversky and Kahneman, 1974; Özdemir, 2020; Waldman, 2020). Lastly, false hierarchy exploits salience, which describes a phenomenon in which human attention, is subconsciously captured by highlighted elements/messages that stand out relative to their surrounding (Özdemir, 2020).

Roach motel affects consumer decision-making by making it easier for users to enter a situation, such as subscribing to a service, but significantly harder for them to exit (Özdemir, 2020). Users are typically required to navigate through multiple pages or obscure menus to cancel accounts, decline an offer or change settings, contrasting sharply with the simplicity of account creation and service purchase (Bhoot *et al.*, 2020; Roffarello *et al.*, 2023). This manipulation takes advantage of consumers' impatience or inertia and seeks to exploit the bias which arises from hyperbolic discounting where individuals care more about their present rewards or feelings than those that could arise in the future.

2.2 Consumer vulnerability

While studies have shown that decision-making biases or heuristics are systematic and predictable, they are often not directly identifiable at the individual level. For this reason, research on consumer susceptibility to harmful manipulation has traditionally focused on observable characteristics, such as age, income, education and other demographics that have been found to be broadly associated with the tendency to exhibit such biases. The underlying assumption is that certain consumer groups (for instance, those who are elderly or in financial distress) are typically characterised by specific biases and thus are more vulnerable to harmful manipulation. Accordingly, consumer 'vulnerability' varies in its scope and intensity between consumer groups with distinct characteristics. This approach has been labelled the 'victim approach', because it is based on the idea that certain groups of people warrant specific or added protection because they are inherently weak and/or insufficiently able to fend for their own interests (Cole, 2016).[5]

However, the view that only specific groups of consumers are subject to certain biases and decision-making traits is increasingly being challenged. Based on their comprehensive survey of US panel data, Stango and Zinman (2020) find that 'biases are

[5]The framing of consumer vulnerability as a diminished capacity to understand or to maximise utility and wellbeing (Craig Smith and Cooper-Martin, 1997) resonates with conceptualizations of consumer vulnerability in market research and some parts of the behavioural literature (Baker et al., 2005).

more rule than exception' and that the median consumer exhibits 10 of 17 potential biases and 'almost everyone exhibits multiple biases'. Moreover, they find that any cross-consumer heterogeneity in biases is poorly explained by a '"kitchen sink" of other consumer characteristics' including demographic characteristics. Of particular relevance to our study is the finding that there was more bias variance *within* classical subgroups that are often used as proxies for consumer vulnerability than *across* them: for example, Stango and Zinman found that there was more bias variation with the highest-education group than between the highest- and lowest-education groups. An important implication of this survey is that all consumers display biases and are thus potentially at risk of having those biases exploited and manipulated.

Developments in the digital economy, the rise of big data analytics and the ability to individually target online users, have stimulated a debate on consumer vulnerability in online environments. Organisations like the OECD, for example, have argued that in the digital age 'vulnerability may be experienced not only by some consumers, but increasingly by most, if not all, consumers' (OECD, 2023).

Some studies argue that our senses of time and space are reshaped in the online environment, and that an (over) abundance of information makes it more likely that consumers will use heuristics to make decisions, make simplified choices and pay less attention than offline (ACM, 2020; CMA, 2020). Others argue that while consumers may believe they have more options and choice in an online environment, in fact, they ultimately seem to be choosing from a much smaller choice set online (Costa and Halpern, 2019; Stucke and Ezrachi, 2021). This is because the consumer experience is mediated by a controlled environment of tailored buttons to press, boxes to check, options to swipe and information to (skim-) read. Technological advances in data collection, processing and analytics increasingly entail a shift in temporal dynamics such that: '[i]n an age of constant "screen time," [...] an offer is always an algorithm away' (Calo, 2014).

The online environment is also characterised by stealth and personalisation, and studies have found that many online consumers are often unaware of the extent to which what they encounter online has been individually tailored to them and can be (re)adjusted to fit a 'persuasion profile' (Susser, Roessler and Nisssenbaum, 2019; CMA, 2020). Indeed, while some studies have found that consumers feel safer online than they do offline (Moran, 2020), many consumers are nevertheless unaware of the cognitive process through which their decision-making is being purposively shaped and influenced (Marchiori *et al.*, 2017; Spencer, 2020).

The new dynamics unfolding online create a significant challenge to the traditional understanding of consumer vulnerability. Indeed, some argue that in digital markets, the notion of consumer vulnerability refers to a 'universal state of defencelessness and susceptibility to (the exploitation of) power imbalances' that weigh heavily in favour of the digital choice architects (Helberger *et al.*, 2021). Calo (2014) goes further and argues that armed with data-driven, dynamically adjustable and personalised choice architectures, online commerce is designed to exploit or even create vulnerabilities. In short, while consumers may perceive that they are in control and safer, and thus less vulnerable, in an online environment, this may be an illusion created by online service providers using technological advancements to fulfil their own aims.[6]

[6]For example, how digital platforms tailor privacy settings and exit options (see CMA, 2020)).

2.3 The effectiveness and effects of dark patterns

A small but growing body of empirical research has examined the use and effects of dark patterns on consumers. Some studies have sought to develop a typology of the different types of dark patterns or focussed on the prevalence of dark patterns (Mills *et al.*, 2023). While these studies have found that the type and frequency of dark patterns can vary across websites, apps and across jurisdictions, taken as a whole they show that 'dark patterns are far from a niche practice' (OECD, 2022). A separate set of studies has focussed on the effectiveness of dark patterns (and particular types of dark patterns) in manipulating consumer decision-making. Important among these studies are Luguri and Strahilevitz (2021), Sin, Harris and Nilsson and Beck (2022) and the study by the European Commission (2022), all of which conclude that certain types of dark patterns can have substantial effects on consumer decision-making.

Of relevance to our study is the emerging evidence on whether dark patterns can have differential (heterogenous) impacts on consumers, according to common metrics such as income, age, education etc. The available research, albeit limited, offers several (at times, inconsistent) insights. Some of the studies focus on how age affects a consumer's susceptibility to dark patterns, particularly children and older consumers (European Commission, 2022). Luguri and Strahilevitz (2021) and Bongard-Blanchy *et al.* (2021) found that lower educated consumers are most susceptible to dark patterns, while the European Commission (2022) found that vulnerable consumers are more likely to make inconsistent choices than average consumers when exposed to dark patterns. In contrast, Di Geronimo *et al.* (2020) found no evidence that the ability to detect dark patterns was correlated with age, employment status or level of education.

In considering why these emerging findings may not provide consistent results, it is important to understand the methodological approaches applied in these studies and how closely they resembled a real-world choice and decision environment. Studies such as Di Geronimo *et al.* (2020) and Bongard-Blanchy *et al.* (2021) used online surveys to assess participants' capability to recognise different dark pattern types and then elicited their views on the effectiveness of different dark patterns. The European Commission (2022) study was also based around a survey but included an online experiment where participants were required to choose between two different digital entertainment service packages. If their choice was consistent with their stated preferences, they received a certain number of points.

In their influential study, Luguri and Strahilevitz (2021) used an online experiment based around a fictitious website and product. In our opinion, this approach is preferable to 'lab experiments' in which subjects are not subjected to dark patterns in a manner that reflects a real-life experience. However, while Luguri and Strahilevitz's experimental design is the closest to our own, our approaches differ in one fundamental aspect which we believe is critical to understanding the results of the two studies. Specifically, in Luguri and Strahilevitz (2021), participants were told that they had been automatically signed up to a costly identity theft protection service and they would need to pay for the service if they did not 'opt out'. Participants were told that the website had been able to pinpoint their mailing address (using their IP address and zip code) and that following 6 months of free theft protection they would be billed monthly to that address. In contrast, in our experiment participants had to consciously 'opt-in' to purchasing the service by entering their credit card or PayPal details. The offer they received did not include any 'free' subscription period and led to an immediate charge. As we explain below, we believe this difference in design is highly material to the credibility and robustness of the experiment and the results, since the use of a payment page more closely approximates reality.

2.4 Our contribution

Our research builds on and expands existing research on dark patterns and online consumer vulnerability in several ways.

First, we required participants to fill in a detailed questionnaire about their personal experiences, attitudes and preferences beyond the main demographics, such age, education and income (e.g., Bongard-Blanchy *et al.*, 2021). This information allowed us to identify specific characteristics and demographic details about participants and thus assign them to different user groups during analysis. This level of granularity is central to our assessment of whether there are heterogenous user effects of dark patterns. To justify the request for such information and to encourage participants to respond truthfully, we informed participants that this information would be used to understand underlying factors that influence people's views about a website design.

Second, to ensure a natural online environment (semi-field experiment), we created a genuine website for a (fictitious) algorithm-driven investment product building on the approach of studies such as Di Geronimo *et al.* (2020) and Bongard-Blanchy *et al.* (2021).[7] However, in our study, all the information about the product, including the offer pop-ups and the payment page, were presented on the website, separately from the survey. We believe the use of an actual product website with all the associated features made the existence of the product and the offer more realistic compared to a lab setting (as used in some of the other studies), and therefore increased the ecological validity of the experiment. In addition, by asking participants to evaluate the website and its content, we familiarised them with the product.

Third, to convince participants of the genuineness of the offer (i.e., that accepting it would entail an actual monetary commitment) and to reflect a real-life scenario, participants understood this to be a paid service from the moment the offer was made. We subsequently used an online payment process ('procced to payment') and typical industry payment page (Figure 1b). Figure 1 includes screenshots of: (a) the landing page of the website; and (b) the payment page (further screenshots of the website are presented in the Appendix C). The incorporation of an immediate payment page into our experiment differs from other experiments, particularly that of Luguri and Strahilevitz (2021), which informed participants of the experimenters' ability to pinpoint participants' billing addresses by means of postcode and IP information they provided in the survey or other non-common methods of payment. A key advantage of our approach is that participants were more likely to believe they are facing a true offer and accepting it may incur a real cost when required to insert payment details. On this point, we

[7] Di Geronimo *et al.* (2020) conduct an 'online experiment in the form of an online survey that included videos of the apps' usage'. Bongard-Blanchy *et al.* (2021) use a survey with images of dark patterns design.

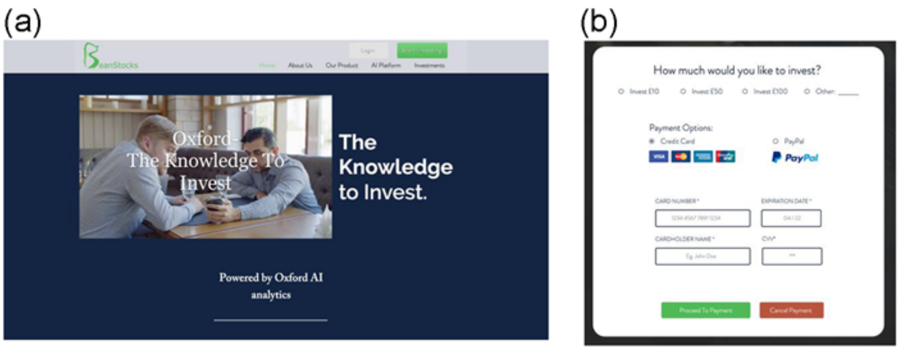


Figure 1. BeanStocks' landing and payment pages.
Note: The left panel (1a) is the landing page of the BeanStocks website. The right panel (1b) is the payment page displayed to all participants that accepted our offer.

note that in Luguri and Strahilevitz (2021) participants' sensitivity to price was very low to non-existent. This, we believe, might reflect the fact that participants did not perceive that they were facing actual costs (i.e., it was costless to accept). By designing our study in a way that enables us to measure the immediate effect of dark patterns on decisions (click to accept), and subsequently the actual commitment to pay (payment choice), we were able to go beyond the present literature by investigating whether dark patterns can significantly affect purchasing decisions when the choice to buy a product and make payment are presented sequentially.

The final way in which we believe our experiment improves on previous designs is through the use of specialised session recording software which allowed us to evaluate the effectiveness of our experiment by viewing full session recordings of the online activity on our website. To the best of our knowledge this has not been used in previous experimental work on dark patterns.

3. Experiment design and procedure

As described in Section 1, our experiment sought to test three questions: first, whether dark patterns were effective in increasing the acceptance rate of a fictitious online financial product; second, whether some identifiable groups of users were more susceptible to these manipulations than others; and third, whether the introduction of friction (in the form of a payment page) following a successful deployment of a dark pattern, reduces its effectiveness and influences users' vulnerability. Our prior, given the existing evidence, and specifically the finding in Luguri and Strahilevitz (2021) was that some demographic factors, (e.g., levels of income or educational attainment) would affect user susceptibility to dark patterns. Our empirical study was pre-registered on the OSF website,[8] and was approved as compliant with relevant ethical and data protection requirements.

[8]https://osf.io/xk7zr/?view_only = 5a921dd5b85f4d20987a691ca8c48a37

3.1 Design

To explore how user autonomy is affected by dark patterns, we used a research design close to a field experiment. Participants were recruited from an existing recruitment platform (Prolific) and instructed to evaluate the website design of a newly launched investment product called 'BeanStocks'. The product offered as part of our experiment was for an online financial investment product, and was similar to financial products offered on other trading platforms, such as the Robinhood app.[9]

Participants were first asked to complete a questionnaire before reviewing the website and were told that information about their personality traits (obtained via a questionnaire) was required to better understand their expressed opinions about the website's design. Participants then replied to several questions about the website design and their views about the investment product.

After participants completed what they considered to be the main task of website evaluation, a pop-up window appeared and informed them that in recognition of their assistance with the website evaluation, they now had an opportunity to buy the product in question. It is at this stage that we subjected the participants to dark patterns. Specifically, the pop-up offer messages presented to participants varied according to the dark pattern condition they were (randomly) assigned to (the next section provides details of each dark pattern condition used in the experiment). Critically, to enhance the realism and accuracy of the experiment, the fact that dark patterns had been pre-embedded in the pop-up offer was not disclosed to participants at the outset. Our aim was to convince participants that they were making a genuine purchase decision online, involving a real payment for a product. At the end of the offer process, we used a credit card payment page to ensure participants' responses reflected a genuine commitment to purchase the product.

The experiment involved an element of deception on our part, but it was nevertheless kept to the minimum necessary. Participants were promptly debriefed after responding to the offer. In the debriefing, they were told the real purpose of the study and informed that no such product offer exists and that they would not be charged (see Figure A4 in Annex C).

We based our experiment on an online financial product for several reasons. First, the use of online financial trading platforms has accelerated in recent years, widening the number of consumers who can directly access such services (ESMA, 2022). Second, the development of online financial platforms has been largely driven by a new category of, often inexperienced and young, users that diversifies the expected profile of investors. This makes the question of whether such younger and inexperienced users are more susceptible to dark patterns one of critical importance. Third, studies have shown that decision-making under uncertainty, such as in financial markets, often involves heuristic processing. Unlike systematic processing that involves careful and deliberate information processing, heuristic processing reduces cognitive efforts

[9]We chose this product for several reasons, including that: (i) it was not just a one-off purchase, like a TV, but involved an ongoing subscription/commitment; (ii) the product was complex, but offered consumers a simple way to 'cut through' such complexity; (iii) customers would have likely been exposed in the past to advertisements and technologies that refer to online investment products, (iv) customers can adjust their spending, by determining the level of investment.

by using simple decision rules to quickly analyse situations (Tversky and Kahnmenn, 1974). While these simplifications, such as buying stocks based on familiarity of companies' names or a friend's recommendation, are often sensible and valuable in uncertain environments, they are also susceptible to predictable biases. For instance, heuristic decisions are prone to persuasive techniques such as framing, anchoring and mere exposure. These effects are further intensified in digital settings (Zhang *et al.*, 2015; Luo *et al.*, 2018).

3.2 Procedure

In determining the geographical location of the study and the fictitious website, we took the decision to base it in the UK. Notwithstanding the fact that the UK is a developed economy, there continues to exist profound disparities in income, education and digital and financial literacy, which are relevant to our study. At the same time, the wide availability of internet access, and the common marketing of online services in the UK, increases the likelihood that users would be broadly familiar with the possibility of using online financial products such as the one we offered in our experiment.

The main online experiment was conducted in five identical sessions on weekday evenings from March to May 2022, with a total of 2500 voluntary UK participants. In each session, participants were recruited from the Prolific platform and electronically provided their informed consent within the survey before participation. In the consent form, participants were informed that the data were collected pseudonymously for research purposes only. As described above, the experiment was implemented in a web browser via Qualtrics (survey). Participants were asked to review the 'BeanStocks' website and required to answer questions as to its functionality and content. On average participants spent fifteen to twenty minutes completing the survey. At all times, participants could withdraw from the experiment without giving any reason by simply closing the web window.

The online experiment was undertaken in four phases. First, participants were invited to participate through an email sent by Prolific or an advertisement shown on that platform, with a recruitment message and a link to the survey. The recruitment message included a summary of the purpose of the study (website evaluation), inclusion and exclusion criteria, the duration of the experiment and contact information of the research team. Participants who chose to accept the invitation were asked to click on a link which took them directly to the Qualtrics platform. They were then provided with a formal consent form to indicate their willingness to participate in the experiment and those who consented began the next part.

In the second part, participants answered demographic questions, including their age, gender, education level, pre-tax monthly household income in the past 12 months and self-assessments about actual and perceived financial position. Following the demographic questions, participants answered several questions about their preferences, knowledge and experiences. Most of the questions consisted of 1–5 Likert scale-type questions that measure participants' digital literacy, online shopping experience, impulse buying tendency, propensity to trust and need for cognition (NFC). In addition, four multiple choice questions were used to test participants' financial knowledge, each of which had only one correct answer. To assess risk preferences and time preferences, participants were asked to report their willingness to take a risk (or to give up something that is beneficial today for more benefit in the future) from points 1 to 10. All these questions were presented in a random order to prevent an order effect arising. Participants were not told what these questions were measuring. We refer to this part of the experiment as the 'personal questionnaire'.

After completing the personal questionnaire, participants were given a link to the website of the product. This was the start of the third part of the experiment which we call 'website evaluation'. Participants were instructed to evaluate the website carefully as this was the declared purpose of their recruitment. They were asked questions about their feelings and views about the design of the website (e.g., on a 1–5 Likert scale or yes/no indicators) and provided their answers on the Qualtrics platform.[10] At this stage, participants were required to have two tabs open: one for the Qualtrics survey and the other for the website of the product. Responding to these three phases constituted a major proportion of the experiment time.

In the last phase, participants were informed that their survey responses were submitted and recorded, and they were now invited to press a button '*Tell Me More*' on the website to open and review an offer pop-up from 'BeanStocks'. Participants were also told that the product offer was made to express our gratitude for their participation, and that the offer was not a part of the survey they had just completed and therefore they did not have to accept the offer to complete the study. Once participants clicked the '*Tell Me More*' button in the website tab, they were either randomly exposed to a dark pattern manipulation or, if they were in the control group, to no manipulation.[11]

We adopted a between-subjects design, such that participants were randomly assigned to different dark pattern treatment conditions, namely: (1) False hierarchy; (2) Roach motel and (3) Confirm shaming. The control group received a neutral offer (shown in Figure 2(b)) which did not include a dark pattern and was designed to be informative. Each of the four dark pattern treatment conditions contained a specific product offer, that was similar to the neutral offer made to the control group, but with an alteration to the two options of choice as explained in Table 1. The False hierarchy condition is presented in Figure 3 as an example, while all the other dark pattern treatment conditions are shown in Appendix C.

All participants, including those in the control group, were informed that if they accepted the offer, they will be directed to a payment page (the '*proceed to payment*'). We also informed all participants that the minimum investment was £10. The use of a neutral offer to the control group allowed us to establish an objective benchmark by which to measure the proportion of participants interested in the BeanStocks trading product, without the use of dark patterns.

Participants that indicated a willingness to accept the product on the offer page – i.e, clicked on the green 'proceed to payment' button – were immediately transferred to a

[10] Many of the questions were based on Elling et al.'s (2012) questionnaire about website evaluation.

[11] Participants who did not click the '*Tell Me More*' did not enter the manipulation stage. As they were not exposed to the manipulation we do not refer to them in the body of the text, and were obliged by the ethics approval to delete their data. In contrast, if a participant dropped out of the experiment post-manipulation (for example by closing the session after being exposed to the dark pattern), we included this in our analysis and treated it as 'reject' at both stages.

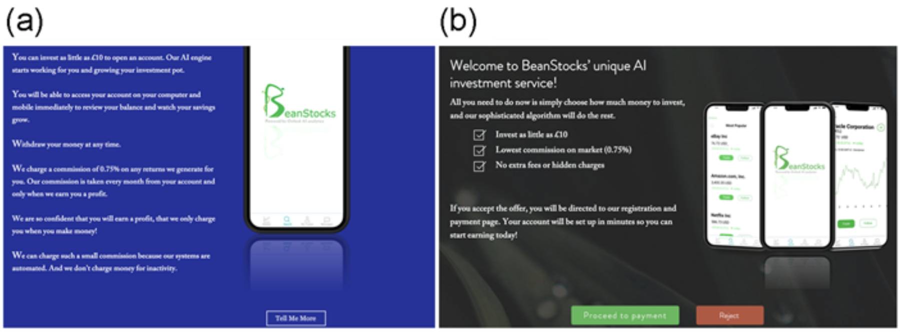


Figure 2. '*Tell Me More*' and the neutral offer pop-up.
Note: Figure 2a shows the last page of the website with the link that launches the offer (and manipulation stage) at the bottom. Figure 2b shows the pop-up window that then appears for the neutral offer (control group). If a participant chose the green option ('Proceed to payment') they were then directed to the payment page (Figure 1(b)). If they clicked on the red option ('Reject') they were directed to the debriefing page.

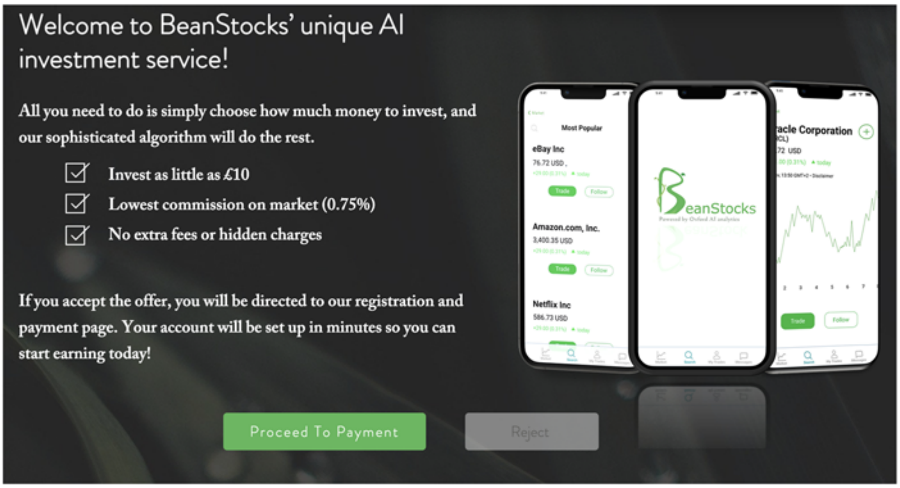


Figure 3. False hierarchy (visual interference).
Note: False hierarchy is a visual interference with the reject option shown less prominently making it less obvious to the user (compared to the red background in the neutral offer shown to the control group). The aim is to intentionally steer users away from certain choices by making them feel that less-visible options are unavailable or disabled (Mathur *et al.*, 2019). Consistent with real-world situations, in our experiment, the text of the offer is identical to the control group, as well as all other elements in the page for comparison. However, the Reject button is less prominent and shown in light grey.

(pseudo) payment page (see Figure 1(b)). Participants that declined the offer (choosing the red 'reject' button in the neutral offer) were directed to the debriefing page without seeing the payment page.[12] We refer to this decision to accept or reject the offer as the 'first choice'.

[12]Their payment choice was recorded as their 'first choice', i.e., decline in both stages.

Figure 4. Experiment flow.

Participants who accepted the 'first-choice' offer were then confronted with a second choice on the payment page where they could choose the amount they would like to invest, or could cancel the payment. We refer to this second decision point as 'payment choice' and consider this stage of the experiment that required a direct commitment to make payment and buy the service, as being critical to the credibility and realism of our results. Immediately after making the choice about the amount of payment (any other clickable elements, besides the amount or cancel options, would provide an error message of 'choose an amount first') a debriefing message appeared ending the experiment (before full credit card details were inserted).

The debriefing message explicitly and clearly informed participants of the full purpose of the study and that both the product and the company were fictitious. Participants were told that they were not going to be charged, nor would they receive the service.[13] In Figure 4, we present the flow of the four parts of the experiment and the main aim of each stage. Table 1 summarises the five conditions and variables used to code the data.

To ensure our results were reliable we used session recording videos that recorded participants' behaviours on the website (i.e., their mouse movements, focus, time per page and such). This allowed us to undertake a qualitative review of user behaviour and to identify malfunctions in the website design, such as lack of attention (proxied by extremely fast submission).

This qualitative examination enabled us to identify a technical problem that affected a fourth dark pattern condition (trick question) that was part of the original experiment. Specifically, while viewing the recordings after the experiment ended, we noticed a problem with the execution of the trick question pop-up window.[14] We therefore decided, as a conservative step, not to rely on this manipulation for the purposes of

[13]To ensure that all participants recognised the product offer was fictitious, they were asked to perform an attention check before receiving the completion code for the participation fee. Participants who withdrew before completing the offer stage would get the debriefing message through Prolific's system without being asked to perform this attention check.

[14]The trick question condition started with a double-negative question at the end of the offer: 'Do you prefer not to accept the offer?' To check whether participants who declined in the first choice were indeed intending to accept the offer, the experiment included an additional chance for them to respond to the offer: a second pop-up which directly asked participants if they would like to join the service. This second pop-up did not appear in any of the other conditions and was tailored to the trick question. The malfunction

Table 1. Treatment conditions and dependent variables

Variable name	Explanations
Condition	There were four conditions in the experiment: (a) Neutral offer (i.e., no manipulation) (Figure 2(b)). (b) False hierarchy – a visual inference with the accept/decline options (Figure 3). (c) Roach motel – two repeated screens with no decline option just 'more information' or 'accept' (Figure A1). (d) Confirm shaming – positive/negative language, supporting/discharging the user choice (Figure A2).
First choice	(a) Participants' first choice about whether to accept the offer in conditions they were assigned to. For the roach motel condition, it is the participant's choice to accept the offer at any stage (i.e., accept or decline in whatever stage). (b) This choice was coded as a binary variable, where 1 refers to a choice to proceed to payment page, and 0 means offer not accepted.
Payment choice	(a) Participants' choice about the amount they want to invest which was shown if their first choice was to accept the offer (see Figure 1(b)) (b) This variable was coded as binary, where 1 was for participants that chose to enter a payment amount, and 0 is for other participants who did not enter an amount or left the experiment. For the trick question condition, participants who accepted the offer in the follow-up question and eventually entered an investment amount were coded as 1.

our study and to only analyse the results obtained for the other three dark pattern conditions for which no technical problems were identified.

4. Sample description and underlying data structure

4.1. Sample description

The main online experiment comprised a total of 2500 voluntary UK participants. The final sample included 2252 complete observations,[15] with ages ranging from 18 to 86 years, (M =39.20 years, SD =14.17). Of this sample, 1126 were female (50.04%), 1089 participants were male (48.40%), 29 were non-binary (1.29%) and 8 did not disclose gender identity (0.36%). As shown in Figure 5, participants also varied in their income levels, with reported pre-tax monthly household income ranging from £0 to more than £7001. Our sample size was sufficiently large to allow us to measure all income groups in a robust manner, regardless of the median income of our participants, which was slightly more affluent than the UK population (Our median income was between £3,000 and £4,000 while the UK median income is around £2,750).

involved many participants not seeing this second pop-up window, which led them to refresh the page or press back space and stop the experiment, or to alter their choice.

[15]We excluded participants who opted out before completing the experiment, suffered malfunctions on the website or failed at least one of two attention checks. We also excluded participants who lived in the same household (detected by their IP addresses) and only include the first participant in that household.

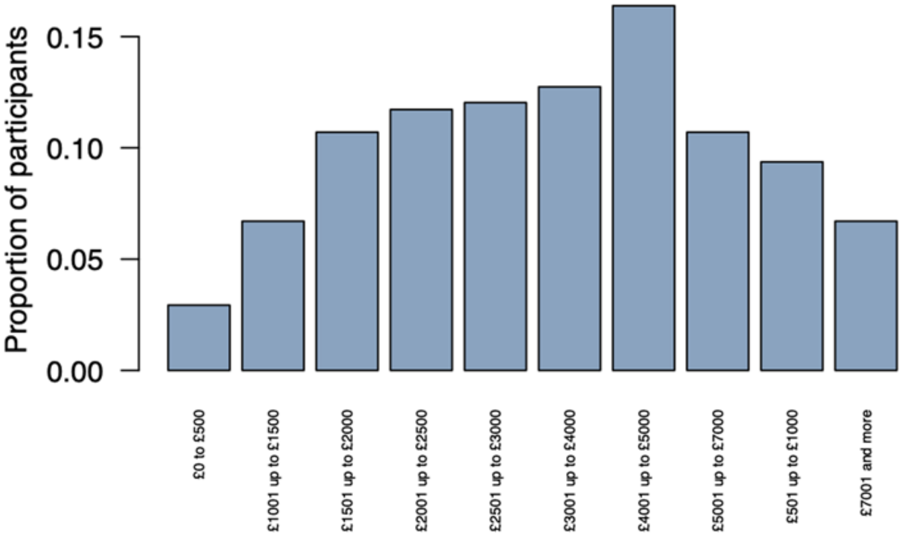


Figure 5. Distribution of income among participants.
Note: The total number of participants in each group: 66 between £0 and £500 (2.93%), 151 between £501 and £1000 (6.71%), 241 between £1001 and £1500 (10.70%), 264 between £1501 and £2000 (11.72%), 271 between £2001 and £2500 (12.03%), 287 between £2501 and £3000 (12.74%), 369 between £3001 and £4000 (16.39%), 241 between £4001 and £5000 (10.70%), 211 between £5001 and 7000 (9.37%) and 151 having more than £7001 (6.71%).

Table A2 in Appendix A includes a list of all variables collected in the personal questionnaire, with relevant literature references and sources. Table 2 provides the basic descriptive statistics for the key variables we used in our analysis.

We performed a two-sample Kolmogorov–Smirnov test using Matlab to understand whether the randomisation was perfectly implemented such that there were no systematic differences in demographic covariates between participants in the different groups. We found most of the demographic data to be distributed equally across all conditions, with only income covariates showing a difference in distributions between samples supposedly randomly assigned to two different conditions (detailed results are summarised in Appendix D).[16] However, as discussed in Section 5, we only find a weak relationship between (objective) income and dark pattern effectiveness. This makes the potential problem of non-randomness less severe. In other words, even though some income covariates are not distributed identically across all conditions, the results of our experiment remain valid and reliable.

[16]The demographic covariates that failed to exhibit randomness are objective income (when comparing the Neutral offer with Confirm shaming, False Hierarchy with Confirm shaming and the Roach motel with Confirm shaming), subjective income (when comparing the Neutral offer with Trick questions) and relative income (when comparing Neutral offer with Trick questions). As the trick question condition was not included in the analysis (as well as in the empirical insights), the analysis that requires careful consideration is the role of objective income in the differences in participants' consumption behaviour between the control and confirm-shaming conditions.

Table 2. Descriptive statistics

Statistic	*N*	Description	Mean	St. dev.	Min	Max
Age	2,252	Self-reported age	39.2	14.2	18	86
Income	2,252	Ten monthly household income segments from £0 to £7000 and more	5.8	2.4	1	10
Education	2,252	Six levels: GCSE; A-level; certificate/diploma of higher education; undergraduate degree; master's degree; doctoral degree	3.5	1.3	1	6
Impulsive tendency	2,252	Tendency to exhibit an unplanned, compelling and hedonic purchasing behaviour.	14.7	4.8	5	25
Digital literacy	2,252	Participants' digital knowledge and experience	76.9	9.1	27	90
Shopping experience	2,252	Participants' online shopping experience	8.2	1.9	2	10
Financial literacy	2,252	Financial knowledge	2.7	0.9	0	4
Trust tendency	2,252	Participants' propensity to trust	13.8	3.9	4	20
NFC	2,252	Participants' degrees of need for cognition (i.e., their tendency to engage in thinking and enjoy solving complex problems	21.1	4.9	6	30
Time preference	2,252	Participants' self-reported time preferences. The higher the value is, the more patient the participant is.	6.7	1.8	0	10
Risk preference	2,252	Participants' self-reported risk preferences. The higher the value is, the participant is more willing to take a risk	5.2	2.2	0	10

4.2 Correlations structure

Using our rich dataset we are able to study consumers' attributes, which are the first step in understanding online vulnerability beyond superficial demographics. Our sample includes both binary, categorical and continuous variables (e.g., gender, education, subjective income, respectively). For the continuous variables, we use Pearson's correlation tests, while for the binary variables (e.g., gender), we conducted a *t*-test, Wilcoxon rank sum test and correlations. For categorical/ordinal variables (e.g., education and income), we use ANOVA, Kruskal–Wallis test and correlations. Table A3 (in the Appendix) summarises the underlying relationships between the eight personal characteristic measures and the key demographic data collected in the personal questionnaire.

Section 5.3 investigates whether dark patterns have heterogenous effects using three standard proxies for consumer vulnerability (age, education and income). Before considering the results of this analysis, it is useful to consider how each of these demographic variables correlated with the various personal characteristics in our data. We found age to be positively correlated with trust, NFC and financial literacy, and negatively correlated with impulsive shopping, digital literacy and risk-taking behaviour. Similarly, we found that higher levels of educational attainment are positively correlated with digital literacy, higher cognition and higher levels of patience. Conversely, our data suggest that educational attainment was negatively correlated with impulsive shopping. Finally, we found a positive correlation between income and trust, cognition, financial literacy, risk taking, experience of online shopping and patience. Appendix B describes the results of our correlation analysis in detail and compares our data with the findings of existing research for each personal characteristic. We go back to these findings while summarizing the key analysis of vulnerability.

5. Results

This section presents the main results of the experiment. First, we present the results obtained for each of the three different dark pattern conditions (False Hierarchy, Confirm Shaming, Roach Motel) controlling for any underlying differences between participants, such as income, age and other personal characteristics. These are the *average* effects of dark patterns on all participants. We then move on to explore whether the effectiveness of the three dark pattern conditions varied among participants with different characteristics using an interaction analysis between treatment and demographics.

5.1 Did dark patterns affect the average acceptance rate for all participants?

To examine if a participants' willingness to accept the offer was affected by the condition that they faced (note that in this analysis, the neutral offer is the baseline condition) a cross-tabulation (contingency table) was computed along with a Pearson's chi-square analysis.[17] The statistical analysis implies that the use of dark patterns *did* affect the acceptance decision of participants in our experiment, i.e., the decision to accept was *dependent* on the condition a participant faced.

However, this statistical evidence only suggests an overall association between the conditions and a participant's choice (either first choice or payment choice). To explore the effectiveness of each dark pattern condition relative to the baseline condition, we implemented several statistical tests for each condition separately.[18] Results are summarised in Table 3.

[17]This statistical approach is commonly used to test dependency among categorical data and is ideal for a large sample size. The null hypothesis in this test is that the conditions of the experiment and choice (first choice and payment choice, respectively) are *independent*. We rejected the null at 5% level of significance (X^{200} (4, 2252) $= 68.4142$, $p < .0001$ for first choice; X^{200} (4, 2252) $= 14.4418$, $p = .006$ for payment choice).

[18]We also ran a permutation test (a.k.a. a re-randomisation test) to examine whether observed mean differences in acceptance rates between the Nature condition and a treatment condition are due to a random chance (detailed results are summarised in the Appendix D). The tested results confirm that the observed differences do come from the treatment effect, not randomness.

Table 3. Average treatment effects

Condition	*N*	(a) Mean acceptance: first choice	(b) Mean acceptance: payment choice	(c) Chi-square test statistic: first choice	(d) Chi-square test statistic: payment choice
Baseline	443	0.4876	0.0790	N/A	N/A
False hierarchy	457	0.7046	0.1510	44.0589***	11.4029***
Roach motel	439	0.6128	0.0957	13.9578***	0.7685
Confirm shaming	451	0.5543	0.1131	3.9889**	2.9846*

*$p < 0.1$, **$p < 0.05$, ***$p < 0.001$.

Columns (c) and (d) in Table 3 show that some dark patterns materially changed the acceptance rates, compared to the baseline. Notably, based on the results of the cross-tabulation analysis, the first-choice acceptance rates are significantly higher for the False Hierarchy and Roach Motel conditions than for the baseline condition. This implies that these types of dark patterns were effective in increasing the acceptance rates for the offer we made to participants. However, when confronted with a second decision point involving payment for the service, the effectiveness of two of the dark patterns (Roach Motel and Confirm Shaming) weakens. Recall, however, that the payment page was identical across all conditions and did not include a dark pattern. Indeed, it was created with the intention of capturing whether using dark patterns at the offer stage are effective in leading a participant to complete the full transaction involving a real payment at a later stage.

To further investigate the effects of dark patterns on our sample participants, we applied both a linear model and GLM (logistic regression) to predict the average likelihood that participants, by condition, would accept the offer at the two time points of the experiment: *first choice* and *payment choice* (Y). The first linear form is presented below:

$$\mathrm{Y} = \mu + \alpha \mathrm{D} + \eta Z_{1..N} + \varepsilon.$$

D is our treatment indicator, which takes binary form (when all dark patterns are analysed together) and a categorical form (when each dark pattern is analysed separately as a treatment condition vs the control group). The coefficient α takes the association between the dark pattern treatment variables, while Z is a vector for the controls, the personal and demographic characteristics from the personal questionnaire. The variables in Z capture any effect which is not related to treatment but directly predicts acceptance. For simplicity we present the results of the linear regression here, and the GLM in Appendix E. Columns (1) and (3) in Table 4 present the results when the three dark pattern conditions are combined to measure the effect of dark patterns vis-à-vis the control group. Columns (2) and (4) separate out each dark pattern treatment group.

Overall, the results indicate stronger acceptance rates for the combined dark patterns group compared to the baseline acceptance rate. This holds for both the first-choice and payment-choice stages, meaning that the dark patterns were not only

Table 4. LM regression full results

	Dependent variable			
	First choice		Payment choice	
	(1)	(2)	(3)	(4)
Dark patterns	0.139*** (0.027)		0.042** (0.017)	
False hierarchy		0.225*** (0.033)		0.075*** (0.021)
Roach motel		0.127*** (0.033)		0.017 (0.021)
Confirm shaming		0.063* (0.033)		0.035* (0.021)
Age	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Income	−0.0004 (0.005)	−0.001 (0.005)	0.005 (0.003)	0.005 (0.003)
Education	0.001 (0.009)	0.001 (0.009)	0.004 (0.006)	0.003 (0.006)
Risk pref	0.005 (0.006)	0.004 (0.006)	0.002 (0.004)	0.002 (0.004)
Financial literacy	0.005 (0.014)	0.007 (0.014)	0.004 (0.009)	0.004 (0.009)
Digital literacy	0.001 (0.001)	0.001 (0.001)	0.002** (0.001)	0.002** (0.001)
Shopping experience	0.006 (0.006)	0.004 (0.006)	−0.002 (0.004)	−0.002 (0.004)
Impulsive	0.004 (0.003)	0.005 (0.003)	0.002 (0.002)	0.002 (0.002)
Time pref	0.010 (0.007)	0.011 (0.007)	−0.0004 (0.004)	−0.0004 (0.004)
Gender (male)	0.030 (0.025)	0.033 (0.025)	0.037** (0.016)	0.038** (0.016)
Trust	−0.0005 (0.003)	−0.0001 (0.003)	−0.001 (0.002)	−0.001 (0.002)
NFC	0.001 (0.003)	0.002 (0.003)	−0.0002 (0.002)	−0.0002 (0.002)
Observations	1,790	1,790	1,790	1,790
R^{200}	0.022	0.035	0.011	0.016
Adjusted R^{200}	0.017	0.028	0.006	0.009

Note: All variables are scaled the mean.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

effective in altering participants' impulsive decision to accept, but also influenced their decision to commit to payment.

None of the demographic and personal characteristics variables are statistically significant at the first-choice stage which suggests that they are not driving the results, although men were more likely to accept the offer at the payment-choice stage. These findings differ from previous studies, such as Luguri and Strahilevitz (2021, 81) who found that 'less education increases vulnerability to small- or moderate-dose dark patterns'. We discuss the moderation effects of education in more detail in the next section.

Looking at each dark pattern treatment condition separately reveals that the False Hierarchy was the most effective dark pattern. Although the Confirm Shaming dark pattern was effective at the first-choice stage it was weaker when participants were given a 'second chance' and directly confronted with a need to pay for their decision. The Roach Motel dark pattern also appears to have caused some backlash among participants that were exposed to this condition, as shown by the gap between first choice and payment choice which is also apparent by the fact that the payment-choice results are not statistically significant.

5.2 Did the effectiveness of dark patterns differ among participants according to their characteristics?

Having examined the *average* effects of dark patterns on all participants, we turned to the question of whether there was evidence that dark patterns had *heterogenous* effects according to certain characteristics of the participants in our sample. Section 5.2.1 explains the modifications we made to the classic interaction model, while Section 5.2.2 presents the results for the effects of dark patterns on three common proxies for consumer vulnerability: income, education and age. We summarize the main findings in Section 5.3.

5.2.1 The interaction models

An interaction model examines whether the relationship between the outcome Y (first choice or payment choice) and a key independent variable D (dark patterns treatment) varies with levels of a moderator X (in our case the demographic variables: income, education and age). Specifically in our case, it might be conjectured based on the standard consumer vulnerability paradigm that dark patterns would be more effective for participants that were older, or with lower levels of income or education attainment. Such conditional hypotheses are common in the social sciences and linear regression models with multiplicative interaction terms are the most widely used framework for testing them in applied work (Brambor *et al.*, 2006). μ and ϵ represent the constant and error terms, respectively. The regression equation for the interaction model is

$$\mathrm{Y} = \mu + \alpha\mathrm{D} + \eta\mathrm{X} + \beta\left(\mathrm{D}^{*}\mathrm{X}\right) + \varepsilon.$$

We modified the Brambor *et al.* (2006) classic model, using a binning estimator as set out in Hainmueller, Mummolo, & Xu (2019).[19] The binning estimator breaks a continuous moderator into several bins represented by dummy variables and then examines the interaction between these dummy variables and the treatment indicator, with some adjustment to improve interpretability. The binning estimator is much more flexible as it jointly fits the interaction components of the classic model to each bin separately, thereby relaxing the linear interaction assumption. Since the bins are constructed based on the support of *X*, the binning ensures that the conditional marginal effects are estimated at typical values of the moderator and do not rely on excessive extrapolation.

[19]The classic interaction model relies on two key assumptions. First, while multiplicative interaction models allow the effect of the key independent variable D to vary across levels of the moderator X, they maintain the important assumption that the interaction effect is linear and follows the functional form given by $\partial Y/\partial D = \alpha + \beta X$. The second assumption is common support, i.e., there are sufficient data when computing the conditional marginal effects. Ideally, to compute the marginal effect of D at a given value of the moderator, X_1, there needs to be (1) a sufficient number of observations which X values are close to X_1, and (2) variation in the treatment, D, at X_1 (e.g., different treatment groups or values). Otherwise, the conditional marginal-effect estimates are based on extrapolation or interpolation of the functional form to an area where there are no or only sparse data and therefore the effect estimates are fragile (King and Zeng, 2006).

5.2.2 Results of interaction models

Having set out our approach, Table 5 summarises the results, for income and education, while Table 6 summaries the interaction model for Age. We continue to study the results of Table 6 graphically to assess the strength of the evidence supporting an age moderation effect.

As shown in Table 5, we do not identify any statistically significant moderation effects of income or education to support the conjecture that the effects of dark patterns monotonically decrease with higher income or educational attainment as the standard vulnerability argument typically assumes (i.e., that susceptibility to dark patterns decreases with higher levels of income or education). While the interaction model supports the main findings from Table 4 (i.e., the dark patterns conditions, and especially the false hierarchy are powerful designs for deception), all of the interaction models in Table 5 are statistically insignificant, as are the income and education indicators.[20]

We also did not observe a clear statistical relationship between financial literacy and dark patterns (see Appendix E). While there was no statistically significant relationship between financial literacy at the first-choice stage, in some specifications we observed a slight positive moderation effect at the payment-choice stage suggesting higher levels of financial literacy could be associated with a stronger treatment effect (see Table A6). This was a surprising result as our expectation was that, given the product we were offering, there would have been a negative relationship between financial literacy and dark pattern susceptibility.[21] While further research is needed to understand the relationship between financial literacy and dark patterns, one possible explanation for our results is that the novel online financial product used in the experiment was unfamiliar to participants who have a general level (which is what the personal characteristic measures) of financial literacy. In other words, the unique fintech context might have meant that even those participants with a general level of financial knowledge were still susceptible to the dark patterns. Indeed, it may have been the case that those with higher levels of financial literacy were more attracted to the product offering given their general interest in investment and finance.[22]

Having found no evidence to support the prior assumptions concerning heterogenous effects on dark patterns according to income and education, we continue to study the moderation effect of age. In contrast to our findings for income and education, Table 6 shows that age does seem to interact with our treatment, i.e., higher age groups were slightly more susceptible to dark patterns at the first-choice stage. However, the moderating effect of age appears to differ by dark pattern treatment. Figure 6 shows a positive moderation effect of age for Confirm Shaming and False Hierarchy, and a negative moderation effect of age for the Roach Motel treatment. This could potentially be

[20] We ran various models to test the results, including models transformed the variables into continuous variables. The results were still not statistically significant.

[21] We used standard tests to assess financial literacy, including asking participants to answer to calculate the compound interest on a £100 savings account over 5 years; to calculate the impact of inflation on the interest earned from a savings account; and to describe what happens to bond prices when interest rates rise.

[22] In the appendix (Table A6), we include the full results of the financial literacy interaction models, with a raw plot of the first-choice results.

Table 5. LM interaction results, income and education

	Dependent variable							
	First choice		Payment choice		First choice		Payment choice	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dark patterns	0.140*** (0.027)		0.042** (0.017)		0.139*** (0.027)		0.043** (0.017)	
False hierarchy		0.227*** (0.033)		0.074*** (0.021)		0.224*** (0.033)		0.075*** (0.021)
Roach motel		0.129*** (0.033)		0.016 (0.021)		0.126*** (0.033)		0.017 (0.021)
Confirm-shaming		0.064* (0.033)		0.035* (0.021)		0.063* (0.033)		0.036* (0.021)
Age	0.008 (0.010)	0.008 (0.010)	0.003 (0.006)	0.003 (0.006)	−0.0004 (0.005)	−0.001 (0.005)	0.005 (0.003)	0.005 (0.003)
Income	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Education	0.001 (0.009)	0.001 (0.009)	0.004 (0.006)	0.003 (0.006)	0.001 (0.018)	0.001 (0.018)	0.009 (0.012)	0.009 (0.012)
Risk pref	0.005 (0.006)	0.004 (0.006)	0.002 (0.004)	0.002 (0.004)	0.005 (0.006)	0.004 (0.006)	0.002 (0.004)	0.002 (0.004)
Financial literacy	0.006 (0.014)	0.008 (0.014)	0.004 (0.009)	0.004 (0.009)	0.005 (0.014)	0.007 (0.014)	0.004 (0.009)	0.004 (0.009)
Digital literacy	0.001 (0.001)	0.001 (0.001)	0.002** (0.001)	0.002** (0.001)	0.001 (0.001)	0.001 (0.001)	0.002** (0.001)	0.002** (0.001)
Shopping experience	0.006 (0.006)	0.004 (0.006)	−0.002 (0.004)	−0.002 (0.004)	0.006 (0.006)	0.004 (0.006)	−0.002 (0.004)	−0.002 (0.004)
Impulsive	0.004 (0.003)	0.005 (0.003)	0.002 (0.002)	0.002 (0.002)	0.004 (0.003)	0.005 (0.003)	0.002 (0.002)	0.002 (0.002)
Time pref	0.010 (0.007)	0.011 (0.007)	−0.0004 (0.004)	−0.0004 (0.004)	0.010 (0.007)	0.011 (0.007)	−0.0004 (0.004)	−0.0003 (0.004)

(*Continued*)

Table 5. (*Continued.*)

	Dependent variable							
	First choice		Payment choice		First choice		Payment choice	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender (male)	0.030 (0.025)	0.032 (0.025)	0.037** (0.016)	0.039** (0.016)	0.030 (0.025)	0.034 (0.025)	0.037** (0.016)	0.039** (0.016)
Trust	−0.0005 (0.003)	−0.0002 (0.003)	−0.001 (0.002)	−0.001 (0.002)	−0.0005 (0.003)	−0.0002 (0.003)	−0.001 (0.002)	−0.001 (0.002)
Nfc	0.001 (0.003)	0.002 (0.003)	−0.0002 (0.002)	−0.0002 (0.002)	0.001 (0.003)	0.001 (0.003)	−0.0002 (0.002)	−0.0003 (0.002)
Dark patterns: income	−0.011 (0.011)		0.002 (0.007)					
False hierarchy: income		−0.008 (0.014)		0.0004 (0.009)				
Roach motel: income		−0.016 (0.014)		0.004 (0.009)				
Confirm shaming: income		−0.013 (0.014)		0.002 (0.009)				
Dark patterns: education					0.001 (0.021)		−0.006 (0.013)	
False hierarchy: education						0.017 (0.025)		−0.001 (0.016)
Roach motel: education						−0.033 (0.025)		−0.018 (0.016)
Confirm shaming: education						0.017 (0.025)		−0.001 (0.016)
Observations	1,758	1,758	1,758	1,758	1,758	1,758	1,758	1,758
R^{200}	0.022	0.036	0.013	0.018	0.021	0.038	0.013	0.019
Adjusted R^{200}	0.014	0.026	0.005	0.008	0.013	0.028	0.005	0.008

Note: All variables are scaled the mean.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 6. LM interaction results, age

	Dependent variable			
	First choice		Payment choice	
	1	2	3	4
Dark patterns	0.138*** (0.027)		0.042** (0.017)	
False hierarchy		0.224*** (0.033)		0.074*** (0.021)
Roach motel		0.125*** (0.033)		0.016 (0.021)
Confirm shaming		0.061* (0.033)		0.035 (0.021)
Age	−0.001 (0.002)	−0.001 (0.002)	−0.0003 (0.001)	−0.0003 (0.001)
Income	−0.0003 (0.005)	−0.001 (0.005)	0.005 (0.003)	0.005 (0.003)
Education	0.001 (0.009)	0.003 (0.009)	0.004 (0.006)	0.004 (0.006)
Risk pref	0.005 (0.006)	0.004 (0.006)	0.002 (0.004)	0.002 (0.004)
Financial literacy	0.005 (0.014)	0.006 (0.014)	0.004 (0.009)	0.004 (0.009)
Digital literacy	0.001 (0.001)	0.001 (0.001)	0.002** (0.001)	0.002** (0.001)
Shopping exp	0.006 (0.006)	0.004 (0.006)	−0.002 (0.004)	−0.002 (0.004)
Impulsive	0.004 (0.003)	0.005* (0.003)	0.002 (0.002)	0.002 (0.002)
Time pref	0.010 (0.007)	0.012* (0.007)	−0.0003 (0.004)	−0.0001 (0.004)
Gender (male)	0.030 (0.025)	0.033 (0.025)	0.038** (0.016)	0.039** (0.016)
Trust	−0.001 (0.003)	0.0001 (0.003)	−0.001 (0.002)	−0.001 (0.002)
NfC	0.001 (0.003)	0.001 (0.003)	−0.0002 (0.002)	−0.0003 (0.002)
Dark patterns: age	0.002 (0.002)		0.001 (0.001)	
Confirm shaming: age		0.005** (0.002)		0.001 (0.001)
False hierarchy: age		0.005** (0.002)		0.002 (0.001)
Roach motel: age		-0.004* (0.002)		0.0003 (0.002)
Observations	1,758	1,758	1,758	1,758
R^2	0.022	0.047	0.014	0.019
Adjusted R^2	0.014	0.037	0.006	0.009

Note: All variables are scaled the mean.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

explained by the fact that the Roach Motel treatment relies on impatience and is therefore possibly less effective on older participants (age is correlated with patience in our sample).

The non-linearity of the LOESS (red) curve in Figure 6 motivates further investigation,[23] and we used binning model plots to relax the linear interaction assumption. Using a binning model, Figure 7 shows that the effects of dark patterns are not fully uniform across the age groups. The oldest age group bin coefficient is lower than the

[23]Using the Wald test, we can reject the NULL hypothesis that the linear interaction model and the five-bin model are statistically equivalent), meaning the binning model might be preferred over the liner interaction.

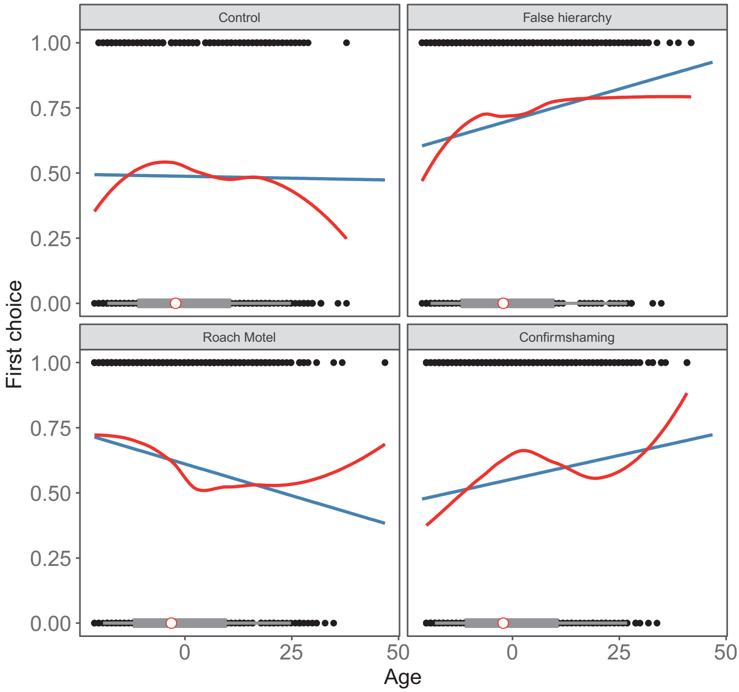


Figure 6. Marginal effect size by age group – raw plot.
Note: Interaction model, showing moderation effects of age in blue, and the red represents a LOESS (locally estimated scatterplot smoothing) approximation. The bottom bar shows the distribution of the moderator age when scaled to the average. The top left panel shows results for the control group, while the other cells show the three dark patterns groups.

previous age group bins for both False Hierarchy and Confirm Shaming. However, the wide confidence intervals leave scope for uncertainty.[24]

Taken as whole, we believe our analysis provides some evidence of a moderating effect of age, which is monotonically increasing as the standard vulnerability argument typically assumes (i.e., that susceptibility to dark patterns increases with age).[25]

5.3 Summary of results

Table 7 presents an overview of the findings in Section 5.2, with the additional associations between demographics and personal attributes discussed in Section 4.2.

[24]The confidence intervals for the binning model marginal effects are wide, running from 0 to ~0.3. Such wide intervals suggest a large degree of uncertainty, yet in the areas where more observations are available (lower ages) the binning model is consistent with the findings of Table 6.

[25]We also used a third model, the Kernel estimator (Hainmueller *et al.*, 2019), which does not rely on linearity at all. Results are consistent with the reported findings.

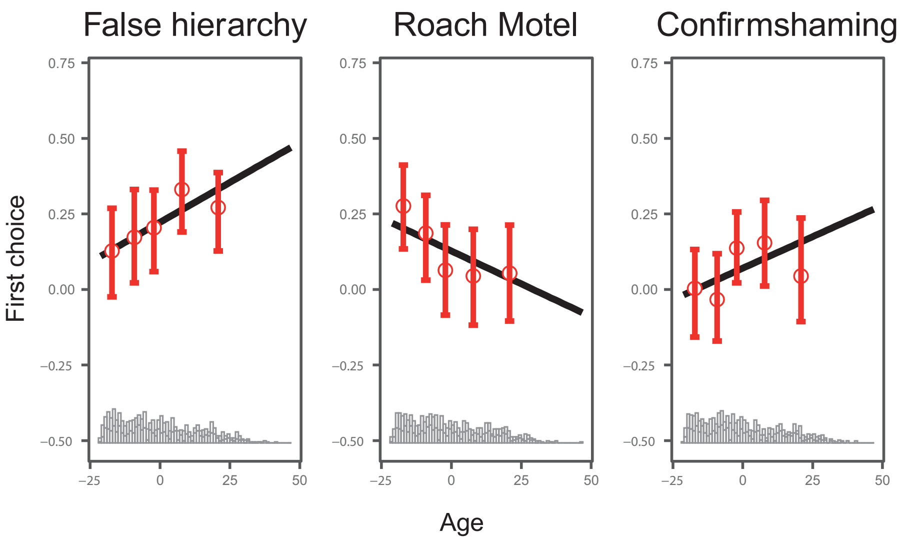


Figure 7. Marginal effect size by age group – binning model.
Note: Interaction model with five roughly equal-size bins, showing moderation effects of age. In the bottom histogram the total height of the stacked bars refers to the distribution of the moderator in the pooled sample and the red and grey shaded bars refer to the distribution of the moderator in the treatment and control groups, respectively. Each bin coefficient is plotted with 95% confidence interval.

Table 7. Summary of the main findings

Condition	*N*	Q1: effect on acceptance	Q2: susceptibility of demographic group	Correlation between age and personal attributes in our data
Dark patterns combined	1347	Strong	Age	+ Propensity to trust + Financial literacy − Digital literacy − Impulsive − Risk preference
False hierarchy	457	Strong	Age (positive)	
Roach motel	439	Strong	Age (Negative)	
Confirm shaming	451	Weak	Age (Positive)	

Note: 'Strong' results refer to statistical significance at 5% or 1% and large enough magnitudes across the estimates in the paper. 'Weak' results refer to unstable or weak statistical significance or a very small magnitude in effect size. The notion '+' and '−' in the last column indicates the direction of the association between age and the personal attribute based on the relevant statistical test.

In summary, we find that dark patterns affected the acceptance decision of participants in our experiment, and that older users were slightly more susceptible to some dark patterns such as false hierarchy and confirm shaming.

While our data are not large enough to allow us to establish the formal psychological and cognitive mechanisms which might have driven these results, it nevertheless suggests that a high tendency to trust and low digital literacy as possible channels for greater susceptibility to dark patterns. In addition, our analysis suggests that for novel online financial products which purport to use AI and other sophisticated techniques like in our case, a general knowledge of finance, low impulsivity and low risk preference did not seem to protect older participants from the effects of dark patterns.

These findings are based on the results of our experiment using a specific product in a particular jurisdiction, and more research is needed before they can be generalised to other products and jurisdictions.

5.5 Limitations of the analysis

Several caveats to our analysis should be acknowledged.

First, our participants were recruited from an online platform (Prolific) which matches researchers with participants. While this site is widely used in academia, and for previous dark patterns studies, concerns have been raised about the representativeness of the pool of participants who use these platforms. There may be a selection bias (volunteer bias) of our sample which may reduce the generalisability of the study results. We tried to mitigate this by gathering relevant data in our questionnaire, and by testing the distributions of our three key variables of interest (age, income, education).

Second, there is also a risk that participants were sceptical of the website experiment and the genuine nature of the offers being made to them. While we designed our website using inspiration from actual online financial products, we did not replicate the full user experience of setting up an account, signing disclaimers, reviewing contract terms etc. Our design was directed by the desire to shorten the participants' journey to fit the tight time and attention window of an online experiment. We note that two design factors may have enhanced the credibility of the offer. First, the inclusion of an immediate payment page which prompted participants to enter their actual credit card information (which we did not record), and second the fact that the product was also described as being part of an innovation advanced by an academic institution.

Third, the experiment was based on an online financial product in the UK, which, as noted above, means that the results may not immediately be transferable to other jurisdictions or to other consumer products, especially products those that are less complex and more familiar to consumers. Accordingly, more empirical work is needed before strong policy implications are reached in other settings. Furthermore, our design only targeted non-mobile users. As the US Federal Trade Commission has noted the type of device (mobile or PC) can potentially affect user interaction which again suggests that caution should be exercised before generalising our findings to all online users.

Finally, our experiment was based on three commonly used dark patterns out of an ever-expanding list. While we sought to address this limitation by undertaking a pilot experiment using many more dark patterns, it is possible that our results may either under-report or over-report the users' vulnerability to other types of dark patterns.

6. Concluding remarks

This study used a novel experimental design to test if dark patterns are effective in manipulating and influencing consumer decision-making, and further, whether the effectiveness of dark patterns varied based on three commonly used proxies of consumer vulnerability (age, income and education). Our findings offer relevant insights that could influence policy debate and regulation at two distinct levels:

6.1 Dark patterns and consumer vulnerability

We find strong evidence that dark patterns are effective across all participants (average effects), but no evidence that the effectiveness of dark patterns is conditioned by income or education. Our results offer some evidence that older age participants are more susceptible to manipulation (heterogeneous effects). As such, our results suggest that the classic general proxies for consumer vulnerability (characteristics such as income or education) may be of limited relevance when considering consumer susceptibility to dark patterns. Instead, our analysis favours the alternative view that *all* online consumers are potentially vulnerable when confronted with dark patterns, with possible heightened vulnerability for older users.

These findings support current policy debate and recent regulatory efforts. As mentioned in the introduction, the EU's recently adopted DSA enshrines a specific ban on the use of dark patterns by online platforms, which appears to put forward an implicit acknowledgment that *all* consumers are vulnerable to falling prey to such practices. In the UK, the CMA has published various papers on which have addressed dark patterns or harmful online choice architectures (CMA, 2022; CMA and ICO, 2023). The CMA has also taken enforcement action against websites which used 'urgency claims', such as countdown timers and 'discounts' to mislead consumers (CMA, 2023). This trend is not limited to Europe. Indeed, while the US FTC Staff Report on dark patterns states that they can have a greater impact on lower-income and other vulnerable populations, elsewhere the agency has acknowledged the need for a wider concept of vulnerability to be adopted when it comes to consumer decisions in digital markets. The FTC has put businesses on notice that every enforcement tool at the agency's disposal would be mobilized to tackle dark patterns (FTC, 2022). According to its consumer protection head, '[i]n today's digital economy, it is simply illogical to put the onus on *individuals* to appreciate the implications of this enormously complex ecosystem, an ecosystem powered by massive data collection and often arcane technology' (Levine, 2022).

Our study contributes to these policy debates by providing empirical evidence to support this trend towards policies that seek to protect all, or a large proportion of consumers, from dark patterns and other forms of online manipulation. As such it is also consistent with the wider but differentiated conception of vulnerability (European Parliament, 2021; OECD, 2023) that involves protecting the average consumer online with prohibitions against dark patterns, while at the same time allowing for enhanced protections for mental or physical infirmity and age.

6.2 Acceptance and payment choices

Our experimental design and the use of a payment page offer important insights as to when dark patterns may be most effective. As we elaborated above, many participants who were influenced by the dark pattern to accept the offer, did not proceed to buy the product once they reached the payment page. This implies that when further action was required by the user, following the successful deployment of a dark pattern, its effectiveness diminishes. Accordingly, the breaking of the acceptance and payment choices into two decision points in our experiment reduced consumer vulnerability. This insight highlights the enhanced vulnerability in instances which involve a 'single click' acceptance.

As a matter of policy, one could distinguish situations where dark patterns are deployed with an 'immediate effect' to syphon off data, encourage choice, or execute a charge when payment details are already stored online (e.g., where there is an existing subscription, a free trial period or the service provider has recorded the payment information, or when the dark pattern is deployed in setting of zero-pricing and no further action is required by the user e.g., cookie consent notices), from those where further action is needed beyond the immediate manipulation. Given our findings, the latter situations offer online consumers an opportunity for further reflection and may prompt them to reconsider their choices and opt out.

If one accepts this distinction, the latter situations appear to be a less urgent problem or at least one for which a tailored response may be more appropriate than a blanket prohibition. On the other hand, when dark patterns achieve their goal through a 'single click' or are deployed by an established platform or ecosystem that already possess the user's relevant payment details, users remain vulnerable and under the 'spell' of the dark pattern. These effects, coupled with the increased centrality of established online platforms, increased asymmetry of power between users and service providers, and the stealth deployment of dark patterns, may well justify blanket protections, as advanced by the European Union in its DSA.

Our study sheds light on the possible benefits which may flow from the slowing down of the decision-making process. Rather than celebrating users' acceptance being 'a click away', it points to the benefits of breaking the payment or acceptance process into two decision points. This insight may support the artificial slowing down of decision-making in instances in which consumer vulnerability could expose it to exploitation. Such may be the case in instances involving agreement to share personal data, accept onerous conditions or pay for services.

References

ACM, 'Guidelines – protection of the online consumer', Boundaries of online persuasion, (2020), https://www.acm.nl/en/publications/information-for-companies/acm-guideline/guidelines-protection-online-consumer [23 November 2024].

Ahern, K. R., R. Duchin and T. Shumway (2014), 'Peer effects in risk aversion and trust', *The Review of Financial Studies*, **27**(11): 3213–3240

Ariely, D. (2009), *Predictably Irrational: The Hidden Forces that Shape Our Decisions*, New York, US: Harper Collins.

Atkinson, A. and F. A. Messy (2011), 'Assessing financial literacy in 12 countries: an OECD/INFE international pilot exercise', *Journal of Pension Economics & Finance*, **10**(4): 657–665.

Badgaiyan, A. J. and A. Verma (2015), 'Does urge to buy impulsively differ from impulsive buying behaviour? Assessing the impact of situational factors', *Journal of Retailing and Consumer Services*, **22**, 145–157.

Baker, H. K., S. Kumar, N. Goyal and V. Gaur (2019), 'How financial literacy and demographic variables relate to behavioral biases', *Managerial Finance*, **45**(1): 124–146.

Baker, S. M., J. W. Gentry and T. L. Rittenburg (2005), 'Building understanding of the domain of consumer vulnerability', *Journal of macromarketing*, **25**(2): 128–139.

Barakat, M. A. (2019), 'A proposed model for factors affecting consumers' impulsive buying tendency in shopping malls', *Journal of Marketing Management*, **7**(1): 120–134.

Bar-Gill, O. 2012, 'Seduction by Contract', (OUP 2012).

Barros, L., T. Klein, A. Shchepetova and T. Hogg (2022), 'The rise of dark patterns: does competition law make it any brighter?', *Competition Law Journal*, **21**(3): 136–144.

Baumgartner, D. and J. Kolassa (2023), 'Power considerations for Kolmogorov–Smirnov and Anderson–Darling two-sample tests', *Communications in Statistics-Simulation and Computation*, **52**(7): 3137–3145.

Bellante, D. and C. A. Green (2004), 'Relative risk aversion among the elderly', *Review of Financial Economics*, **13**(3): 269–281.

Berger, V. W. and Y. Y. Zhou (2014), 'Kolmogorov–Smirnov test: overview'.

Bhoot, M. A., A. M. Shinde and P. W. Mishra (2020, November), 'Towards the identification of dark patterns: an analysis based on end-user reactions', In *Proceedings of the 11th Indian Conference on Human-Computer Interaction*, 24–33.

Bielova, N., C. Santos and C. M. Gray (2023), 'Two worlds apart! Closing the gap between regulating EU consent and user studies', *Harvard Journal of Law & Technology*, **37**(3).

Bongard-Blanchy, K., A. Rossi, S. Rivas, S. Doublet, V. Koenig and G. Lenzini (2021, June), '" 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*, 763–776.

Brambor, T., W. R. Clark and M. Golder (2006), 'Understanding interaction models: improving empirical analyses', *Political Analysis*, **14**(1): 63–82.

Buchan, N. R., R. T. Croson and S. Solnick (2008), 'Trust and gender: an examination of behavior and beliefs in the Investment Game', *Journal of Economic Behavior & Organization*, **68**(3-4): 466–476.

Bucher-Koenen, T., A. Lusardi, R. Alessie and M. van Rooij (2017), 'How financially literate are women? An overview and new insights', *Journal of Consumer Affair*, **51**(2): 255–283.

Calo, R. (2014), 'Digital market manipulation', *George Washington Law Review*, **82**(4): 995–1051.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'donoghue and M. Rabin (2003), 'Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism"', *University of Pennsylvania Law Review*, **151**(3): 1211–1254.

Caner, A. and C. Okten (2010), 'Risk and career choice: evidence from Turkey', *Economics of Education Review*, **29**(6): 1060–1075.

Casacci, S. and A. Pareto (2015), 'Methods for quantifying ordinal variables: a comparative study', *Quality and Quantity*, **49**, 1859–1872.

CMA (2020), 'Online platforms and digital advertising.– Appendix Y: Choice architecture and Fairness by Design' Market study final report', [1 July 2020].

CMA (2022), 'Online choice architecture: how digital design can harm competition and consumers'.

CMA (2023), 'CMA calls on Emma Sleep to Change its Online Sales Practices', Press Release, [7 July 2023].

CMA and ICO (2023), 'Harmful design in digital markets: how Online Choice Architecture practices can undermine consumer choice and control over personal information'.

Cohen, A. and L. Einav (2007), 'Estimating risk preferences from deductible choice', *American Economic Review*, **97**(3): 745–788.

Cole, A. (2016), 'All of us are vulnerable, but some are more vulnerable than others: the political ambiguity of vulnerability studies, an ambivalent critique', *Critical Horizons*, **17**(2): 260–277.

Costa, E. and D. Halpern (2019), 'The behavioural science of online harm and manipulation, and what to do about it (accessed 23 Nov. 2024), https://lavender-lyrebird-381816.hostingersite.com/media-bucket/2019/07/BIT_The-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it_Single-2.pdf'

Croson, R. and U. Gneezy (2009), 'Gender differences in preferences', *Journal of Economic Literature*, **47**(2): 448–474.

Cupák, A., P. Fessler, A. Schneebaum and M. Silgoner (2018), 'Decomposing gender gaps in financial literacy: new international evidence', *Economics Letters*, **168**, 102–106.

Cyr, D. and C. Bonanni (2005), 'Gender and website design in e-business', *International Journal of Electronic Business*, **3**(6): 565–582.

de Holanda Coelho, G. L., P. H. P. Hanel and L. J. Wolf (2020), 'The very efficient assessment of need for cognition: developing a six-item version', *Assessment*, **27**(8): 1870–1885.

Deon, T. (2011), 'The prevalence of impulsive, compulsive and innovative shopping behaviour in the economic retail hub of South Africa: a marketing segmentation approach', *African Journal of Business Management*, **5**(14): 5424–5434.

Di Geronimo, L., L. Braz, E. Fregnan, F. Palomba and A. Bacchelli (2020, April). UI dark patterns and where to find them: a study on mobile applications and user perception. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (1–14).

Dijksterhuis, A., P. K. Smith, R. B. Van Baaren and D. H. Wigboldus (2005), 'The unconscious consumer: effects of environment on consumer behavior', *Journal of Consumer Psychology*, **15**(3): 193–202.

Elling, S., L. Lentz, M. de Jong and H. Van den Bergh (2012), 'Measuring the quality of governmental websites in a controlled versus an online setting with the 'Website Evaluation Questionnaire'', *Government Information Quarterly*, **29**(3): 383–393.

ESMA (2022), 'Final Report on the European Commission Mandate on Certain Aspects Relating to Retail Investor Protection'.

European Commission (2022), 'Behavioural Study on Unfair Commercial Practices in the Digital Environment', Final Report.

European Parliament, 'Vulnerable consumers', Briefing, (2021), https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690619/EPRS_BRI(2021)690619_EN.pdf [23 November 2024].

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman and U. Sunde (2018), 'Global evidence on economic preferences', *The Quarterly Journal of Economics*, **133**(4): 1645–1692.

Farag, H. and C. Mallin (2018), 'The influence of CEO demographic characteristics on corporate risk-taking: evidence from Chinese IPOs', *The European Journal of Finance*, **24**(16): 1528–1551.

Fisch, J. E. and J. S. Seligman (2022), 'Trust, financial literacy, and financial market participation', *Journal of Pension Economics & Finance*, **21**(4): 634–664.

Frazier, M. L., P. D. Johnson and S. Fainshmidt (2013), 'Development and validation of a propensity to trust scale', *Journal of Trust Research*, **3**(2): 76–97.

FTC, 'Bringing dark patterns to light', Staff Report, (2022), https://www.ftc.gov/reports/bringing-dark-patterns-light [23 November 2024].

Gangai, D. K. N. and R. Agrawal (2016), 'The influence of personality traits on consumer impulsive buying behaviour', *International Journal of Marketing and Business Communication*, **5**(1): 35–42.

Gasiorowska, A. (2014), 'The relationship between objective and subjective wealth is moderated by financial control and mediated by money anxiety', *Journal of Economic Psychology*, **43**, 64–74.

Ghani, U., M. Imran and F. A. Jan (2011), 'The impact of demographic characteristics on impulse buying behavior of urban consumers in Peshawar', *International Journal of Academic Research*, **3**(5): 286–289.

Glaeser, E. L., D. I. Laibson, J. A. Scheinkman and C. L. Soutter (2000), 'Measuring trust', *The Quarterly Journal of Economics*, **115**(3): 811–846.

González-Cabrera, C., C. Ugalde, C. Figueroa and J. Pesántez (2019). The impact of media literacy on the intention to share fake information in social networks. In *EDULEARN19 Proceedings* (7123–7131). IATED.

Gray, C. M., Y. Kou, B. Battles, J. Hoggatt and A. L. Toombs (2018, April). The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (1–14).

Gray, C. M., C. Santos and N. Bielova (2023, April). Towards a Preliminary Ontology of Dark Patterns Knowledge. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (1–9).

Gray, C. M., C. Santos, N. Bielova, M. Toth and D. Clifford (2021, May). Dark patterns and the legal requirements of consent banners: an interaction criticism perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (1–18).

Gui, M. and G. Argentin (2011), 'Digital skills of internet natives: different forms of digital literacy in a random sample of northern Italian high school students', *New Media & Society*, **13**(6): 963–980.

Guiso, L. and M. Paiella (2008), 'Risk aversion, wealth, and background risk', *Journal of the European Economic Association*, **6**(6): 1109–1150.

Gutierrez, B. P. B. (2004), 'Determinants of planned and impulse buying: the case of the Philippines', *Asia Pacific Management Review*, **9**(6): 1061–1078.

Hainmueller, J., J. Mummolo and Y. Xu (2019), 'How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice', *Political Analysis*, **27**(2): 163–192.

Hanson, J. and D. Kysar (1999), 'Taking behvioralism seriously: the problem of market manipulation', *NYUL Rev*, **74**, 630.

Hargittai, E. (2010), 'Digital na(t)ives? Variation in internet skills and uses among members of the "net generation"', *Sociological Inquiry*, **80**(1): 92–113.

Hargittai, E. and A. Hinnant (2008), 'Digital inequality: Differences in young adults' use of the Internet', Communication *Research*, **35**(5): 602–621.
Hargittai, E. and S. Shafer (2006), 'Differences in actual and perceived online skills: the role of gender', *Social Science Quarterly*, **87**(2): 432–448.
Harrison, G. W., M. I. Lau and E. E. Rutström (2007), 'Estimating risk attitudes in Denmark: a field experiment', *Scandinavian Journal of Economics*, **109**(2): 341–368.
Helberger, N., M. Sax, J. Strycharz and H. W. Micklitz (2021), 'Choice architectures in the digital economy: towards a new understanding of digital vulnerability', *Journal of Consumer Policy* **45**, 1–26.
Hemerik, J. and J. Goeman (2018), 'Exact testing with random permutations', *Test*, **27**(4): 811–825.
Holt, C. A. and S. P. Sullivan (2023), 'Permutation tests for experimental data', *Experimental Economics*, **26**(4): 775–812.
Hunter, R. F., J. Tang, G. Hutchinson, S. Chilton, D. Holmes and F. Kee (2018), 'Association between time preference, present-bias and physical activity: implications for designing behavior change interventions', *BMC Public Health*, **18**, 1–12.
Jianakoplos, N. A. and A. Bernasek (1998), 'Are women more risk averse?', *Economic Inquiry*, **36**(4): 620–630.
Johnson, S. E., J. A. Richeson and E. Finkel (2011), 'Middle-class yet marginal? The influence of socio-economic status at an elite university on executive functioning', *Journal of Personality and Social Psychology*, **100**(5): 838–852.
Kacen, J. J. and J. A. Lee (2002), 'The influence of culture on consumer impulsive buying behavior', *Journal of Consumer Psychology*, **12**(2): 163–176.
Kahneman, D. and A. Tversky (1996), 'On the reality of cognitive illusions', *Psychological Review*, **103**(3): 582–591.
Khalifa, M. and V. Liu (2007), 'Online consumer retention: contingent effects of online shopping habit and online shopping experience', *European Journal of Information Systems*, **16**, 780–792.
King, G. and L. Zeng (2006), 'The dangers of extreme counterfactuals', *Political Analysis*, **14**(2): 131–159.
Krol, L. R. (2021), 'permutation Test', https://github.com/lrkrol/permutationTest
Lai, C. W. (2010), 'How financial attitudes and practices influence the impulsive buying behavior of college and university students', *Social Behavior and Personality: An International Journal*, **38**(3): 373–380.
Levin, I. P., M. A. Snyder and D. P. Chapman (1988), 'The interaction of experiential and situational factors and gender in a simulated risky decision-making task', *The Journal of Psychology*, **122**(2): 173–181.
Levine (2022), 'Remarks of Bureau of consumer protection director Samuel Levine as prepared for delivery', BEUC: The European Consumer Organisation To Empower, Not to Weaken: Rethinking Consumer Protection in the Digital Age, [27 September 2022].
Lian, J. W. and D. C. Yen (2014), 'Online shopping drivers and barriers for older adults: age and gender differences', *Computers in Human Behavior*, **37**, 133–143.
Limayem, M. and S. G. Hirt (2003), 'Force of habit and information systems usage: theory and initial validation', *Journal of the Association for Information Systems*, **4**(3): 65–97.
Lin, F. T. (2009), 'Does the risk aversion vary with different background risk of households', *International Research Journal of Finance and Economics*, **34**(34): 69–82.
Livingstone, S. and E. Helsper (2010), 'Balancing opportunities and risks in teenagers' use of the internet: the role of online skills and internet self-efficacy', *New Media & Society*, **12**(2): 309–329.
Luguri, J. and L. J. Strahilevitz (2021), 'Shining a light on dark patterns', *Journal of Legal Analysis*, **13**, 43.
Luo, S., B. Gu, X. Wang and Z. Zhou (2018, April). Online compulsive buying behavior: the mediating role of self-control and negative emotions. In *Proceedings of the 2018 1st International Conference on Internet and e-Business* (65–69).
Lusardi, A. and O. S. Mitchell (2008), 'Planning and financial literacy: how do women fare?', *American Economic Review*, **98**(2): 413–417.
Lusardi, A. and O. S. Mitchell (2011), 'Financial literacy around the world: an overview', *Journal of Pension Economics & Finance*, **10**(4): 497–508.
Lusardi, A. and O. S. Mitchell (2014), 'The economic importance of financial literacy: theory and evidence', *American Economic Journal: Journal of Economic Literature*, **52**(1): 5–44.
Marchiori, D. R., M. A. Adriaanse and D. T. De Ridder (2017), 'Unresolved questions in nudging research: putting the psychology back in nudging', *Social and Personality Psychology Compass*, **11**(1): e12297.

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 11K Shopping Websites', In *Proceedings of the ACM on Human-Computer Interaction*, **3**(CSCW), 1–32.

Mathur, A., A. Narayanan and M. Chetty (2018), 'Endorsements on Social Media: an Empirical Study of Affiliate Marketing Disclosures on YouTube and Pinterest', In *Proceedings of the ACM on Human-Computer Interaction*.

Mathur, A., M. Kshirsagar and J. Mayer (2021), 'What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods', In *Proceedings of the 2021 CHI conference on human factors in computing systems*, (pp. 1–18).

Mihic´, M. and I. Kursan (2010), 'Influence of demographic and individual difference factors on impulse buying', *Market*, **22**(1): 7–28.

Mills, S., R. Whittle, R. Ahmed, T. Walsh and M. Wessel (2023), 'Dark patterns and sludge audits: an integrated approach', *Behavioural Public Policy*, 1–27.

Mishra, S. and M. L. Lalumière (2010), 'You can't always get what you want: the motivational effect of need on risk-sensitive decision-making', *Journal of Experimental Social Psychology*, **46**(4): 605–611.

Moran, N. (2020), 'Illusion of safety: how consumers underestimate manipulation and deception in online (vs. offline) shopping contexts', 54 J Consum Aff 890.

Narayanan, A., A. Mathur, M. Chetty and M. Kshirsagar (2020), 'Dark patterns: past, present, and future: the evolution of tricky user interfaces', *Queue*, **18**(2): 67–92.

Nouwens, M., I. Liccardi, M. Veale, D. Karger and L. Kagal (2020 April), 'Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating their Influence', In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

OECD (2022), 'Dark commercial patterns', OECD Digital Economy Papers, October 2022 No. 336.

OECD (2023), 'Consumer vulnerability in the digital age', OECD Digital Economy Papers, No. 355.

Outreville, J. F. (2015), 'The relationship between relative risk aversion and the level of education: a survey and implications for the demand for life insurance', *Journal of Economic Surveys*, **29**(1): 97–111.

Özdemir, Ş. (2020), 'Digital nudges and dark patterns: the angels and the archfiends of digital communication', *Digital Scholarship in the Humanities*, **35**(2): 417–428.

Peretti-Watel, P., O. L'haridon and V. Seror (2013), 'Time preferences, socioeconomic status and smokers' behaviour, attitudes and risk awareness', *The European Journal of Public Health*, **23**(5): 783–788.

Potrich, A. C. G., K. M. Vieira and G. Kirch (2018), 'How well do women do when it comes to financial literacy? Proposition of an indicator and analysis of gender differences', *Journal of Behavioral and Experimental Finance*, **17**, 28–41.

Powell, M. and D. Ansic (1997), 'Gender differences in risk behaviour in financial decision-making: an experimental analysis', *Journal of Economic Psychology*, **18**(6): 605–628.

Preston, A. C. and R. E. Wright (2019), 'Understanding the gender gap in financial literacy: evidence from Australia', *Economic Record*, **95**, 1–29.

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).

Riley, W. B., Jr and K. V. Chow (1992), 'Asset allocation and individual risk aversion', *Financial Analysts Journal*, **48**(6): 32–37.

Rodríguez-de-Dios, I., J. J. Igartua and A. González-Vázquez (2016 November), 'Development and Validation of a Digital Literacy Scale for Teenagers', In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 1067–1072.

Rodríguez-de-Dios, I., J. M. van Oosten and J. J. Igartua (2018), 'A study of the relationship between parental mediation and adolescents' digital skills, online risks and online opportunities', *Computers in Human Behavior*, **82**, 186–198.

Roffarello, M. A., K. Lukoff and L. de Russis (2023 April), 'Defining and Identifying Attention Capture Deceptive Designs in Digital Interfaces', In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19.

Rook, D. W. and S. J. Hoch (1985), 'Consuming impulses', *Advances in Consumer Research*, **12**(1): 23–27.

Rosen, A. B., J. S. Tsai and S. M. Downs (2003), 'Variations in risk attitude across race, gender, and education', *Medical Decision Making*, **23**(6): 511–517.

Santini, F. D. O., W. J. Ladeira, V. A. Vieira, C. F. Araujo and C. H. Sampaio (2019), 'Antecedents and consequences of impulse buying: a meta-analytic study', *RAUSP Management Journal*, **54**, 178–204.

Schooley, D. K. and D. D. Worden (1996), 'Risk aversion measures: comparing attitudes and asset allocation', *Financial Services Review*, **5**(2): 87–99.

Selwyn, N. (2004), 'The information aged: a qualitative study of older adults' use of information and communications technology', *Journal of Aging Studies*, **18**(4): 369–384.

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*, 1–27.

Smith, N. C. and Cooper-Martin, E. (1997), 'Ethics and target marketing: The role of product harm and consumer vulnerability', *Journal of marketing*, **61**(3): 1–20.

Smith, S. M. and I. P. Levin (1996), 'Need for cognition and choice framing effects', *Journal of Behavioral Decision Making*, **9**(4): 283–290.

Spencer, S. (2020), 'The Problem of Online Manipulation', *U Ill L Rev*, **959**, 990.

Stango, V. and J. Zinman (2020), 'We are all behavioral, more or less: a taxonomy of consumer decision making', NBER Working Paper No. 28138, (November 2020).

Stella, G. P., U. Filotto, E. M. Cervellati and E. A. Graziano (2020), 'The effects of financial education on financial literacy in Italy', *International Business Research*, **13**(4): 44–51.

Stucke, M. and A. Ezrachi (2021), *Competition Overdose— How Free Market Mythology Transformed Us from Citizen Kings to Market Servants*, New York, US: Harper Collins.

Styvén, M. E. K., T. Foster and Å. Wallström (2017), 'Impulse buying tendencies among online shoppers in Sweden', *Journal of Research in Interactive Marketing*, **11**(4): 416–431.

Susser, B. R. and H. Nisssenbaum (2019), 'Online manipulation: hidden influence in a digital world', *Geo L Tech Rev*, **4**, 1–45.

Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Connecticut, US: Yale University Press.

Tripp, L. M. (2011), "The computer is not for you to be looking around, it is for schoolwork': challenges for digital inclusion as Latino immigrant families negotiate children's access to the internet', *New Media & Society*, **13**(4): 552–567.

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

Tversky, A. and D. Kahneman (1981), 'The framing of decisions and the psychology of choice', *Science*, **211**(4481): 453–458.

Uslaner, E. M. (2002), *The Moral Foundations of Trust*, Cambridge, UK: Cambridge University Press.

Utz, C., M. Degeling, S. Fahl, F. Schaub and T. Holz (2019 November), '(Un) Informed Consent: Studying GDPR Consent Notices in the Field', In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 973–990.

van Deursen, A. J., E. J. Helsper and R. Eynon (2016), 'Development and validation of the Internet Skills Scale (ISS)', *Information, Communication & Society*, **19**(6): 804–823.

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

Weun, S., M. A. Jones and S. E. Beatty (1998), 'Development and validation of the impulse buying tendency scale', *Psychological Reports*, **82**(3): 1123–1133.

Wibowo, A. and S. Indartono (2017), 'The role of consumer self-efficacy on the effect of demographics background on impulse behavior', *Journal of Business*, **1**(1): 73–84.

Wood, M. (1998), 'Socio-economic status, delay of gratification, and impulse buying', *Journal of Economic Psychology*, **19**, 295–320.

Wu, C. H., S. K. Parker and J. P. de Jong (2014), 'Need for cognition as an antecedent of individual innovation behavior', *Journal of Management*, **40**(6): 1511–1534.

Zarsky, Z. (2019), 'Privacy and manipulation in the digital age', 20 Theoretical Inquiries in Law 157.

Zhang, F., N. J. Yuan, K. Zheng, D. Lian, X. Xie and Y. Rui (2015 September), 'Mining Consumer Impulsivity from Offline and Online Behavior', In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1281–1292).

Zimic, S. (2009), 'Not so 'techno-savvy': Challenging the stereotypical images of the 'Net generation'', *Digital Culture & Education*, **1**(2): 129–144.

Appendix A.

Table A1. Common examples of dark patterns

Name of dark pattern	Description
Activity notifications/messages	Messages which claim that other consumers are viewing the same products, which may be false or misleading.
Bait and switch	Consumer is offered an initial price, which is subsequently increased at a later stage of the transaction.
Comparison prevention	Essential information about products is made hard to find or obscured so that a consumer cannot make a meaningful comparison.
Countdown timer/‘exploding offers’/fake urgency	A consumer is led to believe that a deal or price will expire, which may be false or misleading.
Disguised ads	Where a consumer is enticed to click on something which is actually a concealed ad.
Fake scarcity/low stock messages	Indication of limited stock or product availability, which may be false or misleading.
Forced registration	The consumer is required to register with a website and provide personal details before being able to make a purchase.
Forced disclosure	A consumer is tricked into sharing more information/data than needed.
Hard to cancel	A consumer is encouraged to subscribe to a service, but finds it difficult to withdraw or unsubscribe at a later date.
Hidden costs	Non-optional costs are disclosed late in a transaction or only added to the price at the final stage of a transaction.
Hidden subscription	Interfaces which involve automatic or unanticipated renewal or subscription to a service.
Immortal accounts	Accounts containing consumer information which cannot be deleted or where the consumer cannot unsubscribe.
Misleading reference pricing	Prices or discounts shown are based on a misleading reference price
Nagging	The consumer receives repeated requests to take an action the online provider wants.
Preselection	A default option is preselected by the online provider.
Sneaking	Interfaces which add products to consumers’ shopping basket without their consent.
Testimonials/fake social proof	Purported statements by other consumers about the popularity of quality of a product which can be false or misleading.

Source: Based on OECD (2022) and the examples at https://www.deceptive.design/types

Table A2. Personal questionnaire variables

Variable name	Explanations
Demographics	
Education (1–6)	Six levels: GCSE; A-level; Certificate/Diploma of higher education; Undergraduate degree; Master's degree; Doctoral degree
Education_Interval	Education data were transformed from an ordinal data into an interval data using the approach suggested by Casacci and Pareto (2015). In short, we estimated the probability density function (pdf) of the boundaries of each education category and then computed the 'average' pdf value (i.e., (pdf of lower bound – pdf of upper bound)/percentage of the cases in that category) for each category.
Income (1–10)	Ten monthly household income levels: £0 to £500; £501 up to £1000; £1001 up to £1500; £1501 up to £2000; £2001 up to £2500; £2501 up to £3000; £3001 up to £4000; £4001 up to £5000; £5001 up to £7000; £7001 and more
Income_Interval	Like education data, objective income data were converted into interval data in a similar way. The only difference is that since the boundary values of categories are meaningful, we fitted the mean of each category (namely, 250, 750, 1250, 1750, 2250, 2750, 3500, 4500, 6000 and 8000) into a normal distribution when estimating the pdf values.
Income_Subjective	Participants' subjective assessments about their financial situations. It contains three questions created based on Gasiorowska (2014) and Stella *et al.* (2020). Each question was answered on a five-point scale, 1 being 'strongly disagree' and 5 being 'strongly agree'. The sum of the responses should range from 3 to 15.
Income_Relative	Participants' self-assessments about their position in the UK in terms of their household income. In this question, participants needed to choose a person from a picture of a group of (twenty) people, ordered in terms of their household income. The leftmost person has the least, and the rightmost person has the most.
Covid	Impact of Covid-19 on household income (on a five-point scale) last year.
Housepeople	Numbers of people in the household; Participants were given options, ranging from 1 to 6+. However, some participants who chose 6+ did not indicate the exact number of household members in the entry box. Due to lack of data, we decided not to use this information in the analysis.
Age	Self-reported age collected from a free entry question
Gender	(a) Three types: Male; Female; Non-binary. (b) Participants were also given the option: 'Prefer not to say'. (c) For convenience, in the statistical analysis, female is coded as 1 and male is coded as 2. Non-binary and not disclosure data were not included in the analysis when it comes to gender, due to the small sample size.
Measures for personality characteristics	
Impulsive	(a) This variable measured participants' tendency to exhibit an unplanned, compelling and hedonic purchasing behaviour. (b) The five-item impulse buying tendency scale, developed by Weun *et al.* (1998) was used. (c) Internal consistency: 0.8753

(*Continued*)

Table A2. (*Continued.*)

Variable name	Explanations
Digital	(a) This variable elicited each participant's digital literacy. (b) Six items were selected from Digital Literacy Scale (Rodríguez-de-Dios *et al.*, 2016), based on their factor loading and relevance to our study. (c) To pick up the best items, we jointly evaluate results obtained from González-Cabrera *et al.* (2019), Rodríguez-de-Dios *et al.* (2016), Rodríguez-de-Dios *et al.* (2018) and van Deursen *et al.* (2016). (d) Internal consistency: 0.6906
Shopping	(a) This variable measured an individual participant's online shopping experience. (b) Three questions employed from Khalifa and Liu (2007)'s questionnaire, developed based on Limayem and Hirt's (2003) scale. (c) Internal consistency: 0.8137
Trust	(a) This variable refers to participants' propensity to trust. (b) We employed the four-item propensity to trust scale, developed by Frazier *et al.* (2013). (c) Internal consistency: 0.9205
NFC	(a) This variable measured participants' degrees of need for cognition (i.e., their tendency to engage in thinking and enjoy solving complex problems). (b) The six-item need for cognition scale, developed by de Holanda Coelho *et al.* (2020), was adopted in the experiment. (c) Internal consistency: 0.8840
FinLiteracy	(a) This variable tested participant's financial literacy. (b) Four questions were selected from a financial knowledge questionnaire, proposed by Stella *et al.* (2020) and Atkinson and Messy (2011), based on factor loading, difficulties and relevance to our product. (c) Selected questions range from a beginner's level to an advanced level with an intention to assess the objective ability of the participant to understand concepts like interest rates and inflation. (d) Internal consistency: 0.4214
RiskPref	(a) This variable measured participants' risk preferences. (b) A self-report question regarding risk preference over a 10-point scale, proposed by The Global Preference Survey (Falk *et al.*, 2018), was used. (c) The higher the value is, the more willing to take a risk the participant is.
TimePref	(a) This variable elicited participants' time preferences. (b) Similar to risk preferences, one self-reported question on time preferences over a 10-point scale, proposed by The Global Preference Survey (Falk *et al.*, 2018), was employed. (c) The higher the value is, the more patient the participant is.

Appendix B. Correlation analysis

Table A3 compares the responses to the personal questionnaire with four demographic variables (age, gender, education and income) and previous studies.

Table A3. Personal characteristics and they key demographics

Demo\Measure	Impulsive	Digital	Shopping	Trust	NFC	FinLit	RiskP	TimeP
Age	−0.1196***	−0.2606***	0.0221	0.0929***	0.0980***	0.1639***	−0.1005***	0.0350*
Gender	−0.2212***	0.1131***	−0.0435**	0.0387*	0.1481***	0.2427***	0.1796***	0.1453***
Education	−0.0822***	0.0843***	0.0439**	0.0493**	0.2335***	0.1939***	0.0519**	0.1338***
Income	0.0398*	0.0565**	0.1129***	0.0951***	0.1342***	0.1880***	0.1275***	0.1355***

Note: The numbers reported in Table 3 indicate coefficient values of Pearson's linear correlation between variables, with red cells marking a negative association between the variables. Note that since we have a large sample (>2000), many of the coefficients are statistically significant, but most of them are insignificant in size. We use, but do not report here, additional, more rigorous tests (i.e., a *t*-test, Wilcoxon rank sum test, ANOVA and Kruskal–Wallis test) depending on data types. The null hypothesis for *t*-test: *x* and *y* come from normal distributions with equal means and equal but unknown variances; The null hypothesis for ANOVA: data from several groups (levels) of a factor have a common mean. The results are in line with the correlation matrix. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.001$.

Trust

Our data support a clear positive correlation between propensity to trust and demographic variables such as age (Fisch and Seligman, 2022) and gender (specifically being male) (Glaeser *et al.*, 2000; Uslaner, 2002; Cyr and Bonanni, 2005; Buchan *et al.*, 2008). In addition, we find income to be significantly correlated with trust which is not well-documented in existing literature.

Experience of shopping online shopping

Our data support previous work which has found that gender to be significantly correlated with experience of online shopping. We found that females are more experienced in online shopping than male (Lian and Yen, 2014). However, in contrast with previous research which found that young people are the most active online shoppers (Selwyn, 2004; Lian and Yen, 2014), our data suggest no statistically significant correlation between age and experience with online shopping.

Digital literacy

Our data on digital literacy are consistent with the findings of existing studies that find that higher levels of digital literacy are correlated with being younger[26] (Tripp, 2011; van Deursen *et al.*, 2016), gender (specifically being male)[27] (van Deursen *et al.*, 2016) and higher levels of educational attainment (Gui and Argentin, 2011; van Deursen *et al.*, 2016). Our results also suggest a positive correlation between income and digital literacy, an area which has received relatively little attention to date.[28]

[26]It is important to note that we adopted a self-assessment method to measure digital literacy, which is highly influenced by people's perceptions on their own ability. There are some research findings showing that younger participants did not outperform older ones in all aspects (Livingstone and Helsper, 2010; Zimic, 2009). Especially when it comes to information and evaluation skills, younger participants usually performed worse (Gui and Argentin, 2011).

[27]This might be because men are more likely to rate their abilities higher than women (Hargittai and Hinnant, 2008; Hargittai and Shafer, 2006). Some studies (e.g., Gui and Argentin, 2011) found that male students overall did not perform better than females in digital skills.

[28]Hargittai (2010) showing that higher social economic status is linked to greater internet skills, which is also consistent with our results.

Need for cognition (NFC)

Our data show a positive correlation between cognitive abilities and age, gender (specifically being male), level of educational attainment, and income. However, the existing literature only supports the association between cognitive abilities and education (e.g., Wu *et al.*, 2014; de Holanda Coelho *et al.*, 2020). Factors such as age and gender are often found to be insignificant (Smith and Levin, 1996; Wu *et al.*, 2014). Some caution should be exercised as there are only a relatively small number of papers reporting demographic influences on cognition.

Tendency to impulse buying

Impulsive buying tendency in our data is unsurprisingly negatively correlated with age, being male, and levels of educational attainment. Overall, these results align with existing research (e.g., Rook and Hoch, 1985; Wood, 1998; Kacen and Lee, 2002; Lai, 2010; Badgaiyan and Verma, 2015; Styvén *et al.*, 2017; Wibowo and Indartono, 2017) and largely echoes recent meta-analysis conclusions (e.g., Gangai and Agrawal, 2016; Santini *et al.*, 2019). It is worth noting that although we only find a weakly significant positive relationship between impulsive buying tendency and monthly household income level, past research results are also mixed on this point, with some showing a negative relationship (Wood, 1998), some positive (Mihic´ and Kursan, 2010; Deon, 2011; Barakat, 2019), and some finding no relationship (Gutierrez, 2004; Ghani *et al.*, 2011). This implies that income may be related to this variable in a more specific or complex manner; for example, financial situation at the time of purchase (Gangai and Agrawal, 2016).

Financial literacy

Our data suggest that older adults, males, highly educated people, and the rich are more financially literate. This is consistent with past literature on education (Lusardi and Mitchell, 2014), income levels (Lusardi and Mitchell, 2011), and the gender gap, which is persistent and widespread across countries (e.g., Lusardi and Mitchell, 2008; Bucher-Koenen *et al.*, 2017; Cupák *et al.*, 2018; Potrich *et al.*, 2018; Preston and Wright, 2019). However, we note that inconsistent results have been reported in past studies on the relationship between age and financial literacy, with some studies reporting a positive relationship (Fisch and Seligman, 2022), others a quadratic relationship (in a belled shape; Lusardi and Mitchell, 2014), while other studies report no relationship (Baker *et al.*, 2019).

Time preference

Our data suggests that male, highly educated adults, and richer people are, on average, more patient, which largely aligns with the findings of Falk *et al.* (2018). We note, however, that existing research provide no consistent conclusion on the relationship between income and time preferences, possibly due to varying measurement approaches. For example, Hunter et al. (2018) do not find any significant relationship between income and three different time preferences scales, respectively, while others (e.g., Johnson *et al.*, 2011; Peretti-Watel *et al.*, 2013) observe a greater tendency of low socio-economic status adults (vs high) towards procrastination and the present bias.

Risk preference

Our data are consistent with Falk *et al.* (2018) in finding that women and the elderly tend to be more risk averse than their counterparts.[29] Other demographic covariates such as education and income, unfortunately, tend to be under-reported. To address this we have reviewed the results of other studies which adopt various approaches to measuring risk preferences. Similar to our data, Rosen *et al.* (2003) find that lower

[29]Several other studies which show how risk preferences' interact with age (Riley and Chow, 1992; Harrison *et al.*, 2007; Farag and Mallin, 2018) or gender (Levin *et al.*, 1988; Powell and Ansic, 1997; Jianakoplos and Bernasek, 1998; Rosen *et al.*, 2003; Cohen and Einav, 2007; Croson and Gneezy, 2009; Lin, 2009).

education predicts increased risk aversion, several other studies (e.g., Riley and Chow, 1992; Schooley and Worden, 1996; Bellante and Green, 2004; Lin, 2009) present opposing results.[30] On the relationship between risk preferences and income, our data align with earlier research that identifies risk loving attitudes as an increasing function of absolute income or wealth (Riley and Chow, 1992; Guiso and Paiella, 2008; Caner and Okten, 2010), possibly because potential income risks would motivate less affluent people to play safe (Lin, 2009). We note, however, other studies that do not find a correlation between income and experimentally elicited risk preference (Ahern et al., 2014) or an increased engagement in risky economic behaviour when being allocated insufficient resources in an experiment (Mishra and Lalumière, 2010).

Appendix C. Stimuli description

Trick questions

For this treatment condition the text of the pop-up message was identical to the neutral offer condition (see Figure 2). However, the question which asked participants to accept the offer was framed in a different way with the aim of confusing participants: it used a double-negative statement in the product offer and the response buttons. Specifically, as shown in Figure A1, choosing the option 'No' actually involved accepting the offer given how the question is framed. To follow the common framing approach, and to explore more about participants' real demand, participants who declined the offer (by choosing 'Yes') would then see a neutral follow-up question (with no manipulation) which asked them if they would like to buy the product. This experimental design follows the common practice of the trick question, characterised by offering choices that require careful assessment (Mathur *et al.*, 2021; Waldman, 2020). Double negative is one of the most

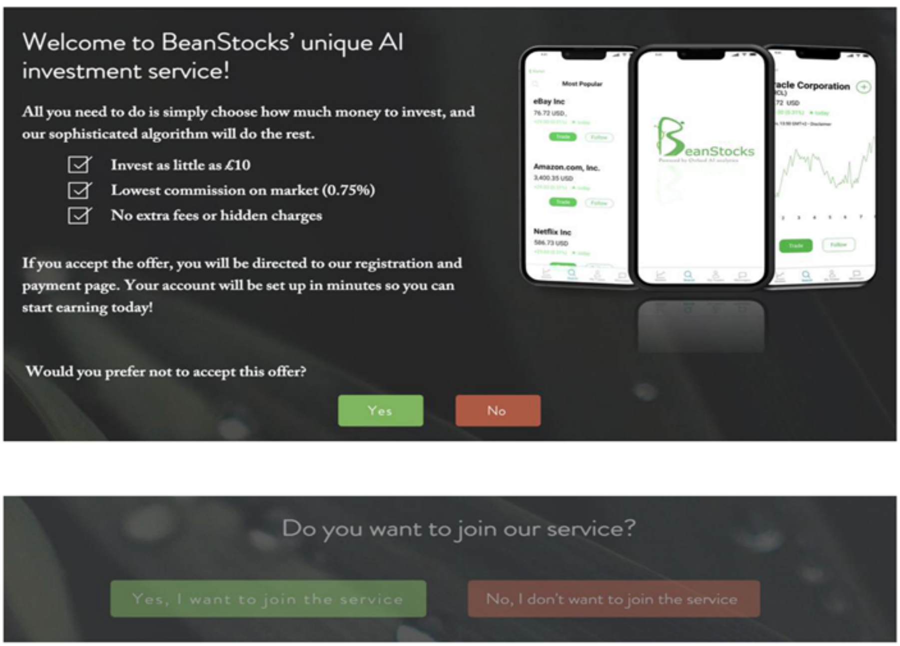


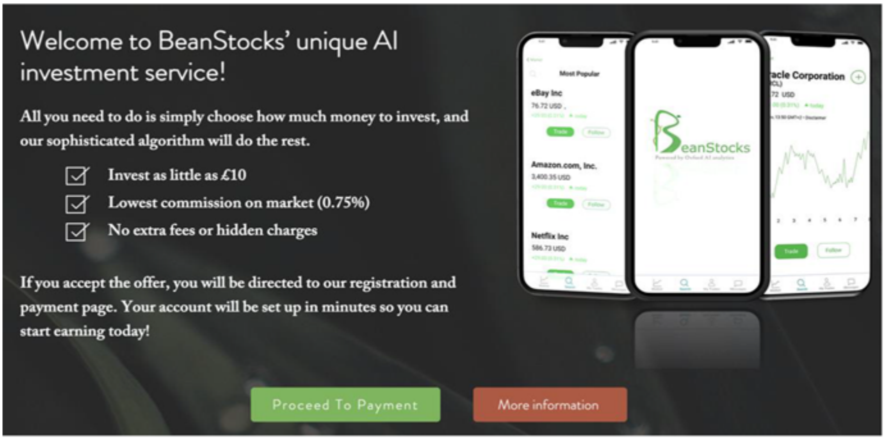
Figure A1. Trick questions treatment condition.

[30]This may be because more risk averse individuals exhibit weaker tendency to pursue a university education (Outreville, 2015).

popular forms of a trick question, while another example provided by Mathur *et al.* (2019) is the option to 'Uncheck the box if you prefer not to receive email updates'.

Roach motel

This treatment condition involved a user interface design that created a situation of 'asymmetric difficulty' for participants in choosing between two options (or among several options). In the first stage, participants see the same message as in the neutral offer condition. However, they were then presented with two options 'Proceed to payment' and 'More information' (instead of 'Proceed to payment' and 'Reject'). If they choose 'More information', they will be given another offer message about the investment product. The message presented in the second stage was different from the one in the first stage. Here, participants were given information about the benefits of the product. Participants needed to choose between 'Proceed to payment' and 'More information'. If 'More information' is chosen, participants will be asked to indicate their reasons for not accepting. At this stage, they will be given an option to accept or reject the offer.



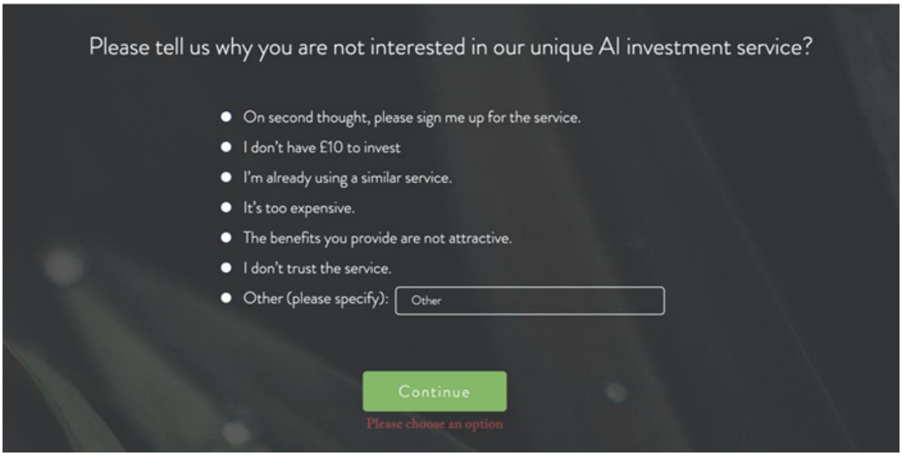


Figure A2. Roach motel treatment condition.

This experimental design followed that of Mathur, Kshirsagar, and Mayer (2021) who used a user interface offers a readily visible choice ('OK') to agree to data tracking and gathering, while the option to decline can only be accessed after navigating through a sequence of seven different displays, initiating from the 'Manage Options' section. The use of a third stage was inspired by Luguri and Strahilevitz (2021), who also asked participants to choose from a range of reasons for not accepting the offer, with a final chance to accept, at the last stage.

Confirm shaming

For this treatment condition the text of the pop-up message was identical to the neutral offer condition, but the wording of the options was different. As shown in Figure A3, participants are presented with two options: 'Yes, I want to start earning high returns!' and 'No, I don't care about securing my future and earning high returns.' Texts of both options were worded in the way to manipulate participants' emotions, making them excited about accepting the offer, and at the same time, engendering a feeling of guilt about declining. Our design is similar to the common practice of confirm shaming which frames the decline option (or any options undesired by firms) intentionally to shame or guilt users into opting in (Mathur *et al.*, 2021; Özdemir, 2020); for example, 'No thanks, I don't want Unlimited One-Day Delivery' (Mathur *et al.*, 2021), 'No thanks, I like paying full price' (Mathur *et al.*, 2019), and 'No thanks, I hate fun & games' (Mathur *et al.*, 2019).

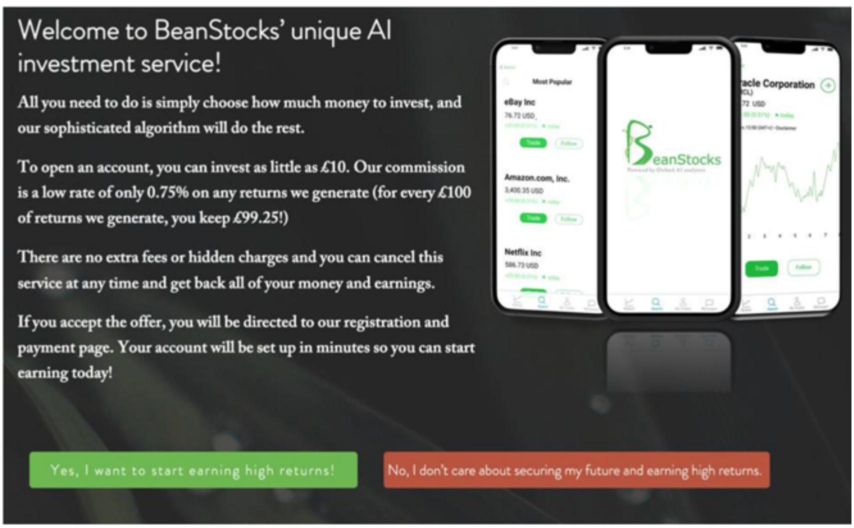


Figure A3. Confirm shaming treatment condition.

Figure A4. De-briefing messages (all conditions).

Appendix D. Randomisation checks

We performed randomisation checks to examine how demographic data were distributed across different conditions. While participants were assigned to each condition randomly, there is nevertheless a potential risk that the demographic profile of participants in the different treatment groups may show significant differences. Figure A5 shows how demographic data are distributed across the different neutral and treatment conditions according to age, education, gender and our income variables. The figures reveal that the patterns of data are highly similar across conditions, implying a strong possibility of random allocations.

A more nuanced statistical analysis was conducted following the graphical investigation. A two-sample Kolmogorov–Smirnov test was employed to examine whether there exists heterogeneity in a demographic covariate between two subgroup (i.e., condition) samples. More specifically, it was used to test if the two dark pattern conditions being compared are from the same (population) distributions over a certain demographic

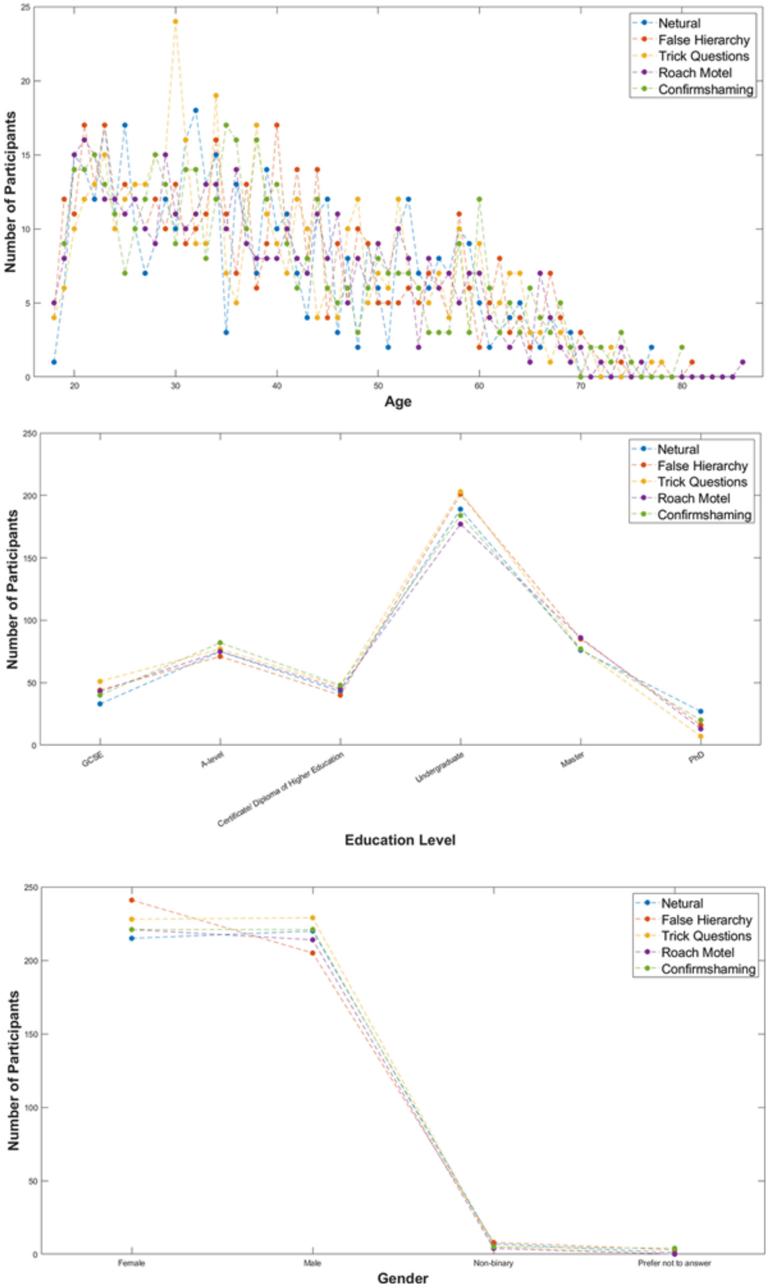


Figure A5. Number of participants in each demographic group by dark pattern conditions.

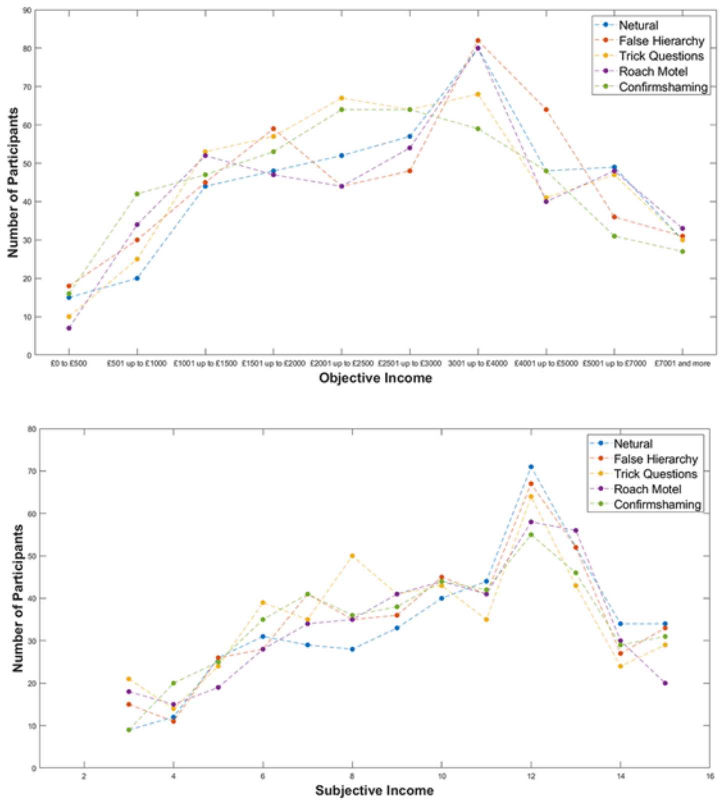


Figure A5. (*Continued.*)

Table A4. The statistical test results for the permutation test

	False hierarchy	Trick question	Roach motel	Confirm shaming	Pooled treatment
First choice					
Control	−0.4533***	0.0358	−0.2533***	−0.1337**	−0.1968***
False hierarchy		0.4906***	0.1945**	0.3147***	
Trick question			−0.2897***	−0.169**	
Roach motel				0.1186*	
Payment choice					
Control	−0.2263***	−0.1727***	−0.0590	−0.1156*	−0.0443**
False hierarchy		0.0543	0.1683**	0.1120*	
Trick question			0.1143*	0.0577	
Roach motel				−0.0569	

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

variable (Berger and Zhou, 2014). Importantly, this test is a powerful tool in situations where traditional parametric assumptions are not satisfied or when the data are skewed (Baumgartner and Kolassa, 2023). The present analysis used the function provided by Matlab.

The results imply distributional invariance, with only a few (i.e., the three income variables) revealing statistical differences between two condition samples. Full results are available upon request from the authors. In addition to checking the randomness of the participant assignment, a permutation test (a.k.a. a re-randomisation test) was employed to investigate whether the observed mean differences in acceptance rates between the control and a treatment condition are due to a chance (or luck) or a treatment effect. Like the two-sample Kolmogorov–Smirnov test used above, the permutation test is non-parametric (Holt and Sullivan, 2023) and its null hypothesis is that the two tested (independent) sample groups come from the same distribution (Hemerik & Goeman, 2018). There are many extensions and variations of this method, the present analysis used the Matlab code developed by Krol (2021) and set to repeat the permutation procedure 10,000 times for better accuracy.

Table A4 shows the test results such as effect sizes (Hedges G) and outcomes of hypothesis testing. The rejection of the null hypothesis implies that the observed mean differences in first choice (or payment choice) stem from the treatment effect (i.e., the dark pattern manipulation), instead of randomness. Accordingly, as revealed in Table A4, the statistical differences in acceptance rates reported in Table 3 in the main paper are not the result of a random chance. This further strengthens the conclusion that dark patterns form an effective strategy for increasing acceptance rates.

Appendix E. Additional results

In this section we include results for the GLM models and additional interaction models for financial literacy (Table A6). First, in the GLM (logistic model) models, the output is on link-scale (logit); thus, the numerical output of the model corresponds to the log-odds. For example, the coefficient of the treatment variable (Dark Patterns) has a numerical value of 0.139 for the first choice. Its positive sign indicates that the chance of observing a first choice increases with treatment (the three dark patterns are considered the treatment group here). The magnitude of the coefficient implies that compared to the control group, treatment results in 0.139 increase in the log-odds ratio of acceptance: rejection of the offer. Therefore, the odds for accepting the offer are 174% higher for the treatment group compared to the control.

In rows two and four in Table A5, each of the three dark pattern treatment conditions are explored separately. According to the results, the most powerful manipulation is False Hierarchy (visual interface), followed by the Roach Motel and Confirm Shaming. All three manipulations are highly effective in the first choice. The Roach Motel condition, which includes two repeated screens with no decline option for the offer (only an option for 'more information') is effective in inducing an acceptance in the first-choice phase (which is considered an 'accept' in any of the screens). But it might cause a 'backlash effect' in the payment phase, where we see acceptance levels drop, with users not being more likely to accept the offer compared to the control group.

Table A5. GLM regression full results

	Dependent variable			
	First choice		Payment choice	
	(1)	(2)	(3)	(4)
Dark patterns	0.139*** (0.027)		0.042** (0.017)	
False hierarchy		0.225*** (0.033)		0.075*** (0.021)
Roach motel		0.127*** (0.033)		0.017 (0.021)
Confirm shaming		0.063* (0.033)		0.035* (0.021)
Age	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Income	−0.0004 (0.005)	−0.001 (0.005)	0.005 (0.003)	0.005 (0.003)
Education	0.001 (0.009)	0.001 (0.009)	0.004 (0.006)	0.003 (0.006)

(*Continued*)

Table A5. (*Continued.*)

	Dependent variable			
	First choice		Payment choice	
	(1)	(2)	(3)	(4)
Risk pref	0.005 (0.006)	0.004 (0.006)	0.002 (0.004)	0.002 (0.004)
Financial literacy	0.005 (0.014)	0.007 (0.014)	0.004 (0.009)	0.004 (0.009)
Digital literacy	0.001 (0.001)	0.001 (0.001)	0.002** (0.001)	0.002** (0.001)
Shopping exp	0.006 (0.006)	0.004 (0.006)	−0.002 (0.004)	−0.002 (0.004)
Impulsive	0.004 (0.003)	0.005 (0.003)	0.002 (0.002)	0.002 (0.002)
Time pref	0.010 (0.007)	0.011 (0.007)	−0.0004 (0.004)	−0.0004 (0.004)
Gender (male)	0.030 (0.025)	0.033 (0.025)	0.037** (0.016)	0.038** (0.016)
Trust	−0.0005 (0.003)	−0.0001 (0.003)	−0.001 (0.002)	−0.001 (0.002)
NFC	0.001 (0.003)	0.002 (0.003)	−0.0002 (0.002)	−0.0002 (0.002)
Observations	1,758	1,758	1,758	1,758
Akaike Inf. Crit.	2,483.369	2,462.254	899.572	895.498

Note: All variables are scaled the mean.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A6. LM interaction results, financial literacy

	Regression results			
	Dependent variable			
	First choice		Payment choice	
	(1)	(2)	(3)	(4)
Dark patterns	0.138*** (0.027)		0.041** (0.017)	
False hierarchy		0.227*** (0.033)		0.076*** (0.021)
Roach motel		0.127*** (0.033)		0.016 (0.021)
Confirm shaming		0.062* (0.033)		0.034 (0.021)
Financial literacy	−0.012 (0.026)	−0.012 (0.026)	−0.038** (0.017)	−0.038** (0.017)
Income	−0.001 (0.005)	−0.002 (0.005)	0.005 (0.003)	0.004 (0.003)
Education	0.001 (0.009)	0.001 (0.009)	0.003 (0.006)	0.004 (0.006)
Risk pref	0.005 (0.006)	0.004 (0.006)	0.002 (0.004)	0.002 (0.004)
Age	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Digital literacy	0.001 (0.001)	0.001 (0.001)	0.002** (0.001)	0.002** (0.001)
Shopping exp	0.006 (0.006)	0.004 (0.006)	−0.002 (0.004)	−0.003 (0.004)

(*Continued*)

Table A6. (*Continued.*)

Regression results				
	Dependent variable			
	First choice		Payment choice	
	(1)	(2)	(3)	(4)
Impulsive	0.004 (0.003)	0.005* (0.003)	0.002 (0.002)	0.002 (0.002)
Time pref	0.010 (0.007)	0.013* (0.007)	−0.0002 (0.004)	0.0002 (0.004)
Gender (male)	0.030 (0.025)	0.032 (0.025)	0.038** (0.016)	0.038** (0.016)
Trust	−0.0004 (0.003)	−0.0003 (0.003)	−0.001 (0.002)	−0.001 (0.002)
NFC	0.001 (0.003)	0.001 (0.003)	−0.0002 (0.002)	−0.0003 (0.002)
Dark patterns: financial literacy	0.023 (0.029)		0.055*** (0.019)	
False hierarchy: financial literacy		0.071* (0.036)		0.103*** (0.023)
Roach motel: financial literacy		−0.042 (0.035)		0.033 (0.022)
Confirm shaming: financial literacy		0.056 (0.035)		0.039* (0.023)
Observations	1,758	1,758	1,758	1,758
R^{200}	0.022	0.042	0.018	0.029
Adjusted R^{200}	0.014	0.033	0.010	0.019

Note: All variables are scaled the mean.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

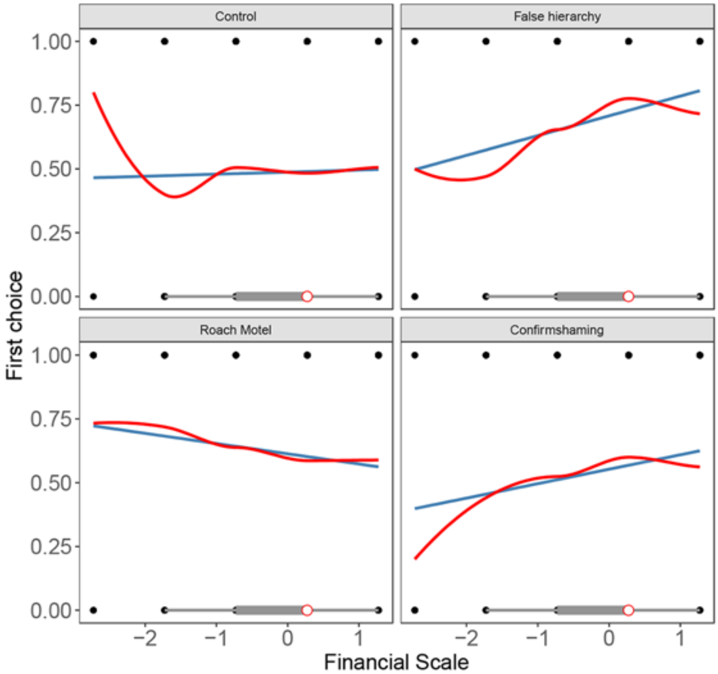


Figure A6. Marginal effect size by financial literacy – raw plot.
Note: Interaction model, showing moderation effects of financial literacy in blue. The red line represents a LOESS (locally estimated scatterplot smoothing) approximation. The bottom bar in each figure shows the distribution of the moderator financial literacy when scaled to the average. The top left panel shows results for the control group, while the other cells show the three dark patterns groups.

Cite this article: Zac A, Huang YC, von Moltke A, Decker C and Ezrachi A (2025), 'Dark patterns and consumer vulnerability', *Behavioural Public Policy*, 1–50. https://doi.org/10.1017/bpp.2024.49