Journal of Economic Behavior and Organization 244 (2026) 107469

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

The transaction test: An experimental method for assessing online interfaces

Francesco Bogliacino [a,*], Leonardo Pejsachowicz [b], Giovanni Liva [c], Francisco Lupiáñez-Villanueva [d]

[a] Dipartimento di Politica Economica, Università Cattolica del Sacro Cuore, Via Necchi 5, 20123 Milan, Italy
[b] Centre d'Économie de la Sorbonne, Université Paris I Pantheon-Sorbonne, 106-112 Boulevard de l'Hopital, 75013 Paris, France
[c] Open Evidence, Carrer Pujades 51-55, 08005 Barcelona, Spain
[d] Information and Communication Department, Universitat Oberta de Catalunya, Poblenou, 154-156, 08018 Barcelona, Spain

ARTICLE INFO

JEL Classification:
C92
D18
D91
L13

Keywords:
Transaction test
Dark patterns
Manipulation
Inconsistency
Time-pressure
Data privacy

ABSTRACT

This article introduces the *transaction test*, a novel experimental method to detect whether a website customization can be deceptive. The method measures changes in choice inconsistency with respect to a plain design. In a highly powered online experiment, we assess three dark patterns and one transparency-based remedy in decisions involving data protection. Compared to a placebo condition, dark patterns increase the likelihood of choice inconsistency by an effect in the range of 6-12 percentage points, depending on experimental conditions. The transparency-based remedy fails to mitigate these effects. A second study implements the transaction test in a market experiment. Here, one of the treatments from the online experiment increases inconsistency by 42 pp. Two additional deceptive designs are tested, respectively favoring consistent and inconsistent choices. The Dark Pattern version raises inconsistency by 18 pp while the Bright Pattern reduces it by 15 pp. Our findings demonstrate how the transaction test, by providing an empirical assessment of the legal concept of unfairness, offers a flexible tool for the regulation of dark patterns and other commercial practices.

1. Introduction

In this article, we introduce the transaction test, an experimental method to assess whether a website customization increases the chance that users' decision-making is inconsistent with individual preferences. Directly inspired by the EU legislation on unfair commercial practices, the transaction test provides a quantifiable and easy to communicate measure of manipulation and can be used as a probatory instrument in cases involving consumer protection.

The transaction test fills a void in the current policy discussion around digital privacy and dark patterns, the jargon term used by computer scientists to describe customizations that try to steer users' choices (Brignull, 2010). Digital apps and tech giants constantly update their websites and platforms to make the user journey easier, increasing the conversion rate from view to purchase, and favoring engagement and retention. Most of these changes are A/B tested, thus presumed to be effective, and raise the legitimate concern that may exploit impulsivity or other behavioral biases. Privacy constitutes the domain where this concern is more apparent (Hartzog, 2018), individual data being the key asset for price discrimination and more effective advertising. While most users state a significant

* Corresponding author.
E-mail address: francesco.bogliacino@unicatt.it (F. Bogliacino).

https://doi.org/10.1016/j.jebo.2026.107469
Received 19 March 2025; Received in revised form 2 February 2026; Accepted 4 February 2026
Available online 10 February 2026
0167-2681/

concern for privacy, they fail to make choices coherent with their statement, a phenomenon known as the privacy paradox (Goldfarb and Que 2023; Acquisti et al. 2016).

Several countries and supranational bodies have been undertaking initiatives to regulate dark patterns and cookies banners,[1] often claiming lack of probatory evidence to regulate practices. Existing studies (Sernac, 2022; Grassl et al. 2021; Miller et al., 2023; Huck & Wallace, 2015; Rasch et al., 2020; Blake et al., 2021; Dertwinkel-Kalt et al., 2020) have focused on whether dark patterns change behavior (e.g. increasing sales) and advocate for regulation based on evidence of this impact. In other words, the current debate infers manipulation from effectiveness. This misses the mark: while some designs may be detrimental, others may be beneficial, by providing useful information or making choices easier. If all that takes to invoke regulation is the argument that a design works, companies may be discouraged from innovation in commercial practices. A more stringent test is required, capable of completing the assessment by testing whether the design leads the consumer to inferior choices. This is what the transaction test does.

We apply the transaction test to two separate studies. In the first, we assess three dark patterns and a protective measure in a high-powered ($N = 7430$) online experiment in six European countries (Spain, Italy, Germany, Poland, Bulgaria, and Sweden), where consumers are presented with framed goods that differ in their degree of data protection. Participants are presented with a choice between two entertainment packages. One of the options requires sharing personal data with a third party and comes with a discount. In all other dimensions, the two options are the same. The discount is intentionally set below the individual's monetary evaluation for data protection (elicited beforehand in a separate task). Participants are incentivized to choose in line with their stated preferences. Although this implementation is not conventional for laboratory experiments, it parallels the standard protocol in UX design experiments, where subjects are asked to perform *tasks* and the researcher tests whether the design facilitates or hinders their completion.

The experiment follows a between-subject design. The first level of randomization concerns the architecture of the main task, which can either be neutral (control), deceptive (dark pattern), or deceptive with remedy (dark patterns and protective measure). We use three dark patterns. *Hidden information* obfuscates the terms of the contract regulating data protection, which nevertheless can be accessed through further actions. *Toying with emotions* (TWE) delivers a promotional message leveraging emotions. *Personalization* is the label used for targeted offers built upon navigation data. Building on previous results in psychological targeting (Matz et al. 2017), we use self-reported profiling on an extroversion-introversion scale to choose a tailored image promoting the product. As a possible protective measure, we test a “cool down” intervention. The cool down is implemented as a confirmation prompt. Participants are exposed to the dark pattern and subsequently presented with a summary of their choice and are asked to either confirm or change it.

Additionally, we manipulate the time availability. We randomly assign participants to a motivated delay or an incentive compatible time pressure environment (Alós-Ferrer and Garagnani 2020; Bilancini, Boncinelli, and Celadin 2022). In designing this manipulation, we were motivated by policy. The current legislation judges the manipulative nature of a practice from the perspective of an “average” consumer, who pays sufficient attention and processes all relevant information; in our protocol, this is represented by the motivated delay treatment. The legislation also introduces special provisions for consumers vulnerable due to structural or situational factors: the incentivized time pressure condition is meant to limit subjects' attention so as to behave as a vulnerable consumer.

In our online experiment, dark patterns raise the level of choice inconsistency by 9 percentage points (pp), from 38 % to 47 %. Among the three, the most effective is hidden information (12 pp). The introduction of the protective measure does not counteract the effect of dark patterns: in aggregate for these treatments, the level of inconsistency remains at 47 % of choices and even when disaggregated by dark pattern, the effect of the measure is neither significant nor sizable.

Time pressure pushes the share of inconsistent choices to 45 %, already in the control. The effect of dark patterns continues to be positive and significant, but smaller in size: on average, subjects in the dark patterns' conditions are inconsistent 51 % of the time, and the results remain the same across the three dark patterns. The cool down reduces the likelihood of inconsistency by a couple of points but the results are not statistically significant at conventional levels.

The first study was conducted under a contract for the European Commission, and the policy requirements partly constrained the design, particularly in how we provided incentives. Our second study implements the transaction test in a market experiment with a conventional induced value protocol (Smith, 1976). Subjects are presented with a choice between two multi-featured packages that vary in terms of data protection and price. Unlike the first study, the value of data protection and other features was not directly elicited but randomly assigned to subjects. Of the two options, one does not sell data to third parties, the other does, but its price tag is lower. As in the previous study, the price difference is intentionally calibrated to be lower than the assigned monetary valuation of data protection. The packages do not differ in any other feature. Participants earn the difference between valuation and price of the selected package. We tested a variant of *hidden information,* and two variants of TWE. These two TWE treatments share the same “manipulative” interface, making some features salient, playing around emotionally charged words, and nudging the choice with a colored button. However, while one version favors the discounted package without data protection (the Dark Pattern version), the other promotes the alternative (Bright Pattern version). Hidden information continues to be very effective, increasing inconsistency to 79 % from a baseline of 36 % in the control. The two TWE treatments affect inconsistency level in different directions. While the Dark Pattern version of TWE makes the choice of the package without data protection 18 pp more likely, the Bright Pattern makes it 15 pp less likely.

This last finding sheds light on the probatory value of the transaction test. The two TWE conditions share the same manipulative design and determine comparable change in behavior. Conventional tests of effectiveness would raise concerns about both versions: they work, and they include elements of manipulation. On the other hand, the transaction test successfully distinguishes an innovation

[1] We cannot review this legal literature here, but some key references are BEUC (2022), Norwegian Consumer Council (2022), UK Competition and Market Authority (2022), SERNAC (2021; 2022), Netherlands Authority for Consumers & Markets (2020), Waldman (2020), Richards and Woodrow (2019), European Commission (2021), Calo (2014)

in design that is welfare improving (Bright Pattern version) from one that is harmful (Dark Pattern version).

Previous research evaluating whether specific online choice architectures are harmful often engage in audits (an approach inspired by Thaler, 2018 and Sunstein 2022). By going through the user's journey to complete a mission, like subscribing to a service, the audit investigates whether the design of the webpage appears to conflict with the legal principles of fairness in commercial practices (e.g. transparency) or whether it raises red flags by "exploiting" biases and heuristics (Mills et al. 2023, Mathur et al. 2019a; Bösch et al. 2016a; Di Geronimo et al. 2020). The transaction test is transparent about the underlying behavioral assumptions it relies on, emphasizing the role of individual preferences and choice inconsistency, and thus operationalizing the legal concept of manipulative practice. It is not grounded on identifying specific biases, an approach that may be questioned as *ad hoc*. In the future, audits may incorporate the transaction test as a common measure of manipulation.

The literature has also tried to assess whether dark patterns change behavior (Luguri and Strahilevitz, 2021; Zac et al. 2025; SERNAC, 2022; Grassl et al. 2021; Sin et al. 2022). Some papers focus on specific practices, like drip pricing (Blake et al. 2021; Dertwinkel-Kalt, Köster, and Sutter 2020; Miller, Sahni, and Strulov-Shlain 2023; Huck and Wallace 2015; Rasch, Thöne, and Wenzel 2020), while others on more general strategies, like obfuscation or confusion (Kalayci & Potters 2011; Kalaycı 2016; Gu and Wenzel 2020; 2015). The transaction test assesses at once whether a practice works and whether it manipulates consumers. As a result, it provides an instrument to complement the existing analyses with a clear policy relevant quantification.

Consumers' and users' choices over their online privacy settings often conflict with their stated privacy preferences. This attitude-behavior gap, commonly referred to as the privacy paradox (Goldfarb and Que 2023; Acquisti et al. 2016) has attracted much attention in the last 10 years, although it has mostly been documented using non incentive-compatible methods (for a recent attempt to develop revealed preference measures of willingness to share data see Benndorf and Normann 2018). Lee and Weber (2024), aiming to understand if this inconsistency is pristine to privacy choices rather than a simple consequence of their demanding nature, elicit revealed preferences for data protection in two dimensions (IQ and Body Mass Index) and examine incentivized behavior in tasks where personal data are exchanged for money. When trading off privacy across different domains, 37 % of subjects present some violations of GARP, a proportion not systematically higher than that in other choice domains. Yet when translating disclosure choices into monetary evaluation, inconsistency rises to 53.7 %, indicating that the difficulty stems from the complexity of the choice rather than from unstable or domain-specific preferences.

This insight suggests that our transaction test, by isolating individuals' ability to make consistent choices in an artificial but fully incentivized setting is particularly well suited to the privacy domain. We note that the baseline share of inconsistency observed in the control condition in both our experiments is in line with Lee and Weber's results for across dimension tradeoff decisions. The large effect sizes due to website customizations support the claim that choices over privacy are highly sensitive to small manipulations (Athey et al. 2017). At the same time, the results of our second experiment suggest that modest changes in choice architecture can also be leveraged to reduce consumer inconsistency.

2. The transaction test

The transaction test requires two steps. In Task I, we identify (through one of three methods) the monetary valuation of a good's feature (M_f). In Task II, subjects are prompted to choose between two versions of the good that differ by the presence of the feature and the price. The price difference between the two versions is the implicit price of the feature (p_f). Purchasing the feature when the price is higher than the monetary valuation ($M_f < p_f$) or not doing so when $M_f > p_f$ violates choice consistency, under conventional assumptions of separability and monotonicity in money. A transaction test assesses whether an interface for Task II increases the level of inconsistency with respect to a transparent design.

There are three possible implementations of the transaction test, depending on the method used for task I. If the monetary valuation can be elicited under incentive compatibility, the two tasks can be incentivized separately. If monetary valuations can be recovered only through stated preferences, we can incentivize consistency: calling $I\left(M_f, p_f\right)$ the indicator function equal to one if the participant with preference M_f facing price p_f makes an inconsistent choice, the task pays the participant $a\ I\left(M_f,p_f\right) + \left(1 - I\left(M_f,p_f\right)\right)\ A$, with $a < A$. Finally, if the aim of the application requires controlling the distribution of the preferences, the experimenter can randomly assign the valuation to subjects and compensate them with the outcome of their choices (valuation minus price), as in conventional market experiments.

Formally, consider a choice architecture d as a decision-setting for Task II. The architecture includes the information provided upfront (including graphical aspects, cues, etc...). Define $\bar{d}$ to be the transparent design, with neutral information and absent explicit cues. A transaction test assesses the following null hypothesis:

$$H_0 : E\left[I\left(M_f, p_f\right)\middle|d\right] \leq E\left[I\left(M_f, p_f\right)\middle|\bar{d}\right]$$

Notice that the transaction test is different from a test of effectiveness and can deliver different policy recommendations. To see this, consider a market for an online service, where the key product involves data protection, and a sizable share of consumers display status quo bias. A new offer becomes available, with a discount that exceeds monetary evaluation for most of the market, i.e. $F\left(p_f\right) > \frac{1}{2}$, where $F(\cdot)$ is the cumulative distribution function of the monetary valuation. Under neutral design ($\bar{d}$), status quo bias will cause a significant amount of inconsistency, as consumers will tend to stick to the initial offer, even though switching would be beneficial on average. A dark pattern-like design d that sets the new offer as default will significantly increase subscriptions to the new package, because default leverages status quo bias (Thaler and Benartzi 2004). A test of effectiveness of the new design would reject the null and its policy recommendation would be that the design be blocked. On the other hand, the outcome of the test of inconsistency would be the result of two opposing forces.[2] We will observe a sizable increase of inconsistency in the upper tail of the distribution, as a subset of $1 - F\left(p_f\right)$ will now purchase the new offer. Nevertheless, an even larger share of consumers within the lower tail will now correct their inconsistency. Given this distribution, the transaction test for the new design will fail to reject the null, recommending that the regulator should abstain from action. In anticipation of government intervention, companies will be disincentivized from developing the new design, if the agency in charge relies on a test of effectiveness, instead of the transaction test. Outside of statistical margins of error, the transaction test does not make policy recommendations that stifle innovation, while at the same time protecting consumers from real harm.

Our transaction test is coherent with the legal discipline in Europe and the US. The discipline of dark patterns in Europe falls within the domain of the Unfair Commercial Practice Directive, which regulates "any act, omission, course of conduct or representation, commercial communication including marketing, by a trader, directly connected with the promotion, sale, or supply of a product to consumers" (THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE UNION. 2005). The Directive states that practices are unfair if they breach professional diligence, are deceptive, or coercive, and if they cause or are likely to cause an average consumer to take a transactional decision that they would not have taken otherwise. The US discipline states that "an unfair trade practice is one that causes or is likely to cause a financial loss to consumers, not compensated by the benefits, and is not reasonably avoidable by a consumer" (Federal Trade Commission. 1984). Our transaction test translates the abstract concepts defined by both disciplines for unfairness into practical application. Under the EU perspective, our test serves as a joint evaluation of the deceptiveness and effectiveness of the practice. Under the US perspective, the test identifies the share of transactions with a net loss.

3. Experiment I: Experimental design

3.1. Implementation of the transaction test

We will now describe the concrete implementation of the transaction test in the experiment. Participants are prompted to choose between two entertainment packages (framed goods). Version 1 of the package includes a price (p_1), data protection, and other $n-2$ features. Version 2 costs p_2, sells data to third parties, and includes other $n-2$ features. The difference in price $p_1 - p_2$ is calibrated to be 2.5 euros lower than the monetary valuation of data protection, elicited in a separate task. Participants should always choose version 1 of the package, as it better suits their preferences.

The other $n-2$ characteristics are identical between the two packages, but there are two (randomized) versions of the description, to provide a meaningful choice and avoid revealing our research question. The format of the price can change as well: prices are shown in the "pay for twelve months" or "pay for three months" formats, randomized across options. Fig. 1 shows the interface of the main choice in the control condition. The position of each option on the page was randomized.

Participants were incentivized to choose consistently. Before seeing the main choice, they were told that their earnings would depend on making choices in agreement with the response to the questions on monetary valuation. They were provided with an example and a comprehension question with feedback.

3.2. Elicited monetary valuation

We elicit the monetary valuation of data protection on an eleven points scale, from 0 to 5 EUR or more, where each point corresponds to 50 cents. To ensure valid elicitation, we adopt the following measures. First, we explain to the participants that their responses would influence their earnings in the experiments. Second, we introduce four distractors to avoid experimenter demand; in total, we elicit monetary valuation for data protection, interruption by commercials, access to movie pre-release, unsubscribe in one click, and access to a companion service (music platform). Third, we follow Diaz et al. (2021) and introduce an incentivized recall: participants earn one additional point if they can repeat the answer to one randomly chosen question.

3.3. Randomization

Fig. 2 shows the randomization of participants into different conditions and the number of observations per treatment.

Participants are first randomized into Time Pressure (Vulnerable Consumer) or Motivated Delay (Average Consumer). In the Vulnerable Consumer treatment, participants are endowed with 30 seconds (equivalent to 1.5 points) and incur a monetary deduction for each second spent while making their choice. In this way, the time pressure is incentive compatible, but we do not exclude any participants from the analysis (Alós-Ferrer and Garagnani 2020; Bilancini, Boncinelli, and Celadin 2022). In the Average Consumer treatment, participants earn an additional point if they provide a text of at least 32 characters explaining their choice. This randomization allows us to differentiate the impact of dark patterns and the remedy on vulnerable consumers, i.e. someone whose

[2] Practically speaking, for the example at hand, the agency asked to rule on the case could either run a transaction test for different distributions of monetary valuation or with an estimation of the true distribution.

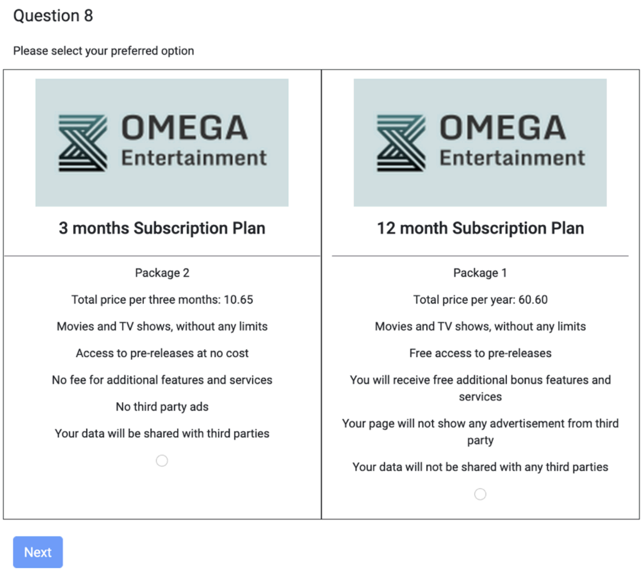

Fig. 1. The main task in the online experiment.

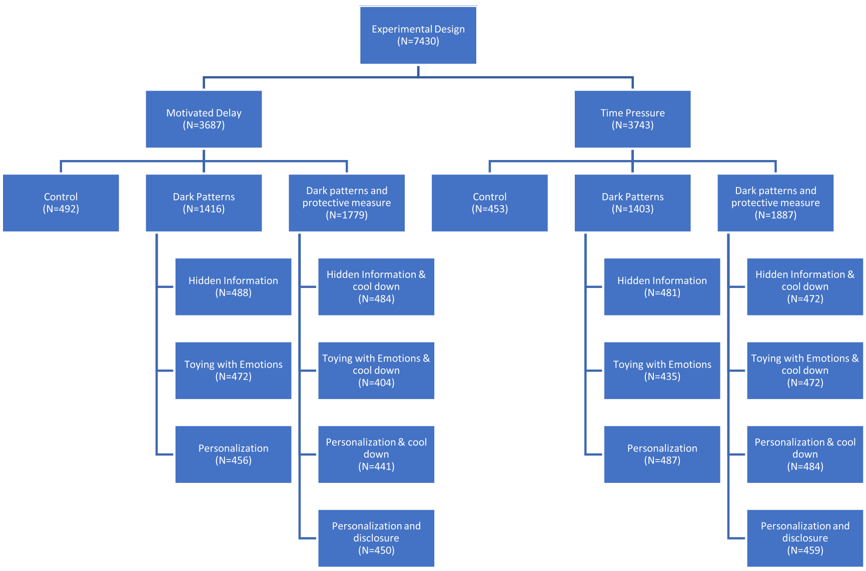

Fig. 2. Experimental Design.

capacity to process information has been partially impaired, and on average consumers, i.e. someone that pays enough attention and processes information.

The second randomization concerns the web interface for the main task. Participants face the main choice under one of three conditions: control, dark pattern, or dark pattern and protective measure. The dark pattern condition is further randomized into three sub-treatments: hidden information, TWE, and TWE with personalization (personalization hereafter). The dark pattern and protective measure condition is further subdivided into four conditions: hidden information and cool down, TWE and cool down, personalization and cool down, and personalization and disclosure.

In the cool down, the participant is shown the selected option and is asked to confirm it. The cool down page always comes after exposure to a dark pattern. Only for personalization, we include an additional sub-treatment, labelled Personalization and disclosure, where the cool down page displays an additional sentence to disclose the use of personal data for the packaging of the offer ("Your data have been collected and used to prepare this personalized offer to you").

3.4. Experimental treatments: Dark patterns and protective measures

This subsection describes the different experimental treatments for Factor II: dark patterns and protective measures. The control is shown in Fig. 1.

The Dark Patterns always steer the consumer towards the package without data protection. The Hidden Information interface has three dots instead of the data protection information, which becomes visible only by clicking on them. This treatment is displayed in Fig. 3-left panel. The TWE condition introduces a blue label stating "Don't waste time looking for what to watch. Choose this offer and we will give you personalized suggestions to new content you'll love. We have prepared this personalized offer just for you". The personalization, on top of the blue label with the TWE sentence, has a picture. The picture is personalized by gender and extroversion questions (Matz et al. 2017). If the participant is classified as an extrovert, the picture has a group of friends having fun watching TV. If classified as an introvert, the picture has a male/female (depending on gender identification) relaxing in front of the TV. The pictures were purchased from Getty®. The treatment is represented in Fig. 3-right panel.

In the conditions with cool down, after being exposed to a dark pattern and making their choice, participants see a box stating "This is a summary of the offer you selected", showing the selected package and the options to confirm or change selection. The disclosure treatment comes only after personalization: the participant is exposed to the dark pattern, makes the choice, sees the cool down, but with an additional sentence "Your data has been collected and used to prepare this personalized offer for you".

3.5. Experimental sequence

Participants go through the following experimental sequence. After stating their preferences and performing the incentivized recall, participants reveal their gender and their Big Five (Openness, Conscientiousness, Agreeableness, Extraversion and Neuroticism) personality profile (Norman 1963) in a standardized set of questions and then move to the main task. The main task includes the explanation of the incentive systems for the main task, with a comprehension question with feedback, and the main choice. A post experimental questionnaire follows, including socio-demographics, three questions on preferences towards risk, patience, and trust (Falk et al. 2018), and a short questionnaire measuring the sensitivity to persuasion (Kaptein et al. 2012). The questionnaire asks for the level of agreement with six statements, using a Likert scale. The six principles of persuasions (Cialdini 2009) are commitment, reciprocity, scarcity, likeability, authority, and social pressure.

3.6. Procedure

We sent an invitation to an online panel in Europe. We define quotas by country, gender, and age group (18-24, 25-54, 66-65). The target sample was computed at 7200 observations. Countries included are Spain, Italy, Germany, Poland, Bulgaria, and Sweden. A native speaker translates the protocol in each language, further revised by separate experts.

In the power analysis, we estimate an average inconsistency in the baseline of 20 % and an overall 40 % standard deviation. Considering significance at 10 % (one side) and 80 % power, we wanted to identify a minimum sizeable effect of around 15 % of a standard deviation. This gives at least 400 observations per cell.

Each participant received a fixed fee for completing the questionnaire plus the incentives from the main task and from the monetary evaluation task. Average variable incentives are in line with the payments on other online platforms. The completion time was around ten minutes.

All the incentives are computed in points that are exchangeable in the local currency, to ensure replicability. One point is equal to 0.50 euros. Participants receive one point if they correctly recall the stated monetary evaluation in one randomly chosen question and 10 points if they choose consistently (1 point if they fail to). In the motivated delay, they receive 1 point if they write at least 32 characters to explain their choices; in the incentivized time pressure, they are endowed with 30 seconds (corresponding to 1.5 points) to make their choice. If they spend more than 30 seconds, they are not prevented from making their choice but do not receive additional incentives.

The full experimental protocol, with images of all the treatments is available in the Supplementary Online Appendix. The companion report of the study includes additional information and analysis (European Commission et al. 2022). Ethical Committee at UOC granted the IRB approval on October 26th, 2021. The hypotheses and the analysis plan were pre-registered on Aspredicted #84620.

3.7. Analysis plan

Our outcome variable Y_i is a dummy equal to one if the participant i chose inconsistently. The outcome per participant can be written in the following form:

$$Y_i = \alpha + \sum_j \sum_k d_i^{j,k} \beta^{j,k} + \varepsilon_i$$

Where i is the participant, $j=1,.., 8$ is the condition with respect to factor one (the choice architecture, see the Experimental treatments subsection above), $k=1,2$ is the condition with respect to factor two (motivated delay v. time pressure), $d_i^{j,k}$ is the treatment dummy (equal to one if the participant i is in the condition j,k and zero otherwise) and ε_i is any unobservable variable affecting the outcome. Control and Motivated Delay is the omitted category. $\beta^{j,k}$ tells us how many percentage points more likely to be inconsistent a

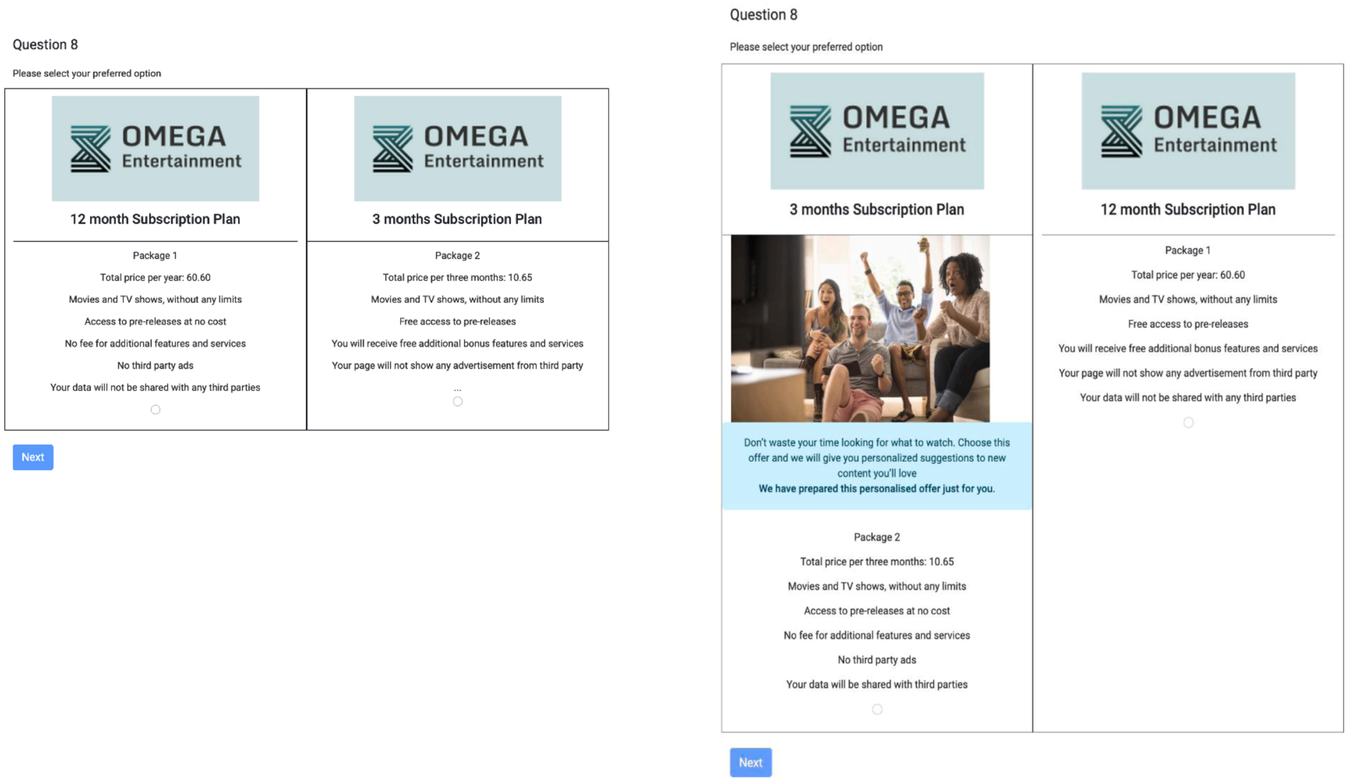

Fig. 3. The hidden information and personalization treatments.

participant is, when in experimental condition j,k, with respect to the omitted category. All tests are one sided, as formally explained in Section 2.

The main hypotheses to be tested are four. (H1) Dark patterns induce more inconsistency in average consumers; (H2) dark patterns induce more inconsistency in vulnerable consumers; (H3) protective measures reduce inconsistency with respect to dark pattern in average consumers; (H4) protective measures reduce inconsistency with respect to dark pattern in vulnerable consumers.

The equation above is estimated by Ordinary Least Squares. In the estimation, we control for randomization of the options on the page, for the annual/quarterly price, and for description of alternative features.

4. Experiment I: Results

7430 participants took part in the online experiment. The descriptive statistics are reported in Table 1.

4.1. Aggregate evidence

Fig. 4-left panel reports the average level of inconsistency for participants in the Motivated Delay condition, which corresponds to the condition of the average consumer. We aggregate the Dark Patterns and Dark Patterns & Protective Measure in two groups. In the control, the level of inconsistency is 37.80 %. Dark Patterns increase inconsistency by 9 percentage points (pp) to 47.25 %. The presence of the protective measure does not reduce the likelihood of inconsistent choices, which remains at the same level (46.60 %).

In Fig. 4-right panel, we report the average level of inconsistency for those participants in the time pressure condition, subjected to temporary vulnerability. We group the results of the Dark Patterns and Dark Patterns & Protective Measure. The level of inconsistency is significantly higher under incentive-compatible time pressure than motivated delay: vulnerable consumers have a greater likelihood of being inconsistent than average consumers (45.47 % versus 37.80 %). The effect of Dark Patterns is still significant, but it is about two-thirds of what we detected under motivated delay (6 pp), and the level of inconsistency reaches 50.89 %. The introduction of the protective measure induces a 2 pp reduction in the likelihood of inconsistency (48.60 %).

4.2. Granular evidence

The results of the main OLS regression described in the Analysis Plan are reported in Table 2 column (1) below. Column (2) adjusts for socio-demographics, gender, age, civil status, employment status, income brackets, and education. Column (3) includes the response to the privacy valuation. Results are consistent across the three specifications.

To better understand the effect of the individual treatments, we display the coefficients in Fig. 5. In the top panel, we plot the effects of the dark patterns and dark patterns with protective measure with respect to the control for average consumers (motivated delay). The graph includes the confidence interval at 90 % to visualize the level of significance in one-sided tests (since differences are all in the expected direction). In the bottom panel, we make the same plot for vulnerable consumers. In this case, we adjusted the reference group: the plotted coefficients report the difference in inconsistency with respect to the level under the control design and time pressure.

Under motivated delay, Hidden information increases this likelihood by 12.25 pp ($t = 3.90$, $p < 0.001$, all tests are one-sided), TWE increases the likelihood by 6.02 pp ($t = 1.90$, $p = 0.028$), and personalisation increases the likelihood by 9.84 pp ($t = 3.08$, $p = 0.001$). Personalisation does not significantly add to the effect of TWE alone (3.81 pp, $t = 1.17$, $p = 0.123$).

Under time pressure, Hidden information increases this likelihood by 5.80 pp ($t = 1.80$, $p = 0.036$), TWE does not significantly increase the likelihood (4.16 pp, $t = 1.26$, $p = 0.103$), and personalisation increases the likelihood by 5.50 pp ($t = 1.71$, $p = 0.043$). We can test whether personalization outperforms TWE: this is not the case ($t = 0.41$, $p = 0.340$).

As can be seen, time pressure decreases the effect of dark patterns across the board. A possible interpretation of this evidence is that an instance of crowding out is taking place. This study and the other laboratory experiments in this paper (as we will discuss in the final Section) show that around one third of subjects make inconsistent choices even without deceptive design. Due to the binary scale of the measure, these subjects will not be affected by dark patterns, because they already choose the worst outcome. Among the remaining two-thirds, there is a share of participants who get tricked by dark patterns and a share that does not, probably because of a different level of attention. When we introduce vulnerability, the baseline level of inconsistency increases, and we suspect that this change takes place entirely within the share which is susceptible to dark patterns. As a result, a form of crowding out of the effect is taking place and the resulting effectiveness of dark pattern is reduced.

As we already know from the aggregate analysis, the protective measure does not significantly reduce inconsistency. We can visualize the effect separately under each dark pattern. We start from motivated delay. The cool down decreases the likelihood of inconsistent preferences under hidden information by 0.29 pp ($t = 0.09$, $p = 0.463$), under TWE by 2.53 pp ($t = 0.76$, $p = 0.224$), and under personalisation by 1.05 pp ($t = 0.01$, $p = 0.375$). Cool down with disclosure decreases the likelihood of inconsistent preferences under personalisation by 0.11 pp ($t = 0.03$, $p = 0.486$). Overall, we cannot reject the null hypothesis that protective measure has no impact.

The effect becomes larger under vulnerability but fails to reach significance. The cool down decreases the likelihood of inconsistent choices under hidden information by 1.67 pp ($t = 0.52$, $p = 0.299$), under TWE by 2.24 pp ($t = 0.69$, $p = 0.246$), and under personalisation by 0.68 pp ($t = 0.21$, $p = 0.415$). The cool down with disclosure decreases the likelihood of inconsistent preferences with respect to personalisation by 3.52 pp ($t = 1.09$, $p = 0.137$). Therefore, this study did not detect a statistically significant effect of introducing a cool down with or without a disclosure message.

Table 1
Descriptive statistics.

Total observations	7430
Gender	
Male	48.63 %
Female	51.02 %
Other	0.35 %
Age	43 (sd 15)
Education	
Primary education	6.72 %
Secondary education	40.05 %
At least some tertiary education	11.33 %
Completed tertiary education	29.30 %
Marital status	
Single	37.04 %
Married/civil union	53.38 %
Divorced/widowed	9.58 %
Household yearly income	
9.999 Euro or below	18.28 %
10.000 Euro – 29.999 Euro	41.71 %
30.000 Euro – 49.999 Euro	23.18 %
50.000 Euro – 149.999 Euro	15.90 %
150.000 Euro or above	0.94 %
Labour market status	
Employed	64.41 %
In search of job	9.34 %
Students/retired/housekeeper	21.61 %
Other labour market status	4.63 %
Country	
Bulgaria	16.51 %
Germany	17.07 %
Italy	16.57 %
Poland	16.68 %
Spain	16.62 %
Sweden	16.55 %

Note: share of participants.

In Table 3, we report the results of the Romano and Wolf (2005) procedure to control for multiple hypothesis testing. With respect to original tests, the only non robust effect is TWE, which is not significantly affecting vulnerable consumers.

4.3. Robustness check and validity of the design

In this subsection, we present the results of the statistical analysis conducted on various components of the study. These include balancing of covariates, assessment of the time manipulation, plausibility of the preferences elicitation, assessment of the correct recall of the stated monetary valuations, and plausibility of the measurement of personality traits.

We perform a multinomial logit regression to assess whether socio-demographics (gender, age, marital status, employment status, education, income level) and attitudes towards persuasion predict assignment to treatment cells. Out of the 180 tests of hypothesis, only six are significantly different from zero and no covariate appears significant more than once, which is compatible with random assignment.

Time manipulation worked. The average time to decide under time pressure was 24.05 seconds against 144.49 seconds in the motivated delay condition ($t = 29.54$, $p < 0.001$).

SOM, Section V, Fig. 1, Panel A shows the histogram of the monetary valuation for data protection. The modal response is 5 euros or more (24.01 %), followed by 2 euros (14.79 %). Since 5 euros or more is the largest option, 24.01 % of the respondents answer the rightmost category. This is larger than the closest options; consequently, a larger support would have changed the shape of the distribution. For these subjects the discount offered was small compared to their willingness to pay, potentially softening the effect of dark patterns.

At the end of the first instrumental task, respondents were incentivized to correctly recall the stated monetary valuation. 100 % of participants recall their response within a one-euro difference from the original stated value (while the difference between monetary valuation and price in the choice task was 2.5 euros by construction). The difference between the first stated valuation and the second stated valuation for the same question represents a measure of “recall error”. This measure is not statistically different across experimental conditions (Kruskal-Wallis, $\chi^2 = 8.013$, $p = 0.331$). Comprehension, as measured by a question in the instructions (the text of the question and the feedback are reported in the SOM), does not represent an issue either, as it is not significantly different across conditions (Pearson’s $\chi^2 = 2.806$, $p = 0.902$). Table 1 in the SOM replicates Table 2 above adding both comprehension and recall as further controls in all three specifications and the results are identical.

Finally, we analyze the answers to the personality traits questions (Woods and Hampson 2005). Since personalization requires

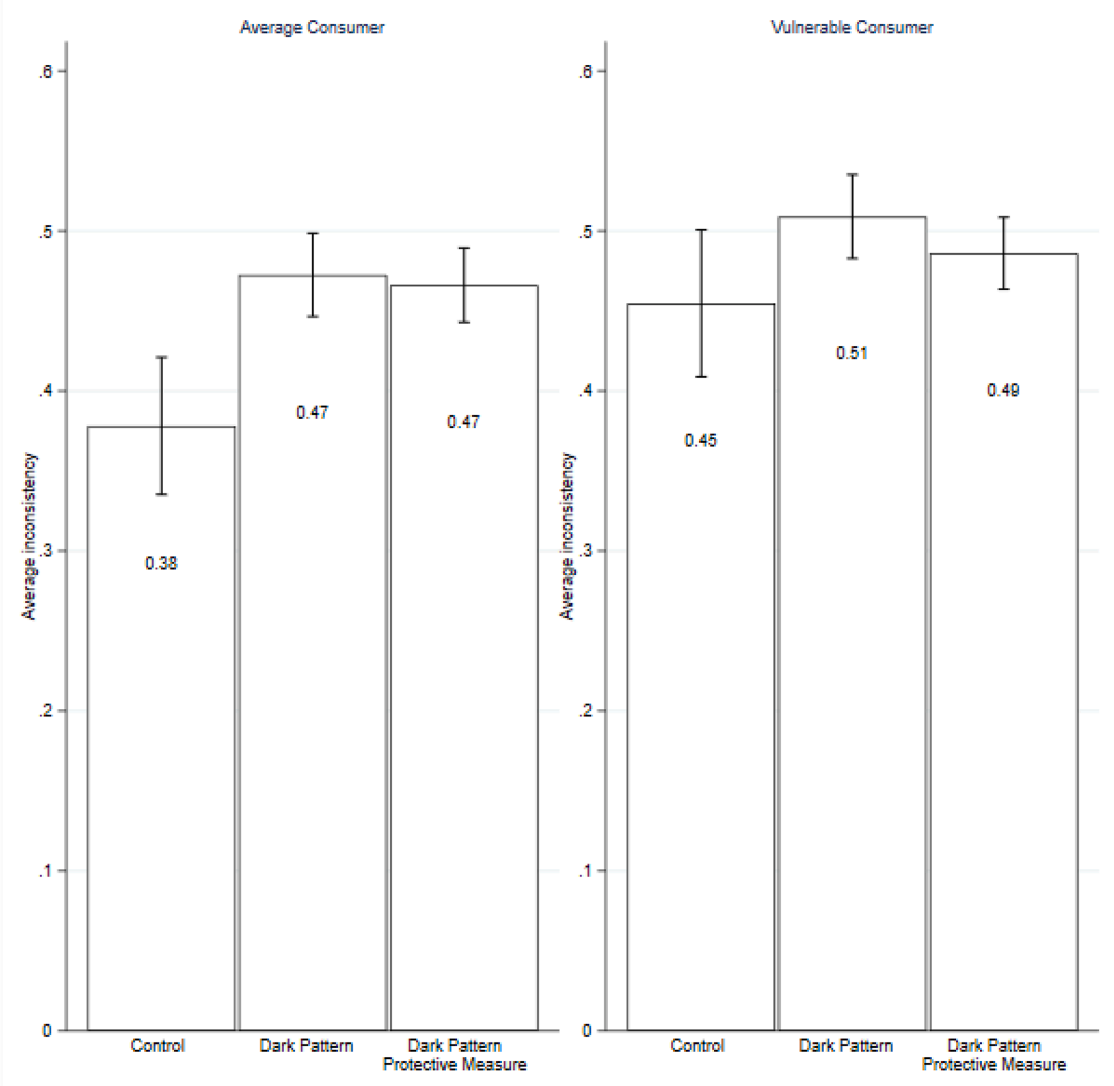

Fig. 4. The average level of inconsistency broken down by experimental conditions.
Note: Average consumers are the participants in the motivated delay condition; vulnerable consumers only include participants in the incentivized time pressure condition. The graph shows the means by condition and the confidence interval at 95 %.

separately identifying those that score low and high on this dimension, we can check whether the distribution has most of participants in the middle or displays fat tails. Had the distribution been approximately normal, (1) most of the response would be around the mean and (2) measurement error would shift the type of a significant portion of people, classifying a lot of slightly introvert as extrovert and vice versa. Under this scenario, personalization can fail to change the behavior, because of mistargeting (sending the wrong message) and not necessarily because of ineffectiveness (the message does not work).

SOM, Section V, Fig. 1, Panel B reports the empirical distribution for the response to the extraversion item, together with a normal distribution. It is shown that the distribution has much more mass on the tails, rejecting the hypothesis of normality. This weights in favor of the instrument. The limited value added of personalization may be due to the choice of images or to the absence of treatment effect but is not due to measurement error in personality traits.

5. Experiment II: Experimental design

5.1. Objectives

In the previous Sections, we documented how to apply the transaction test to website customizations involving choices over privacy settings. The preferences used to assess consistency were elicited in a separate task by means of a self-reported question. This may raise

Table 2
The impact of dark patterns and protective measure on the main outcome.

VARIABLES	(1) Inconsistent Choice	(2) Inconsistent Choice	(3) Inconsistent Choice
Hidden Information (A)	0.12***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)
TWE (A)	0.06*	0.06*	0.06**
	(0.03)	(0.03)	(0.03)
Personalization (A)	0.10***	0.10***	0.10***
	(0.03)	(0.03)	(0.03)
Hidden + Cool Down (A)	0.12***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)
TWE + Cool Down (A)	0.03	0.04	0.04
	(0.03)	(0.03)	(0.03)
Personalization + Cool Down (A)	0.09***	0.09***	0.09***
	(0.03)	(0.03)	(0.03)
Personalization + Disclosure (A)	0.10***	0.10***	0.10***
	(0.03)	(0.03)	(0.03)
Control (V)	0.08**	0.08**	0.08**
	(0.03)	(0.03)	(0.03)
Hidden Information (V)	0.13***	0.13***	0.13***
	(0.03)	(0.03)	(0.03)
TWE (V)	0.12***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)
Personalization (V)	0.13***	0.13***	0.13***
	(0.03)	(0.03)	(0.03)
Hidden + Cool Down (V)	0.12***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)
TWE + Cool Down (V)	0.10***	0.09***	0.09***
	(0.03)	(0.03)	(0.03)
Personalization + Cool Down (V)	0.12***	0.12***	0.12***
	(0.03)	(0.03)	(0.03)
Personalization + Disclosure (V)	0.10***	0.09***	0.09***
	(0.03)	(0.03)	(0.03)
v1	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)
v2	-0.12***	-0.12***	-0.12***
	(0.01)	(0.01)	(0.01)
v3	-0.02**	-0.02**	-0.02**
	(0.01)	(0.01)	(0.01)
female		-0.02*	-0.02*
		(0.01)	(0.01)
age		0.00	0.00
		(0.00)	(0.00)
married		0.03**	0.03**
		(0.01)	(0.01)
education		0.03**	0.03**
		(0.01)	(0.01)
income		0.01	0.01
		(0.01)	(0.01)
employed		0.01	0.01
		(0.01)	(0.01)
Privacy valuation			0.00
			(0.00)
Constant	0.45***	0.41***	0.40***
	(0.02)	(0.03)	(0.03)
Observations	7,430	7,430	7,429
R-squared	0.02	0.02	0.02

Note: OLS regression. Regressors are dummies for experimental conditions. (A) "motivated delay", (V) "time pressure". v1: order of presentation of the packages, v2: price is annual on the left and quarterly on the right, v3: order of the formats of additional features in the package. Robust standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

two concerns that we should address: the role of measurement error and the possibility that the treatment changes preferences. While we address these critiques to the stated preferences approach in the discussion section, in the current section we prove the robustness of the transaction test by removing the channel through which these two factors operate (*design by elimination;* Niederle, 2025). In fact, here we present an experiment where preferences are randomly *assigned* to participants. Except for recall, this eliminates the possibility

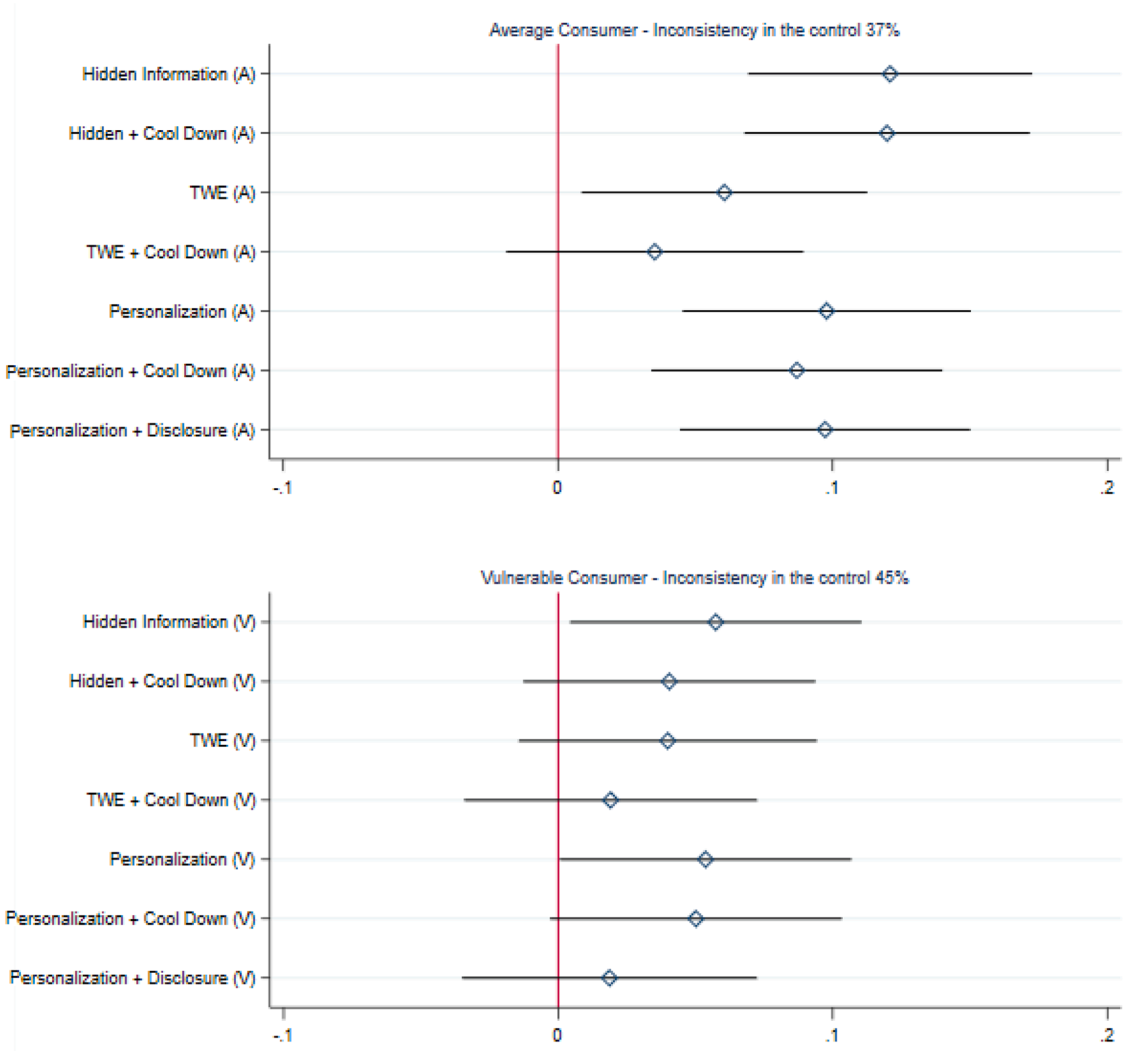

Fig. 5. The impact of Dark Patterns and Protective Measures on the Average and Vulnerable Consumers.
Note: Average consumers are the participants in the motivated delay condition; vulnerable consumers only include participants in the incentivized time pressure condition. In each graph, the reference group is the control design: in the top panel, the effects are calculated with respect to the control in the motivated delay; in the bottom panel, the effects are calculated with respect to the control in the time pressure. We plot the confidence interval at 90 % to help visualize the one-sided test (because of the definition of transaction test). Authors' elaboration from the regression displayed in Table 2.

of measurement error.[3] Moreover, there can be no updating of preferences.

In the implementation of the previous experiment, the cost of inconsistency, though sizable, is constant across participants. Nevertheless, one may wish to manipulate the cost of inconsistency ($|M_f - p_f|$) to learn about its impact on susceptibility to dark patterns and to further analyze the effect of choice architecture on payoffs.

In this second part of the paper, we present the design and discuss the results of a market experiment where the preferences are assigned, and the cost of violation varies over a range. Additionally, we introduce two treatments which are based on the same type of manipulation (a version of TWE) but are expected to work in opposite directions. Analyzing these two experimental conditions, we can see whether the transaction test is able to discriminate between a welfare improving innovation and one that hurts consumers.

5.2. Implementation of the transaction test

Participants play the role of buyers. They are asked to choose between two entertainment packages: version 1 of the package

[3] Notice that, as we discussed in the Robustness checks in Section 4, recall did not constitute a sizable problem in Experiment I.

Table 3
Retest of the hypotheses, Romano and Wolf p-values.

	Romano Wolf p-value
H1: Dark Pattern v. Control, Average Consumer	
Hidden Information	0.001
TWE	0.026
Personalization	0.003
H2: Dark Pattern v. Control, Vulnerable Consumer	
Hidden Information	0.026
TWE	0.120
Personalization	0.048
H3: Protective measure v. Dark Pattern, Average Consumer	
Hidden + Cool Down	0.516
TWE + Cool Down	0.775
Personalization + Cool Down	0.631
Personalization + Disclosure	0.491
H4: Protective measure v. Dark Pattern, Vulnerable Consumer	
Hidden + Cool Down	0.713
TWE + Cool Down	0.750
Personalization + Cool Down	0.592
Personalization + Disclosure	0.855

Note: OLS regression. Romano Wolf procedure, bootstrapping with 999 repetitions. One sided.

includes a price (p_1), data protection, and other $n-2$ features, version 2 costs p_2, sells data to third parties, and includes other $n-2$ features. The difference in price, $p_1 - p_2$, is random and is always lower than the monetary valuation of data protection.

There are six features: content streaming, data protection, absence of advertisement, access to new releases, unsubscribe in one click, and additional access to a music platform. Participants receive (randomly assigned) valuations for all six features, but the packages they choose from may include a limited subset of these. Here the assignment of valuations is replacing the elicitation of stated preferences in the previous experiment.

Subjects' variable payoff is the difference between (the sum of) their valuations and the price of the package. As in the previous experiment, participants should always choose version 1 of the package to maximize their payoff.

Fig. 6 shows the interface of the main choice in the control condition. The position of each option on the page was randomized. The descriptions of the other features come, as before, in two versions and are randomized. Unlike what happens in the previous experiment, here the price is always expressed in points.

5.3. Treatments

Participants are assigned to one of four conditions: Hidden Information, Toying With Emotions in the Dark Pattern version (TWE Dark Pattern), TWE Bright Pattern, and Control.

Each package appears as an orange container with a logo, a price, and a set of characteristics. The control condition is plain and informative, with the two packages side by side and all the description of the features neutrally presented, as depicted in Fig. 6. Hidden information replaces the description of the data privacy feature on both sides with three dots. Hovering over the dots changes the cursor, and clicking reveals the description.[4] The TWE treatment comes in two formats. In the Dark Pattern version, displayed in Fig. 7, we manipulate the respondent to choose the package without data protection. Following conventional techniques used in online marketplaces, we format in the same way a tag with a catchy sentence, the button that selects the option, and the description of the feature that we want to make salient (in this case, the price). Additionally, where the salient package displays color, the alternative one displays white. The tag includes the sentence "Selected for you! Enjoy the benefits of this offer."

The Bright Patterns version, displayed in Fig. 8, "nudges" respondents towards the package with data protection. The logic of the design is the same: the interface makes salient the tag, the button, and a feature. However, this time, data protection is made salient and the opposite choice is nudged.

5.4. Experimental procedure

Participants go through eight pages: informed consent, captcha, general instructions, specific instructions (with comprehension question), assignment of valuations, intermediary questionnaire, main choice, and feedback. As in the previous experiment, the intermediate questionnaire asks about personality profile, age, and gender. The experiment omits elicitation of stated preferences, replacing it with the value assignment.

The experiment was programmed using Lab.js and Open Lab (Henninger et al. 2022a; Shevchenko 2022) and run on the Prolific

[4] With this task, the subject is primed to look at all characteristics, removing the description of data protection only from one package will nudge subject to choose the alternative. The condition is shown in Figure 10 of the SOM.

Fig. 6. The main task in the market experiment.

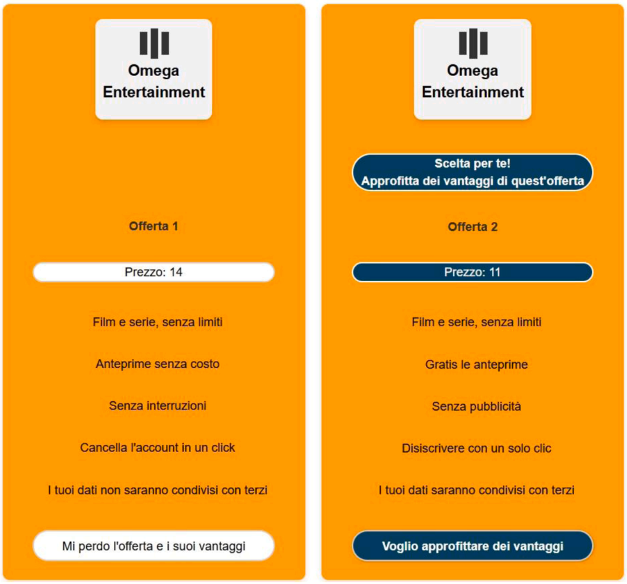

Fig. 7. The Toying-with-emotion treatment (Dark Pattern).

platform in January and July-August 2025. The IRB at UOC provided an extension of the original approval and all participants provided informed consent. The instructions are available in the SOM, Section IV. We run the experiment on an Italian sample since all authors are native speakers, and Italy was included among the countries of the previous study.

5.5. Hypothesis and data analysis

This experiment constitutes a test of H1 and H3, as formulated in Section 3. The null hypothesis is that Dark patterns increase the level of inconsistency with respect to a plain design. The Bright Pattern was designed to reduce inconsistency with respect to the plain design, so we reverse the direction of the inequality for the null hypothesis related to this treatment. All tests are one sided (as formalized in Section 2), in the proper direction.

We report both nonparametric tests and linear regressions.

6. Experiment II: Results

We collected 409 observations over seven sessions, 55.01 % of subjects self-describe as male, while the modal category for age is 25-34.

Out of 99 participants assigned to the control, 36 (36.36 %) choose the package without data protection. The average outcome is essentially the same as the one in the previous study, despite the change in the incentive system. The extent of inconsistency in the

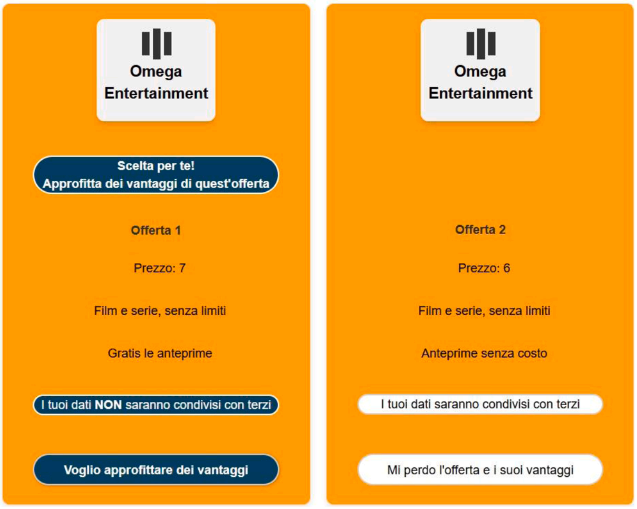

Fig. 8. The Towing-with-emotion treatment (Bright Pattern).

control is sizable, qualitatively similar to the "privacy paradox", the tendency of platform users to engage in online settings that violate their stated preferences for data protection. As we will discuss in the next Section, it is also in line with the level of inconsistency elicited in non-privacy domains. Even assuming measurement error, the first observation from this experiment is that there is significant space to improve consumer decisions.

In the hidden information condition, 79 out of 100 (79.00 %) choose the package without data protection. This constitutes a remarkable 43 pp effect (Pearson χ^2=37.07, $p < 0.001$, all tests are one sided). As a second observation, we note that the effectiveness of Hidden Information has been replicated in the market experiment, despite the change in the incentive system. The size of the effect is significantly higher. This experiment confirms that Hidden information represents the most effective dark pattern, which makes us suspect that attention plays a critical role.

In the TWE (Dark Pattern) treatment, 55 out of 103 choose the package without data protection (Pearson χ^2=5.91, $p = 0.007$), while in the TWE (Bright Pattern), only 22 out of 107, showing a reduction of almost 16 pp (Pearson χ^2=6.348, $p = 0.006$). Our third observation digs into the nature of the transaction test. An experiment measuring behavioral change of website customizations would detect significant impact by our TWE treatments. So does the transaction test. But, since the transaction test is explicit on the behavioral assumptions over decision-making, it can differentiate its recommendations over the two versions of TWE, even though their appearance and underlying manipulative technique is the same.

6.1. *Robustness check and additional analyses*

Comprehension of the task does not affect our results. Looking at the attempts to reply correctly, which we recorded, we find that 89.73 % of participants need maximum 1 feedback. The distribution of attempts across conditions is not statistically different (Kruskal-Wallis, $\chi^2 = 0.686$, $p = 0.876$). The likelihood of answering the question correctly at the first attempt is also not different across conditions (Person's $\chi^2 = 1.234$, $p = 0.745$).

A linear probability regression controlling for order on the page, the randomized description of the additional features, and the number of attempts confirms the result, as reported in Table 4, column (1). Fig. 9 plots the results. Following the same robustness check as in Table 4 we also control for socio-demographics (column 2) and for the (randomly assigned) privacy valuation[5] (column 3), without finding any significant change.

Since in this experiment there is random variation in the discount, inconsistency has a variable cost for consumers. We can measure the impact on consumer welfare by using this monetary measure. In the control, participants earn on average 4.1 points. Hidden information reduces the payoff to 2.9 points (Mann Whitney Wilcoxon rank sum test, z=4.570, $p < 0.0001$, all tests are one-sided). In terms of share of standard deviation in the control, the effect of Hidden Information is 64 % of a standard deviation. TWE (Dark Pattern) reduces the payoff to 3.5 points (z=2.303, $p = 0.010$), 32 % of a standard deviation. Subjects in TWE (Bright Pattern) obtain a slightly higher but not statistically different payoff, 3 % of a standard deviation (z=-0.458, $p = 0.323$). In Table 4, column (4)-(6) we

[5] Valuations are balanced across experimental conditions, as expected given randomization. We perform Kruskal-Wallis test to assess equality of distribution and the hypothesis is not rejected for content streaming ($\chi^2 = 2.467$, $p = 0.481$), data protection ($\chi^2 = 2.307$, $p = 0.564$), absence of advertisement ($\chi^2 = 1.634$, $p = 0.651$), access to new releases ($\chi^2 = 3.640$, $p = 0.303$), unsubscribe in one click ($\chi^2 = 6.860$, $p = 0.076$), and additional access to a music platform ($\chi^2 = 2.284$, $p = 0.515$). In the SOM, Section VII, we replicate the analyses of Table 4, including value-Unsuscribe among the controls and the results are identical.

Table 4
The impact of dark patterns and protective measure on the main outcomes.

VARIABLES	(1) Inconsistency	(2) Inconsistency	(3) Inconsistency	(4) Payoff	(5) Payoff	(6) Payoff
Hidden Information	0.43***	0.43***	0.43***	-1.12***	-1.12***	-1.07***
	(0.06)	(0.06)	(0.06)	(0.22)	(0.22)	(0.22)
TWE (Dark Pattern)	0.18***	0.18***	0.18***	-0.57**	-0.57**	-0.57**
	(0.07)	(0.07)	(0.07)	(0.25)	(0.24)	(0.24)
TWE (Bright Pattern)	-0.15**	-0.14**	-0.14**	0.04	0.05	0.06
	(0.06)	(0.06)	(0.06)	(0.24)	(0.24)	(0.23)
left	-0.04	-0.04	-0.04	0.05	0.05	0.09
	(0.05)	(0.05)	(0.05)	(0.16)	(0.16)	(0.16)
standardDescription	-0.04	-0.04	-0.04	0.14	0.14	0.17
	(0.05)	(0.05)	(0.05)	(0.16)	(0.16)	(0.16)
Attempts CQ	0.01	0.01	0.01	-0.03	-0.03	-0.03
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
female		-0.01	-0.01		0.03	0.02
		(0.05)	(0.05)		(0.16)	(0.16)
age		0.00	0.00		0.04	0.03
		(0.02)	(0.02)		(0.07)	(0.07)
valuePrivacy			0.00			0.33***
			(0.03)			(0.10)
Constant	0.38***	0.37***	0.36**	4.07***	3.97***	2.62***
	(0.06)	(0.08)	(0.14)	(0.21)	(0.28)	(0.51)
Observations	409	409	409	409	409	409
R-squared	0.20	0.20	0.20	0.08	0.08	0.11

Note: OLS regression. Left: order of presentation of the packages. Standard Description: formats of additional features. Attempts CQ: number of attempts at the comprehension question. Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

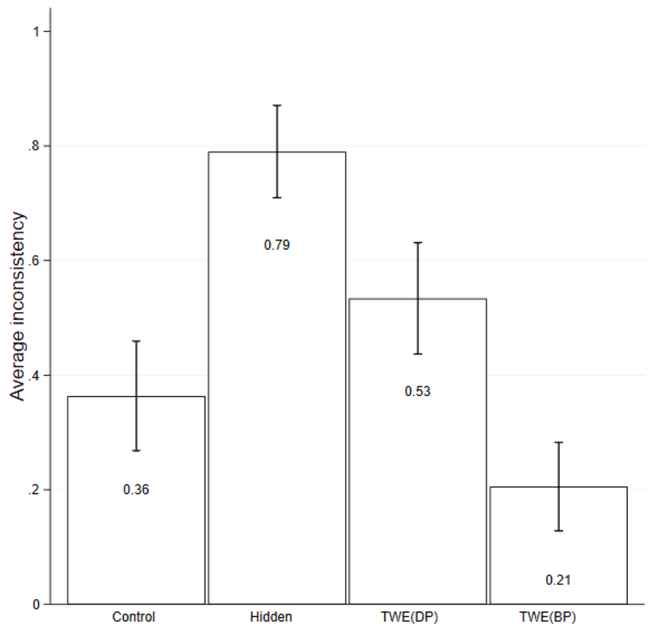

Fig. 9. The average level of inconsistency broken down by experimental conditions.

report the linear regressions using the payoff as outcome variable, with the same set of controls as for the inconsistency outcome.

We briefly report the p-value calculated using Romano and Wolf regressions with 999 repetitions for both outcomes, with one-sided tests. Using inconsistency as main outcome, we found $p = 0.001$ for hidden information, $p = 0.010$ for TWE (Dark Pattern), and $p = 0.023$ for TWE (Bright Pattern). For the payoff outcome, we found $p = 0.001$ for hidden information, $p = 0.033$ for TWE (Dark Pattern), and $p = 0.751$ for TWE (Bright Pattern) (Fig. 9).

7. Discussion and concluding remarks

7.1. An overview of the tested hypotheses

We formulated four hypotheses: H1 and H3 apply to average consumers, H2 and H4 to vulnerable consumers. H1 stated that Hidden Information and TWE (with and without personalization) would increase inconsistency in average consumers. Evidence from the two experiments largely confirms that this is the case. H3 stated that a cool down protective measure would reduce inconsistency when the design includes Hidden Information or TWE. Data from our first study showed an effect which is small in size and not statistically significant. However, a TWE designed to increase the choice of data protection succeeded in reducing inconsistency.

H2 stated that Hidden Information and TWE (with and without personalization) would increase inconsistency in vulnerable consumers. H4 stated that a cool down protective measure would reduce inconsistency when the design includes Hidden Information or TWE. These two hypotheses were tested with experiment I. Data support H2. On vulnerable consumers, the effect of the protective measure is relatively larger than in average consumers but still fails to reach statistical significance.

We summarize the evidence in Table 5 below.

7.2. The transaction test as a policy instrument

The transaction test is transparent, its results are easy to communicate, and the method is flexible. In this way, regulatory agencies have access to an instrument that can be applied as probatory evidence in cases involving consumer protection against alleged manipulation by companies. The test protects the incentives to innovate in *front end design*, without loosening the guarantees for consumers.

The measure of inconsistency is dichotomic. Thus, it is simple to understand but treats equally any violation. In our first study, this is not a concern, because the violation is large and is directly reflected into the incentive system (pay for consistency): the harm to consumer is a linear transformation of inconsistency. As we show in the market experiment, other measures are possible, such as paying in proportion to the size of the violation. A possible alternative would be introducing some measure of money pump (Echenique et al. 2013) or a measure of welfare loss with respect to the maximum achievable payoff (which could be particularly relevant if the task is generalized allowing for multiple packages). The reason why our first study did not include such complications is that the policy request that motivated the study imposed some constraints. Priorities are often different between the policy domain and academia. For instance, the manipulation of the price became less important than the "realism" of the package, or the test of different sentences in the protective measures. We believe in evidence-based policy, and these are trade-offs that we are willing to face. Our experience is that the use of stated preferences and the use of a simple dichotomic outcome make the test more persuasive in non-academic communication.

Findings from experiment II illustrate how the transaction test improves on tests which measure behavioral change. There, together with a conventional dark pattern version of TWE, we introduced a treatment designed to steer choices towards data protection (TWE Bright Pattern) and showed that it significantly and substantially reduces inconsistency. A test based on behavioral change would have detected significant impacts of both TWE Dark Pattern and TWE Bright Pattern. So did the transaction test. But, since the transaction test is explicit on its behavioral assumptions, it was able to differentiate its recommendations between the two versions of TWE, even though their appearance and underlying manipulative technique are the same.

Part of the credibility of the transaction test relies on its flexible structure, whereby relevant framings can be introduced to elicit the

Table 5
Summary of the hypotheses.

Treatment	Change in the share of inconsistency (percentage points)	Source
H1: Impact of Dark Pattern, Average Consumer		
Hidden Information	12.25	Experiment I
TWE	6.02	Experiment I
Personalization	9.84	Experiment I
Hidden Information	42.64	Experiment II
TWE	17.03	Experiment II
H2: Impact of Dark Pattern, Vulnerable Consumer		
Hidden Information	5.80	Experiment I
TWE	n.s	Experiment I
Personalization	5.50	Experiment I
H3: Impact of Protective Measure, Average Consumer		
Cool Down v. Hidden	n.s.	Experiment I
Cool Down v. TWE	n.s.	Experiment I
Cool Down v. Personalization	n.s.	Experiment I
Cool Down-disclosure v Personalization	n.s.	Experiment I
TWE (Bright Pattern) v. Control	-15.79	Experiment II
H4: Impact of Protective Measure, Vulnerable Consumer		
Hidden + Cool Down	n.s.	Experiment I
TWE + Cool Down	n.s.	Experiment I
Personalization + Cool Down	n.s.	Experiment I
Personalization + Disclosure	n.s.	Experiment I

social expectations implicit in a real-world setting that the policy question addresses. In our experience, these drivers of external validity include some element of *overload*, achieved through distracting stimuli, an intermediate task etc. We stress that this is essentially done for external validity, while our findings show that these are not the key driver behind the effect of dark patter. Had that be the case, the introduction of incentivized time pressure would have led to larger – not smaller – effect sizes.

7.3. Welfare harm

In experiment I, due to how incentives are computed, namely that consistency is paid ten times an inconsistent choice, monetary losses due to dark patterns are proportional to the treatment effect on the rate of inconsistency. In experiment II, the strongest dark patterns costs participants 25 cents on the euro, while the loss is lower for other dark patterns. This illustrative example of how to conduct welfare analysis shows the flexibility of the transaction test and the possibility to perform cost-benefit assessments. Nevertheless, this is out of the scope of this paper, whose core aim is to translate existing regulatory provisions into an experimental methodology.

Specific features of the choice domain influence the potential welfare consequences. Platforms deploy dark patterns at scale, thus extracting rents on each transaction, which may be small for the individual consumer but add up to significant welfare harm due to the scale of operation. Within this context, revealing the presence of even small individual welfare losses becomes particularly relevant because the situation resembles a collective action problem: customers may notice dark patterns and feel annoyed by them, but may fail to act because the individual loss may not justify the effort. A recent contribution by two of the authors explores institutions facilitating coordinated action to circumvent this problem (Bogliacino et al. 2025).

7.4. Measurement error

The structure of the transaction test may raise questions over the potential bias introduced by the first task when preferences are self-reported. Moreover, the use of two elicitation methods (first in the preference and then in the choice tasks) may inflate the rate of inconsistency, as has been previously observed in the debate on preference reversal (Grether and Plott 1979) or the privacy paradox (Solove 2021). We address this critique in what follows.

In the first experiment, participants are incentivized to behave consistently, they are paid a larger sum, A, if they choose consistently with previously stated preferences, M_f, and a smaller sum, a, if they do not. Even assuming that M_f is reported with error as $M_f^* + \varepsilon$, with M_f^* being the true value, the incentive system only depends on the reported value, M_f, and not the true value, M_f^*. Thus, recall error, rather than noise, matters. As we reported in the robustness checks in Section 4, recall cannot substantively affect our findings: in an intermediate task, we directly measure recall through an incentivized task. We find that all our participants recall within a 40 % margin of the price difference they will face in Task II.

Moreover, when we move from stated M_f in experiment I to randomly assigned M_f in experiment II, any source of measurement error besides recall is removed by design. The market experiment thus shuts down this channel. The transaction test reaches the same conclusions in both settings: the second study largely confirms the main results from the first study, detecting larger effects.

Finally, the transaction test is an assessment in change, not in level. It does not judge website customizations by the share of people who choose consistently, but by its relative performance with respect to a control condition. Randomization is taking care of potential confounds, especially for sample sizes as the ones in our first experiment.

Another potential source of measurement error can be due to the exact instructions used at the moment of choice in the second task. We used a very simple text: "Now make your choice. Select the package you prefer", which does not remind subjects about the incentive structure. Although the task had been clearly explained and tested for comprehension, our phrasing here may have increased the baseline level of inconsistency. Nevertheless, instructions were kept constant across conditions, thus not affecting the estimation of treatment effects.

7.5. Dark patterns affecting consumers' preferences

The results of the transaction test would be biased if the introduction of a dark pattern changes consumer preferences.

The transaction test measures consistency in terms of a specific characteristic's value (M_f, p_f), while goods are vectors with many characteristics. Most dark patterns, and certainly those studied here, modify the consumer's choice journey, not the narrative around specific characteristics. For instance, website customizations trick consumers over luggage options, data privacy, or automatic renewal, but they are not meant to persuade consumers over luggage, privacy or renewal. As a result, preference change is seldom an issue.

Both experiments in this paper shut down the channel of preference change. Incentives systems transparently depend on *correctly recalling* the stated preferences or the assigned monetary valuations. In fact, the very issue of preference change materializes only in settings where both Task I and Task II are incentivized based on outcomes (see below).

7.6. Helping consumers make consistent choices

The extent of inconsistency in the control is the same in the first and second study. It is also very similar to the rate of inconsistency between elicited WTP and data-money exchange (37 %) in the work by Lee and Weber (2024) and to the rate originally found in the experiments over preference reversal (Grether and Plott 1979): if we aggregate over P- and $-bet, the rates in the three experiments are 33 %, 37 %, and 33 %. Even allowing for some measurement error, there is space for improving choices, through better design and through protective measures. The experimental evidence presented in this article documents that a transparency-based remedy does not provide an effective solution, at least for the setting we analyzed. The "cool down" is an implementation of the idea that a "sludge" can also be favorable to consumers (Sunstein 2022) and can be related to the concept of decision-points (Soman et al. 2019; 2010). The behavioral assumption underlying this approach is that impulsivity of the decision-maker drives the effectiveness of the choice architecture. In fact, Sin et al. (2022) listed cool down (reflection) as a behavioral intervention against purchase impulsivity, together with distraction and postponement.

The results from our experiments do not seem to support this System I-System II (Kahneman 2011) interpretation of deceptive design (Luguri and Strahilevitz 2021b; UK Competition and Market Authority 2022; Mathur et al. 2019b). The dual process model posits that participants tend to make impulsive decisions most of the time, even though they are equipped with an alternative deliberative system that processes all the information, because the latter is only activated when the incentives are strong enough. Under a dual process explanation, two stylized facts should hold. First, the impact of dark patterns should be stronger for vulnerable consumers than for average consumers. Second, the best protective measure should indeed be a reflection intervention (Sin et al., 2022). Instead, in our data, the impact of dark patterns for average consumers is one and a half to two times higher than for vulnerable consumers and the cool-down intervention was ineffective. Future research should explore alternative behavioral explanations of the efficacy of manipulative practices.

7.7. Separate incentive compatible elicitation of preferences and choices

The two experiments reported in this paper address framed goods. A companion experiment reported in the SOM assesses the robustness of the transaction test in a domain where both preference elicitation and choice can be incentivized. As we indicated in Section 2, this is one of the three alternative implementations of the transaction test.

In this additional lab experiment, the main choice involves lotteries with binary outcomes (a reward and a loss), which can be incentivized at both stages of the transaction test. Fixing the loss, the "contractual terms" of the binary lottery become the probability of winning and the size of the reward. In the adapted first stage of the transaction test, we use the Multiple Price List (MPL) task to elicit the trade-off between these contractual terms. Participants complete four versions of this task. In the adapted second stage of the transaction test, the subjects perform a set of dichotomic choices over randomly selected rows from the MPLs. The second stage can be repeated multiple times with different conditions, without revealing our aim. By selecting random rows from the MPLs, we observe choices at varying distances from the indifference point, both below and above. We randomly assign participants to different sequences of conditions for the binary choices, which include a neutral control, a version of TWE, and a "default" treatment (a colored button). We could not introduce hidden information or personalization in this environment.

Our findings indicate that the share of inconsistency in the control reaches 37 %. Under dark patterns, the share increases by 4 pp, mostly in the TWE treatment. While subjects become less inconsistent in the control in later rounds (from 41 % to 33 %), they do not under dark patterns (from 40 % to 41 %). Inconsistency for binary choices close to the indifference point is similar across conditions (45 %). Moving away from the indifference point, the level of inconsistency drops for both control and dark patterns, but significantly more for the former (32 % against 38 %). A detailed explanation of the design and the results is presented in SOM, Section VI.

7.8. Concluding remarks

This article introduces the transaction test: a tool to determine whether a practice steers consumers to make transaction decisions that contradict their individual preferences, operationalizing the legal definition of unfair commercial practice. The transaction test allows for a controlled assessment of the practice and transforms the measurement of manipulation into a straightforward empirical test. Our market experiment shows that the transaction test is a general and robust approach. Online platforms continuously introduce new deceptive designs and deploy them at scale (Luguri and Strahilevitz 2021b; Mathur et al. 2019b; Bösch et al. 2016b; Cara 2019, BEUC, 2022). Our method helps researchers and policymakers to decide when they represent a potential threat.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

Funding: *EU Consumer Programme (2014-2020) under a service contract with the European Innovation Council and SMEs Executive Agency (EISMEA) acting under the mandate from the European Commission. Behavioural Study on Unfair Commercial Practices in the Digital Environment (European Commission - Framework Contract with reopening of competition - Behavioural studies),* a French government grant managed by the *Agence Nationale de la Recherche* under the framework of the *Investissements d'avenir programme reference ANR-17-EURE-001, internal funding from Università Cattolica del Sacro Cuore, from PredictBy and from Open Evidence.* We thank the Editor - Daniele Nosenzo - and three anonymous referees for their comments and suggestions. We are grateful to Egelyn Braun, Hannah Nohlen, and Silvia Pella for comments during the project. A special thanks to Marcello Negrini and Agne Kajackaite for commenting on a previous draft. We thank Cristiano Codagnone, Teresa Rodríguez de las Heras Ballell and Lucie Lechardoy for help and discussion during the writing of the companion report. A special thanks to Alba Boluda for excellent research assistance throughout the project. We thank Setecem for programming the online experiment and Schlesinger group for data collection. We thank the lab at PSE and in particular Maxim Frolov for assistance for the experiment reported in the SOM. We are grateful for comments received at BEBES in Bogotá, the 4th Winter Workshop in Bergamo, the 12th BEEN Meeting, the BEEC conference in Bogota, the 4th IBPPC in London and during seminars at the Behavioral Public Policy course at UPenn, at PSE-Paris I, Charles University, Universidad del Desarrollo, Universidad de Navarra, and an internal workshop at the Università Cattolica. The information and views set out in this article are those of the author(s) and do not necessarily reflect the official opinion of the Commission/Executive Agency. The Commission/Executive Agency do not guarantee the accuracy of the data included in this study. Neither the Commission/Executive Agency nor any person acting on the Commission's/Executive Agency's behalf may be held responsible for the use which may be made of the information contained therein.

The Open Science Framework (OSF) page of this project (https://osf.io/wuanh/) contains additional data and code of the experiments. A previous version of this manuscript circulated under the title: "Testing for Manipulation: Experimental Evidence on Dark Patterns". The authors declare no conflict of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2026.107469.

Data availability

Data are available

References

Acquisti, A., Taylor, C., Wagman, L0, 2016. The economics of privacy. J. Econ. Lit. 54 (2). https://doi.org/10.1257/jel.54.2.442. Preprint, American Economic Association, June 1.

Alós-Ferrer, Carlos, Garagnani, Michele, 2020. The cognitive foundations of cooperation. J. Econ. Behav. Organ. 175, 71–85. https://doi.org/10.1016/j.jebo.2020.04.019.

Athey, S., Catalini, C., & Tucker, C. (2017). The digital privacy paradox: Small money, small costs, small talk (23488; NBER WORKING PAPER SERIES).

Benndorf, Volker, Normann, Hans Theo, 2018. The willingness to sell personal data. Scand. J. Econ. 120 (4), 1260–1278. https://doi.org/10.1111/sjoe.12247.

BEUC. 2022. "' Dark patterns ' AND the Eu consumer law acquis".

Bilancini, Ennio, Boncinelli, Leonardo, Celadin, Tatiana, 2022. Social value orientation and conditional cooperation in the online one-shot public goods game. J. Econ. Behav. Organ. 200, 243–272. https://doi.org/10.1016/j.jebo.2022.05.021.

Blake, Tom, Moshary, Sarah, Sweeney, Kane, Tadelis, Steve, 2021. Price salience and product choice. Mark. Sci. 40 (4), 619–636. https://doi.org/10.1287/mksc.2020.1261.

Bogliacino, F., Charris, R., Codagnone, C., Folkvord, F., Montealegre, F., Lupiáñez-Villanueva, F., 2025. Unfair commercial practices in a pit market: evidence from an artefactual field experiment. Behav. Public Policy 9 (2), 443–460. https://doi.org/10.1017/bpp.2022.33.

Bösch, Christoph, Erb, Benjamin, Kargl, Frank, Kopp, Henning, Pfattheicher, Stefan, 2016. Tales from the Dark side: privacy dark strategies and privacy dark patterns. Proc. Priv. Enhancing Technol. 2016 (4), 237–254. https://doi.org/10.1515/popets-2016-0038.

Brignull, Henry. 2010. "Deceptive Design." Https://Www.Deceptive.Design/.

Calo, Ryan., 2014. Digital market manipulation. George Wash. Law Rev. 82 (4), 995–1051. https://doi.org/10.2139/ssrn.2309703.

Cara, Corina., 2019. Dark Patterns in the Media: A systematic review. Netw. Intell. Stud. VII (14), 105–113.

Cialdini, Robert B., 2009. Influence: Science and Practice, 5th ed. Pearson Education.

Dertwinkel-Kalt, Markus, Köster, Mats, Sutter, Matthias, 2020. To buy or not to buy? Price salience in an online shopping field experiment. Eur. Econ. Rev. 130, 103593. https://doi.org/10.1016/j.euroecorev.2020.103593.

Diaz, Lina, Daniel Houser, John Ifcher, and Homa Zarghamee. 2021. "Estimating social preferences using stated satisfaction: novel support for inequity aversion." *Ssrn*, no. 14347. https://doi.org/10.2139/ssrn.3846691.

Echenique, F., Lee, S., Shum, M., 2013. The money pump as a measure of revealed preference violations. J. Polit. Econ. 119 (6), 1201–1223. https://doi.org/10.1086/674077.

European Commission, 2021. Guidance on the Interpretation and Application of Directive 2005/29/EC of the European Parliament and of the Council Concerning Unfair Business-to-Consumer Commercial Practices in the Internal Market. Off. J. Eur. Union.

European Commission, Directorate-General for Justice and Consumers, F Lupiáñez-Villanueva, A Boluda, F Bogliacino, G Liva, L Lechardoy, and T de las Heras Ballell. 2022. *Behavioural study on unfair commercial practices in the digital environment : dark patterns and manipulative personalisation : final report*. https://doi.org/10.2838/859030.

Falk, Armin, Becker, Anke, Dohmen, Thomas, Enke, Benjamin, Huffman, David, Sunde, Uwe, 2018. Global evidence on economic preferences*. Q. J. Econ. 133 (4), 1645–1692. https://doi.org/10.1093/qje/qjy013.

Federal Trade Commission. 1984. *FTC Policy Statement on unfairness*.

Geronimo, Linda Di, Braz, Larissa, Fregnan, Enrico, Palomba, Fabio, Bacchelli, Alberto, 2020. 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. CHI '20. New York, NY, USA. Association for Computing Machinery. https://doi.org/10.1145/3313831.3376600.

Goldfarb, A., Que, V.F., 2023. The Economics of Digital Privacy. Annu. Rev. Econ. 15, 267–283. https://doi.org/10.1146/annurev-economics.

Grassl, Paul, Schraffenberger, Hanna, Borgesius, Frederik J.Zuiderveen, Buijzen, Moniek, 2021. Dark and bright patterns in cookie consent request. J. Digit. Soc. Res. 3 (1), 1–35.

Grether, David M, Plott, Charles R, 1979. Economic theory of choice and the preference reversal phenomenon. Am. Econ. Rev. 69 (4).

Gu, Yiquan, Wenzel, Tobias, 2015. Putting on a tight leash and levelling playing field: an experiment in strategic obfuscation and consumer protection. Int. J. Ind. Organ. 42, 120–128. https://doi.org/10.1016/j.ijindorg.2015.07.008.

Gu, Yiquan, Wenzel, Tobias, 2020. Curbing obfuscation: empower consumers or regulate firms? Int. J. Ind. Organ. 70, 102582. https://doi.org/10.1016/j.ijindorg.2020.102582.

Hartzog, Woodrow., 2018. Privacy's Blueprint: The Battle to Control the Design of New Technologies. Harvard University Press, Cambridge, Mass.

Henninger, Felix, Shevchenko, Yury, Mertens, Ulf K., Kieslich, Pascal J., Hilbig, Benjamin E., 2022. Lab.Js: A free, open, online study builder. Behav. Res. Methods 54 (2), 556–573. https://doi.org/10.3758/s13428-019-01283-5.

Huck, Steffen, Wallace, Brian, 2015. The Impact of Price Frames on Consumer Decision Making A Price Frame Refers to the Way a Price Is Presented. Mimeo.

Kahneman, Daniel. 2011. "Thinking fast, Thinking slow." *Interpretation, Tavistock, London*.

Kalayci, Kenan, Potters, Jan, 2011. Buyer confusion and market prices. Int. J. Ind. Organ. 29 (1), 14–22. https://doi.org/10.1016/j.ijindorg.2010.06.004.

Kalaycı, Kenan., 2016. Confusopoly: competition and obfuscation in markets. Exp. Econ. 19 (2), 299–316. https://doi.org/10.1007/s10683-015-9438-z.

Kaptein, Maurits, De Ruyter, Boris, Markopoulos, Panos, Aarts, Emile, 2012. Adaptive Persuasive Systems: A study of tailored persuasive text messages to reduce snacking. ACM Trans. Interact. Intell. Syst 2 (2). https://doi.org/10.1145/2209310.2209313.

Lee, Yi-Shan, Weber, Roberto A., 2024. Revealed privacy preferences: are privacy choices rational? Manag. Sci. https://doi.org/10.1287/mnsc.2022.00807. June.

Luguri, Jamie, Strahilevitz, Lior Jacob, 2021. Shining a light on dark patterns. J. Leg. Anal. 13 (1), 43–109. https://doi.org/10.1093/jla/laaa006.

Mathur, Arunesh, Acar, Gunes, Friedman, Michael J., Lucherini, Elena, Mayer, Jonathan, Chetty, Marshini, Narayanan, Arvind, 2019. Dark patterns at scale: findings from a crawl of 11K shopping websites. In: Proceedings of the ACM on Human-Computer Interaction, 3. https://doi.org/10.1145/3359183.

Matz, S.C., Kosinski, M., Nave, G., Stillwell, D.J., 2017. Psychological targeting as an effective approach to digital mass persuasion. Proc. Natl. Acad. Sci. U. S. A. 114 (48), 12714–12719. https://doi.org/10.1073/pnas.1710966114.

Miller, Klaus, Navdeep S Sahni, and Avner Strulov-Shlain. 2023. "Sophisticated consumers with inertia: long-term implications from a large-scale field experiment." https://ssrn.com/abstract=4065098.

Mills, Stuart, Whittle, Richard, Ahmed, Rafi, Walsh, Tom, Wessel, Martin, 2023. Dark patterns and sludge audits: an integrated approach. Behav. Public Policy. https://doi.org/10.1017/bpp.2023.24.

Netherlands Authority for Consumers & Markets. 2020. "Boundaries of online persuasion".

Niederle, Muriel (2025) Experiments: why, how, and a users guide for producers as well as consumers. NBER working paper 33630.

Norman, Warren T., 1963. Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality ratings. J. Abnorm. Soc. Psychol. 66 (6), 574–583. https://doi.org/10.1037/h0040291.

Norwegian Consumer Council. 2018. "Deceived by design".

Rasch, Alexander, Thöne, Miriam, Wenzel, Tobias, 2020. Drip pricing and its regulation: experimental evidence. J. Econ. Behav. Organ. 176, 353–370. https://doi.org/10.1016/j.jebo.2020.04.007.

Richards, Neil, Hartzog, Woodrow, 2019. The pathologies of digital consent. Wash. Univ. Law Rev. 96, 1461.

Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73 (4), 1237–1282. https://doi.org/10.1111/j.1468-0262.2005.00615.x.

SERNAC. 2021. "Informe de Resultados de levantamiento de Dark Patterns." Vol. 2021.

SERNAC. 2022. "Consentimiento en El Uso de Cookies: Evidencia Experimental Sobre El Impacto de La Privacidad por Defecto y Los Patrones Oscuros en Las Decisiones de Los Consumidores".

Shevchenko, Yury., 2022. Open Lab: A web application for running and sharing online experiments. Behav. Res. Methods 54 (6), 3118–3125. https://doi.org/10.3758/s13428-021-01776-2.

Sin, Ray, Harris, Ted, Nilsson, Simon, Beck, Talia, 2022. Dark patterns in online shopping: do they work and can nudges help mitigate impulse buying? Behav. Public Policy 1–27. https://doi.org/10.1017/bpp.2022.11. May.

Smith, Vernon L., 1976. Experimental Economics: induced value theory. Am. Econ. Rev. 274. https://doi.org/10.1017/cbo9780511528354.008.

Solove, D.J., 2021. The myth of the privacy paradox. George Wash. Law Rev. 89 (1), 1–51. https://perma.cc/6NMX-83JF.

Soman, D., Cowen, D., Kannan, N., & Feng, B. (2019). Seeing sludge: towards a dashboard to help organizations recognize impedance to end-user decisions and action.

Soman, D., Xu, J., Cheema, A., 2010. Decision Points: A Theory Emerges. Rotman Management Magazine, pp. 1–4.

Sunstein, Cass R., 2022. Sludge Audits. Behav. Public Policy 6 (4), 654–673. https://doi.org/10.1017/bpp.2019.32.

Thaler, Richard H., 2018. Nudge, not sludge. Science 361 (6401), 431. https://doi.org/10.1126/science.aau9241.

Thaler, Richard H., Benartzi, Shlomo, 2004. Save more tomorrow: using behavioral economics to increase employee saving. J. Polit. Econ. 112 (S1), S164–S187. https://doi.org/10.1086/380085.

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE UNION. 2005. *DIRECTIVE 2005/29/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL "Unfair Commercial Practice Directive"*.

UK Competition and Market Authority. 2022. "Evidence review of online choice architecture and consumer and competition harm".

Waldman, Ari Ezra, 2020. Cognitive biases, dark patterns, and the 'privacy paradox. Current Opinion in Psychology. Elsevier B.V. https://doi.org/10.1016/j.copsyc.2019.08.025

Woods, Stephen A, Hampson, Sarah E, 2005. Measuring the big five with single items using a bipolar response scale. Eur. J. Personal. 19 (5), 373–390. https://doi.org/10.1002/per.542.

Zac, Amit, Huang, Yu Chun, Moltke, Amédée Von, Decker, Christopher, Ezrachi, Ariel, 2025. Dark patterns and consumer vulnerability. Behav. Public Policy. https://doi.org/10.1017/bpp.2024.49.