the behaviouralist

citizens advice

Online Choice Architecture and Vulnerability: Understanding the Impact of Reference Pricing on Consumer Behaviour

Research Report December 2025

Table of contents

Foreword by The Behaviouralist

Consumers now make many of their everyday choices in digital environments. Smartphones and computers are routinely used to purchase clothes, book accommodation, select utility suppliers, connect with friends, and consume news. While these developments have brought substantial benefits, they have also created new challenges. In particular, there is an urgent need to evaluate the “online choice architecture” (OCA) practices employed by firms, and to determine whether such practices harm consumers or distort competition.

Ideally, policymakers would assess the impact of an OCA by running natural field experiments on the relevant platform—randomly removing the practice and measuring the resulting changes in user behaviour. Although feasible in some circumstances, platform operators are often unwilling or unable to facilitate such tests. Fortunately, a credible and practical alternative exists: high-fidelity simulation experiments. These experiments draw on behavioural science, user-experience design, and economics to create realistic, interactive replicas of digital environments. Researchers recruit representative users, randomly assign them to different OCA conditions or policy interventions, and observe their behaviour. The strong environmental validity of this method allows for the generation of highly robust, context-specific evidence.

This report uses a simulation experiment to examine the effects of three types of OCA practices: reference pricing, sensory manipulation, and time-bound elements. We hope that readers will gain a deeper understanding of how these practices shape consumer behaviour and welfare, and that this study will encourage further methodological innovation and greater empirical scrutiny of the many OCA practices that remain understudied.

the behaviouralist

Jesper Akesson
Managing Director

Foreword by Citizens Advice

Design isn't neutral. For years, Citizens Advice has been working to show that the design of consumer interfaces - known as online choice architecture or OCA - plays a big role in the decisions we make. We can measure the harm this causes in many different ways - with people stuck paying for subscriptions they don't want or need, paying more than they intended while shopping online, and wasting countless hours trying to obtain refunds for goods or services they didn't even mean to purchase in the first place. But knowing and recognising that OCA can harm consumers still leaves us with a lot more questions. How and why do OCA practices impact consumer decisions? Do they impact everyone in the same way? Are some groups of consumers more likely to experience harm?

Last year we began exploring these questions, and published a report written by the Behavioural Insights Team (BIT) exploring the relationship between OCA and vulnerability. This research uncovered a gap in the evidence when it comes to understanding how consumers in vulnerable circumstances are impacted by OCA practices. But without this evidence, regulators are limited in how they can understand the effect of OCA practices and how effectively they can protect consumers from harm. This risks some harmful practices going undetected - despite the risk of disproportionate harm to consumers in vulnerable circumstances.

This research aims to fill in some of the knowledge gaps we identified in our previous report, and adds to the evidence available to regulators to combat OCA harms. With The Behaviouralist, we investigated the impact of reference pricing. We wanted to understand more about how this practice affects consumer decision making, particularly for consumers in vulnerable circumstances. We were specifically interested in understanding what - if any - impact reference pricing has on consumers with mental health problems and experiencing financial insecurity. The experiences of these groups are of particular interest to Citizens Advice as they make up a big portion of the people we help through our frontline services, but also because of the way mental health or financial insecurity may shape the kinds or scale of detriment people experience. Mental health problems and financial insecurity are prevalent in the general population too, and we wanted to give regulators real evidence on how to protect these groups.

With so much consumer spending happening online, it is vital that regulators know who is losing out because of OCA and act to level the playing field. There's still much we don't know about OCA practices and their impacts, but we hope this provides regulators with a solid foundation to build upon.

Chloë Maughan

Principal Policy Manager, Consumer Policy

Executive Summary

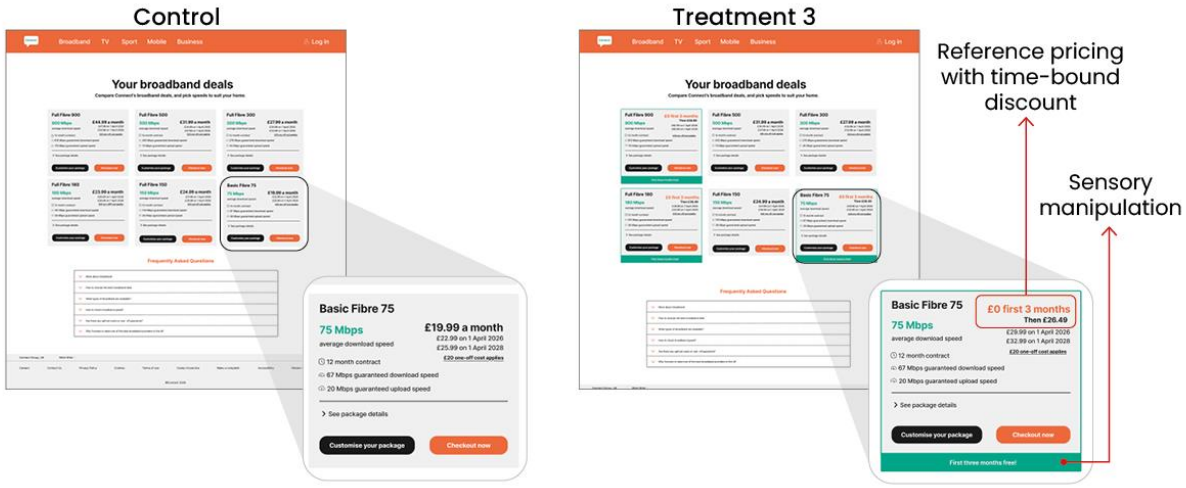
This study examines how three common online sales practices – reference pricing, sensory manipulation, and time-bound elements – influence people's choices and well-being when shopping online. Reference pricing refers to showing a higher previous or future price next to the current price (for example, "was £22.99, now £19.99"). Sensory manipulation refers to using eye-catching design on certain deals (for example, bold colours or large fonts). Time-bound elements include countdown timers (for example, "20 seconds left on this offer") and time-limited discounts (for example, "first 3 months free").

These practices are widely used, particularly in three types of online marketplaces: broadband plans, gym memberships, and hotel bookings. We replicated these marketplaces to test how the practices work under real-world conditions. Specifically, we set out to measure whether they reduce decision quality, whether people facing financial insecurity or poor mental health are affected differently, and whether these practices discourage shopping around, which matters for competition and consumer welfare. Our findings are intended to support better design and oversight of online marketplaces, so the benefits of digital commerce are widely accessible while the risks of confusion or manipulation are reduced.

To study the effects of these practices, we built a realistic simulation experiment that took place across these three marketplaces. Participants were asked to shop and select the best offer in each one. A nationally representative sample of over 8,000 adults took part, including large sub-samples of people facing financial insecurity and poor mental health. Participants were entered into a lottery for each correct choice so they were motivated to behave much more similarly to how they would when shopping online.

People were randomly assigned to one of four groups: (1) reference pricing only; (2) reference pricing plus sensory manipulation; (3) reference pricing plus sensory manipulation and a time-bound element; or (4) a control group that saw none of these practices. Examples of how the control group vs. the reference pricing with sensory manipulation and time-bound elements looked in the broadband context is illustrated in Figure 1. We measured decision quality (did they pick the best-value option), time spent shopping, confidence in selected offers, shopping-around (whether they visited a competitor site), and the monetary amount spent shopping.

Figure 1. Control Group vs Treatment 3



What we found

- **Reference pricing reduced consumer decision quality**. People shown "was/now" prices were less likely to choose the best-value option.
- **Adding sensory manipulation and time-bound elements typically strengthened the negative effect of reference pricing,** but this effect varied by market.
- **Vulnerable people started from a lower baseline decision quality**. People experiencing poor mental health or financial insecurity were less likely to pick the best-value option in the control group, but their choices were not disproportionately affected by reference pricing, relative to others.
- **There is suggestive evidence that a combination of reference pricing, sensory manipulation and time-bound elements made vulnerable people** spend more time picking a broadband plan, relative to others.
- **Broadband decisions were hard for everyone, even without reference pricing**. The complexity of broadband plans seemed to be the reason.
- **Reference pricing reduced shopping-around behaviour**. Reference pricing on its own reduced the visits of participants to a competitor site.
- **Confidence and accuracy moved in opposite directions,** with people who picked the best-value option feeling less confident in their selection. Even in the absence of reference pricing, sensory manipulation and time-bound elements, people who were more confident generally made poorer choices.
- **Reference pricing countered the advantages of bargain-hunting and financial literacy.** Self-described bargain-hunters and those with high financial literacy started out making more correct choices, but reference pricing, especially when combined with sensory manipulation and time-bound elements, eroded that advantage.
- **Reference pricing tends to make consumers overspend, not underspend.** This effect got stronger when combined with sensory manipulation and time-bound elements.

Based on these results, we recommend that regulators do the following:

- **The Competition and Markets Authority** (CMA) should conduct further investigations into the use and impact of reference pricing across markets. Previously, the CMA have done market-specific investigations on reference pricing and published relevant principles for firms (CMA, 2024), but these focused primarily on whether or not the price comparison is genuine. Our research shows that reference pricing can harm consumer decision making regardless of whether the information shown is fairly presented. We recommend that the CMA explore the use of reference pricing across markets and that they expand and reframe their guidance for firms around consumer harm rather than whether or not information is genuine.

- **Ofcom** should investigate broadband pricing practices to address the serious difficulties consumers face in making the right choice in this market. They should consider our experimental findings on the harm caused by OCA practices, but also the findings from our control group which show how difficult it is for consumers to make good decisions in this market in the first place.
- **All the regulators we engaged** throughout the design of this experiment should continue to conduct research on the impact of OCA practices, especially on vulnerable consumers, to address the gap in existing evidence.

1 Introduction

With a growing share of economic activity moving online, online choice architecture (OCA)—the way information, products, and services are presented in digital marketplaces —has become an increasingly important focus for research and policy (CMA, 2022; BIT, 2024). A body of literature documents widespread use of tactics such as reference pricing (e.g., "was £22.99, now £19.99"), sensory manipulation (e.g., using bold colours, or large fonts on certain deals), and time-bound elements (e.g., "only 20 seconds left on this offer") in digital markets (Mathur et al., 2019; Di Geronimo et al., 2020). What remains less well understood are the causal effects of these specific practices on consumer choices, both among the general population and specifically for vulnerable groups.

This study aims to answer two questions. First, how do prevalent OCA practices, specifically reference pricing and its combination with sensory manipulation and time-bound elements, affect the probability of consumers purchasing objectively optimal offers? And, conversely, do these practices increase the likelihood of detrimental choices, such as selecting decoy options? Do the effects differ for vulnerable participants, particularly those experiencing financial insecurity or poor mental health? Second, how do these practices affect consumer behaviours related to competition, proxied by consumers' propensity to shop around?

Guided by desk research on typical online marketplaces, we implemented a preregistered online randomised controlled trial that mirrored real shopping journeys across three marketplaces: broadband plans, gym memberships, and hotel bookings. A large UK sample (nationally representative in terms of age, gender, and region) was recruited alongside vulnerable samples experiencing financial insecurity or poor mental health. All participants completed the three shopping tasks. Each marketplace was paired with a specific shopping scenario and participants were instructed to select the offer that best fits that scenario. In every marketplace, participants saw four to six offers, including (i) an optimal offer that met all scenario requirements at the lowest necessary cost and (ii) an inferior decoy offer. In this report, we define 'decoys' as options that are dominated (or close to dominated), in line with Which? (2024). In other words, there is no set of (rational) preferences that would justify purchasing the decoy option in either marketplace, given the other offers available. Participants could also click a competitor banner to shop around in an alternative marketplace with similar offers, allowing us to examine effects on competitive behaviour.

Participants were randomly, with equal probability, assigned to one of four conditions: control (that did not feature any of the 3 OCA practices), T1 with reference pricing only ("was £22.99, now £19.99" on selected suboptimal deals), T2 featuring the reference pricing above plus sensory manipulation (bold colour frames of those reference-priced options), and T3 having all elements of T2 plus either a countdown timer or a time-limited discount such as "first 3 months free," depending on the market. All other aspects of layout, journey, and offer attributes were held constant across arms.

First, we found that OCA practices harmed decision quality. Relative to control, exposure to any treatment group lowered the probability of selecting the optimal options by 1.7 percentage points and increased decoy selection by 2.0 percentage points. Sensory manipulation amplified the effects of reference pricing. The time-bound elements had mixed effects by market: in gyms, adding a countdown timer further reduced correct choices; in hotels, we found no significant effects, while in broadband, a “first 3 months free” discount paradoxically raised correct choice rates relative even to control. Across conditions, broadband was generally difficult: even in the control condition only about 1 in 10 participants chose the optimal plan, suggesting that evaluating broadband tariffs might be intrinsically hard, even independent of reference pricing, sensory manipulation, and time-bound elements.

Second, participants experiencing financial insecurity or poor mental health under-performed the general population. Vulnerable consumers were 5.0 - 7.9 percentage points less likely to choose the right products than the general population and were also equally affected by reference pricing and the other OCAs (when compared to the general population). In other words, we did not detect disproportionate harm from OCA among vulnerable groups but found that they start from a lower baseline of correct choice, making them systematically more exposed to poor outcomes online.

Third, we found some evidence that OCAs can hinder market competition, but only under specific conditions. Participants in T1 (pure reference pricing) were 2.3 percentage points less likely to click on a banner to access a competitor site, consistent with reduced shopping around. However, this finding did not extend to more complex treatment conditions T2 and T3. There was also considerable heterogeneity in shopping-around behaviours, with individuals reporting higher stress or poor mental health especially unlikely to browse alternatives, while self-identified bargain-hunters and participants with high levels of financial literacy more likely to click on alternative marketplace banners. Taken together, the results indicate that common OCA practices can lead participants to make detrimental choices, and that they also have a potential to harm market competition.

Our study adds to the existing evidence on online choice architecture and vulnerability in five concrete ways. First, it builds on theoretical foundations in BIT (2024), by experimentally testing whether vulnerable customers disproportionately suffer poor outcomes as a result of OCA. Also, building on Lupiáñez-Villanueva et al. (2022), Luguri & Strahilevitz (2021) and Zac et al. (2023), our study moves beyond demographic proxies of vulnerability, capturing financial insecurity via income levels and mental-health status via a validated MHI-5 scale. Second, we move beyond stated intentions to revealed choices in realistic online marketplaces. Instead of hypothetical ratings or simple accept/decline tasks, participants completed full shopping journeys on realistic marketplace replicas, and we evaluated objective, welfare-relevant outcomes: selecting the cheapest option that met pre-specified needs, the uptake of decoy offers, and shopping-around patterns. This design substantially strengthens inference relative to classic reference pricing studies (e.g., Lichtenstein, Burton & O’Hara, 1988; Lichtenstein, Burton & Karson, 1991; Jensen et al., 2003; Krishnan et al., 2013; Kan et al., 2014) and complements existing work on visually manipulative design (e.g., Milosavljevic et al., 2012; Clement et al., 2015; Lupiáñez-Villanueva et al., 2022; Zac et al., 2025).

Third, we estimate incremental and joint effects of tactics used together—reference pricing on its own (T1), with added sensory manipulation (T2), and with an additional timing element (T3)—reflecting how OCA practices typically appear in combination on actual websites (CMA, 2022). Fourth, we extend the literature to competition outcomes by measuring "shopping-around" (click-outs to, and purchases on, a rival site), directly testing the concern that OCA hinders market competition. Fifth, we study multiple markets, broadband, gyms and hotels, showing both common patterns and important context-specific differences; notably, broadband proved generally difficult, with about 90% error rate even in the control group, highlighting the role of product complexity independent of OCA. Together, this research provides policy-relevant, ecologically valid evidence: sensory manipulation amplifies the harms of reference pricing, time-bound elements have mixed effects by market, subtle reference prices can reduce comparison shopping, and vulnerable consumers, while not disproportionately harmed by the selected OCA practices, start from a lower baseline of correct choice, underscoring the need for regulatory attention (CMA, 2022; BIT, 2024).

The rest of the report is structured as follows. Section 2 details the experimental design, treatments, outcomes, sampling, and validity checks. Section 3 presents the primary results on decision quality and heterogeneity by vulnerability, as well as secondary outcomes on shopping around and market competition. Section 4 discusses our findings and places them within the existing literature.

2 Methodology

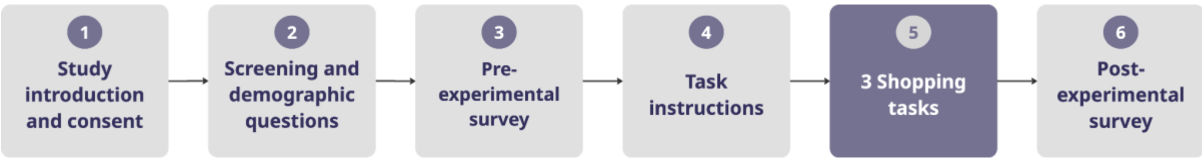
The methodology employed in this study consisted of an online experiment that investigated how reference pricing and its combination with other online choice architecture (OCA) techniques— reference pricing, sensory manipulation, and time-bound elements and their combinations—influences consumer decision-making, specifically whether its presence makes consumers make detrimental choices, and how these influences and outcomes may vary across different markets. The study focused on consumers in the general population and consumers in vulnerable circumstances, particularly individuals facing financial insecurity, and individuals experiencing poor mental health.

The Competition and Markets Authority (CMA, 2022) identified numerous, potentially harmful OCA practices and grouped them into three broad categories—choice information, choice structure, and choice pressure. Our experiment drew from each category: reference pricing choice information, sensory manipulation in choice structure, and scarcity-based choice pressure. We selected these tactics to mirror common marketplace designs and because several have been the focus of CMA investigation. For example, discount claims closely related to reference pricing were central to the CMA's 2019 hotel booking investigation. It has been emphasised by the CMA (2022) that these design tactics are rarely used in isolation online, yet most empirical studies to date examine them one at a time. By testing reference pricing both in isolation and in combination with sensory manipulation and time elements, we estimate incremental and joint effects. To our knowledge, this is among the first online experiments that investigates their interactions in a realistic marketplace setting, enhancing the ecological validity of the evidence.

A key part of this research was determining if vulnerable groups are disproportionately negatively affected by the online choice architecture practices under study. Secondary research questions examined the broader consequences of these online choice practices, including potential harm to market competition. Finally, a tertiary analysis investigated the underlying mechanisms through which reference pricing, sensory manipulation and discount timers operate to influence consumer behaviour.

3 Experiment Overview

Figure 2. Experimental structure overview



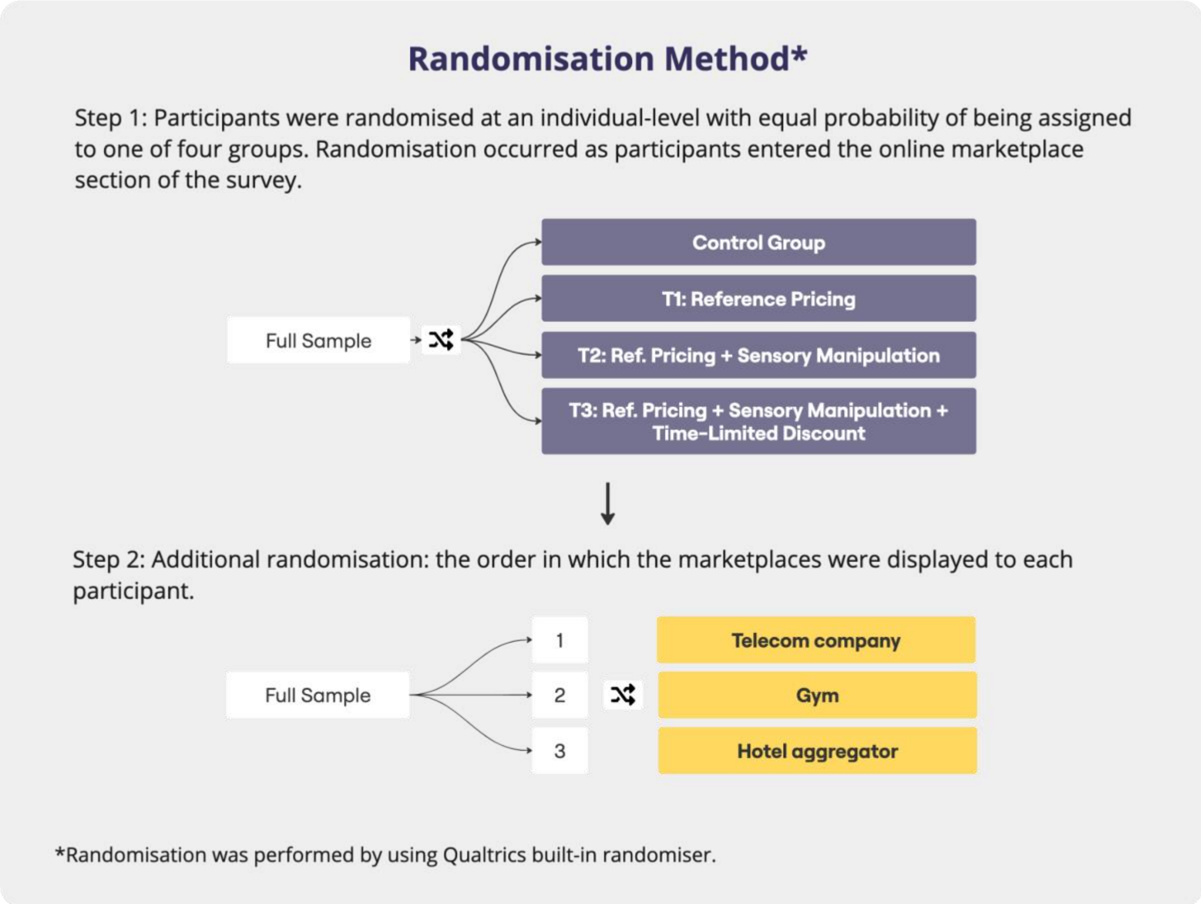
A total of 8,046 participants (UK adults with internet access) took part in the experiment. Of the 8,046 participants, 7,008 were nationally representative of the UK adult population, 1,779 met the criterion of financial insecurity (household income below £15,000), and 4,259 respondents met the criterion of poor mental health (MHI-5 score equal or below 68). The experiment was conducted online and coded using Qualtrics, a software platform that allowed us to create a highly realistic virtual environment to simulate three online marketplaces: a telecommunications service platform, a gym membership platform, and a hotel booking platform.

The online experiment began with a welcome screen detailing eligibility, the right to withdraw, and that compensation depended on completing the survey and passing attention/comprehension checks.

Next, participants were asked to answer demographic questions (age, gender, education, UK region, and local urbanisation rate). This was followed by pre-experimental questions on their finances (income, employment, benefit receipt), mental health (MHI-5 inventory, diagnosis, disability), shopping habits (online shopping frequency, stress levels, bargain hunting), financial literacy (Lusardi and Mitchell's 2008 3 question inventory), and time discounting preferences (2 standard switching decision task questions).

After answering these questions, participants were instructed to complete three shopping tasks, one in each marketplace: selecting a broadband deal, a gym membership, and a hotel stay. Participants were presented with a hypothetical set of needs and were instructed to choose the cheapest option that met every requirement listed. They were also informed that, in addition to the compensation received for their participation in the study, they were given one entry into a £300 prize draw for each marketplace task they answered correctly. Comprehension of these instructions was tested with one comprehension question, followed by a second question if the first was answered incorrectly. Participants who failed both comprehension questions were screened out of the experiment.

Figure 3. Randomisation method



Each participant was then blindly randomly allocated with equal probability to one of four experimental conditions showing slightly different versions of the marketplaces. The order in which the marketplaces were shown to each participant was also randomised in all treatments. Once participants were assigned to an experimental group, they remained in that condition for all three shopping tasks:

- **The control group** was presented with marketplaces without reference pricing, sensory manipulation, or discount timers.
- **The "reference pricing" group (Treatment group 1, or T1)** was presented with marketplaces that featured reference pricing.
- **The "reference pricing group + sensory manipulation" (Treatment group 2, or T2)** was presented with marketplaces that featured reference pricing and sensory manipulation.
- **The "reference pricing group + discount timers" (Treatment group 3, or T3)** was presented with marketplaces that featured reference pricing, sensory manipulation and discount timers.

Marketplaces and experimental conditions were selected after careful analysis, as explained in the next section. In addition, all marketplaces featured a banner that, when clicked, redirected users to a competitor website offering equivalent products/services but did not feature reference pricing, sensory manipulation, or discount timers in any of the groups.

Following the three marketplace tasks, participants were asked to answer one question including three simple geography trivia questions as an attention check.

Finally, a post-experimental survey assessed their marketplace task experience: stress levels, post-task mood, task difficulty per marketplace, decision confidence, value-for-money perception, feelings of deception, noticing reference prices, and the influence of reference prices.

3.1 Marketplaces and Experimental Conditions

Marketplaces and experimental conditions were selected and developed in collaboration with Citizens Advice. In addition, early consultations were held with experts from multiple competition and consumer protection organisations, such as the Competition and Markets Authority (CMA), Ofcom, the Financial Conduct Authority (FCA), the Department for Business and Trade (DBT), and others. Consultations were conducted on various aspects of the experimental design (e.g., population of interest, outcome variables), the types and measurement of harm relevant to OCA practices, and the evidence required from a regulatory and legislative perspective. In this section, we provide more detail on the selection and design of the marketplaces and experimental conditions.

3.1.1. Marketplaces Selection

First, five potential marketplaces were shortlisted to be included in this research. This selection was informed by previous research on OCA practices (Citizens Advice, 2022; Competition & Markets Authority, 2022; Behavioural Insights Team, 2024), their policy relevance and research interest. Telecom companies were deemed the most relevant, followed by gyms, hotel and flight search aggregators, and retailers.

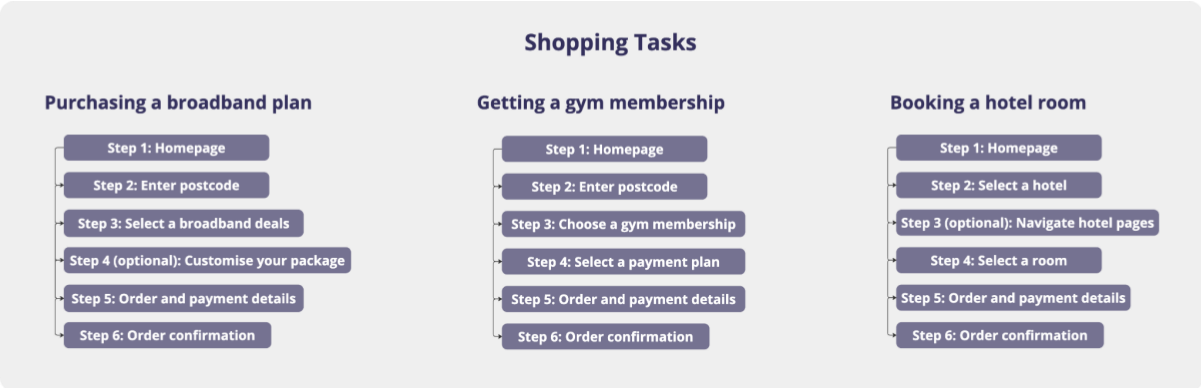
Next, desk research was conducted to collate real-life examples of the OCA practice under study–reference pricing–including any examples where reference pricing was used in combination with other practices (e.g., sensory manipulation, time-bound discounts). We reviewed and collated evidence from a range of marketplaces (e.g., telecoms, gyms, hotel booking sites) and product categories (e.g., phone plans, broadband plans, etc.). This allowed us to explore how reference pricing is deployed in real-life environments, its prevalence among these marketplaces, the complexity of their interfaces and user journeys (e.g., the multiple screens consumers have to navigate when purchasing in these marketplaces), and their suitability and technical feasibility to be replicated in an experimental setting. For example, during the desk research, we found that reference pricing was not as prevalent in flight aggregator sites compared to other practices, such as drip pricing. We also found that phone market user journeys were notably complex, often involving ten or more steps (e.g., one journey featured fourteen webpages requiring some kind of choice from users). In this way, some marketplaces were excluded because of their difficulty in being replicated in an experimental setting as they required an elaborate logic with multiple decision points and detailed navigation paths. Collated evidence can be found in Appendix D.

As a result of a structured comparison between the collated evidence, the criteria outlined above, and the input by Citizens Advice and stakeholders, three shopping scenarios were selected: purchasing a broadband deal, a gym membership, and a hotel stay. In the next section, we provide further details on how each shopping task was designed.

3.1.2. Marketplaces and Shopping Task Design

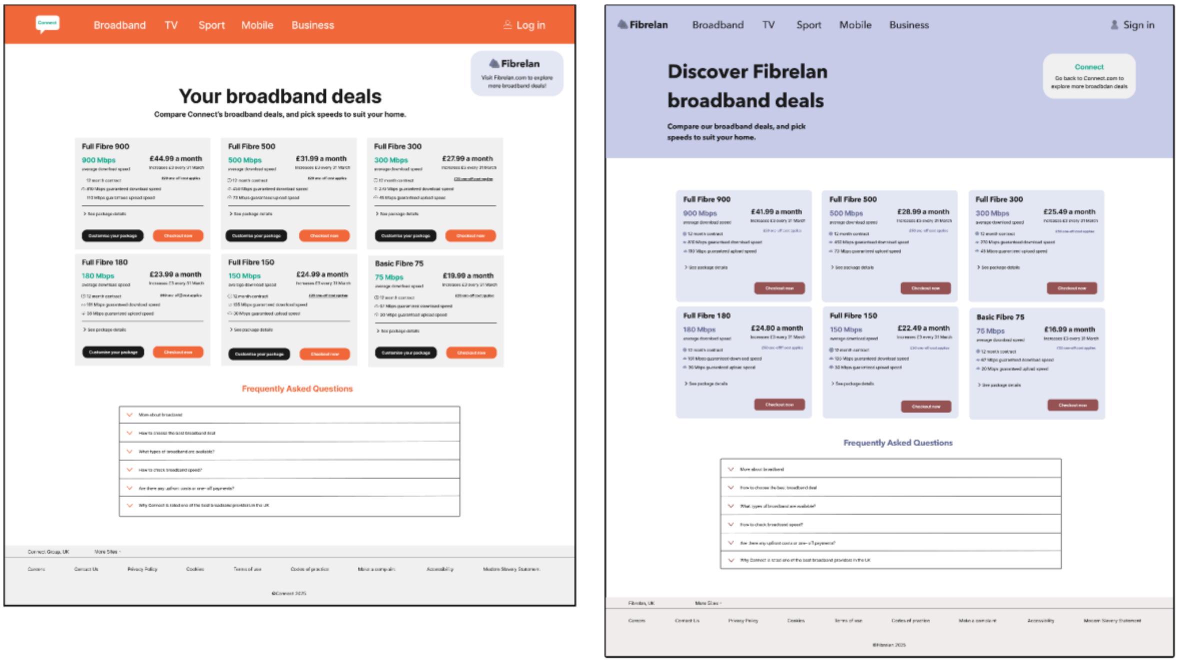
Marketplaces and purchase tasks were designed in such a way that the information that participants saw–and the decisions that they were asked to make–mirrored real-life purchasing decisions and online environments. Immersing participants in highly realistic environments prompts them to unconsciously reveal their 'true' preferences and act as they would in 'the real world'. The three marketplaces had similar user flows, meaning the series of interactions or steps required to complete the shopping task were the same in length and structure, as shown in Figure 4 below. User flows can be found in Appendix D.

Figure 4. Structure of Shopping Tasks



In addition, participants were given the option to continue shopping around and visit alternative marketplaces if they wished to do so (e.g., a different telecom company or travel agency). As shown in Figure 5 below, these alternative marketplaces consisted of an interface that was very similar to the 'main' marketplaces and didn't vary across treatment groups. In this way, we recorded whether participants bought something within the initial marketplace or whether they decided to shop around.

Figure 5. Product choice screens displayed in the experiment for the 'main' (left) and 'alternative' (right) telecom marketplaces



The products shown in each marketplace were designed to mimic real-life examples of products (e.g., a 'Full Fibre 180' broadband plan). Participants were presented with a range of products during the ‘select a product’ step. To complete the shopping task, they had to compare the task requirements against the characteristics of the available products and choose the least expensive option that fulfilled all the criteria outlined in the task instructions. Products offered always included one 'correct' option (the cheapest option that met all task requirements), one 'decoy' option (an option that no rational consumer should choose), and two to four 'incorrect' options (a range of options that could be attractive to consumers but that didn’t respond to the task instructions). Those participants assigned to one of the OCA practice-informed treatment groups were shown OCAs in some of the products, to steer them towards certain options. OCAs were consistently placed in the 'decoy' option and in one or two additional 'incorrect' options, making inferior products more attractive alongside the 'correct' product.

Table 1. Marketplace characteristics

Marketplace	Telecom	Gym	Hotel Booking Site
Type of product	Broadband plans	Gym membership	Hotel stays
Num. of products shown	6	4	6
Variable product characteristics	• Average Download Speed (Mbps) • Guaranteed Speed (Mbps) • Upload Speed • Access to additional perks and benefits	• Membership name • Gym access times • Book classes in advance (7 vs 14 days) • Freeze memberships in advance	• Hotel name • Hotel photographs • Stars • Rating • Price • Distance from the center • Room size

	• Monthly price • One-off cost	• Access to all gym locations • Monthly subscription fee • Monthly joining fee • Annual subscription fee • Annual joining fee	• Type of bed • Breakfast included • Free cancellation
Num. of products with OCAs (only 2 to 4 exp. groups)	3	3	2

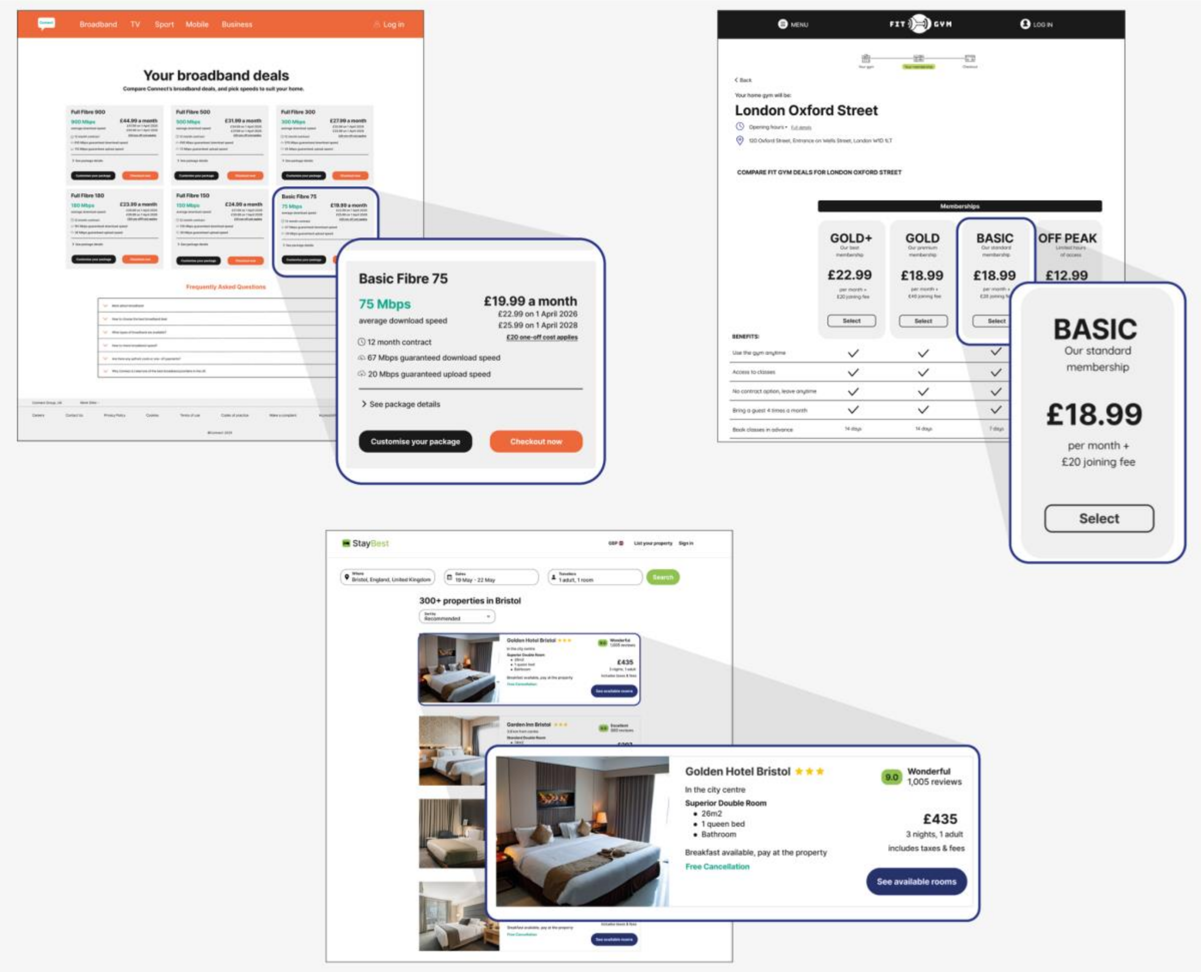
3.1.3. Experimental Conditions

In this study, people were randomly assigned to one of four groups: (1) reference pricing only; (2) reference pricing plus sensory manipulation; (3) reference pricing plus sensory manipulation and a time-bound element; or (4) a control group that saw none of these practices. Some characteristics were constant across conditions (e.g., products offered), and others were tailored to each treatment condition, such as including a reference price or adding visual cues to specific products, as outlined in the previous section. In the next paragraphs, we describe the experimental groups in more detail.

3.1.3.1. Control Condition

Control conditions were designed to represent a typical online shop interface for each of the three tasks: selecting a broadband deal, a gym membership, and a hotel stay. These conditions, to which no intervention was applied, served as the comparison group for the evaluation of the experimental conditions.

Figure 6. Control conditions for the three marketplaces: telecom, gym and hotel booking site

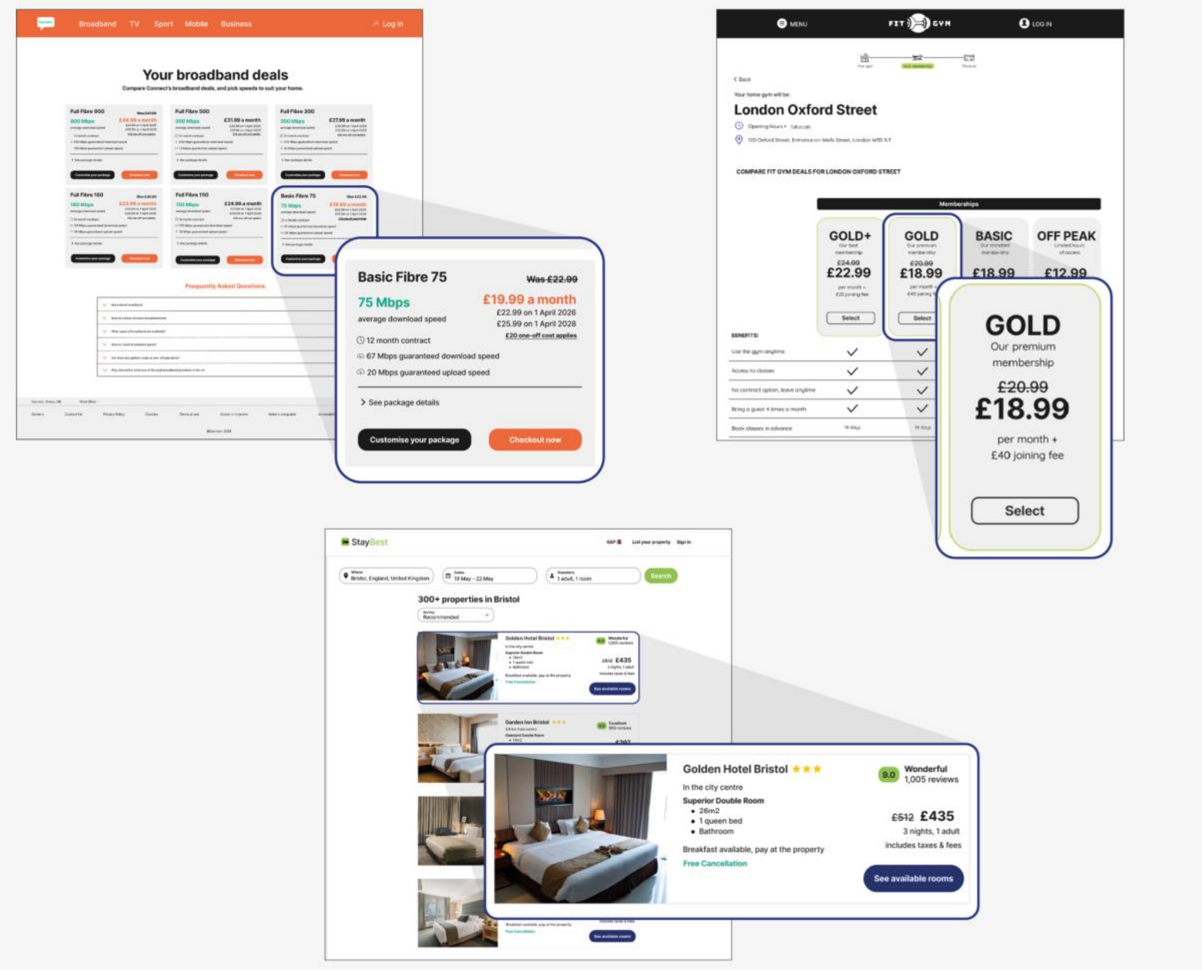


3.1.3.2. Reference Pricing (Treatment 1)

The reference pricing conditions were largely equivalent to the control conditions, except for the introduction of reference pricing in the product-selection screen. This page featured reference pricing, involving the display of a previous price alongside the current price to make the current price look more attractive as shown below.

The broadband and gym marketplaces offered subscription-based services with a pricing structure consisting of two components: an upfront cost and a monthly subscription fee. This pricing model can be classified as a type of OCA known as partitioned pricing. In partitioned pricing, individual price components are presented separately without disclosing the total or estimated overall cost to consumers. This design choice may have influenced participants' decisions in addition to reference pricing. However, we made this choice for two main reasons: first, to reflect real-life scenarios where OCAs rarely exist in isolation, and second, to help us define superior and inferior choices without making it too easy for participants in the experiment to identify them.

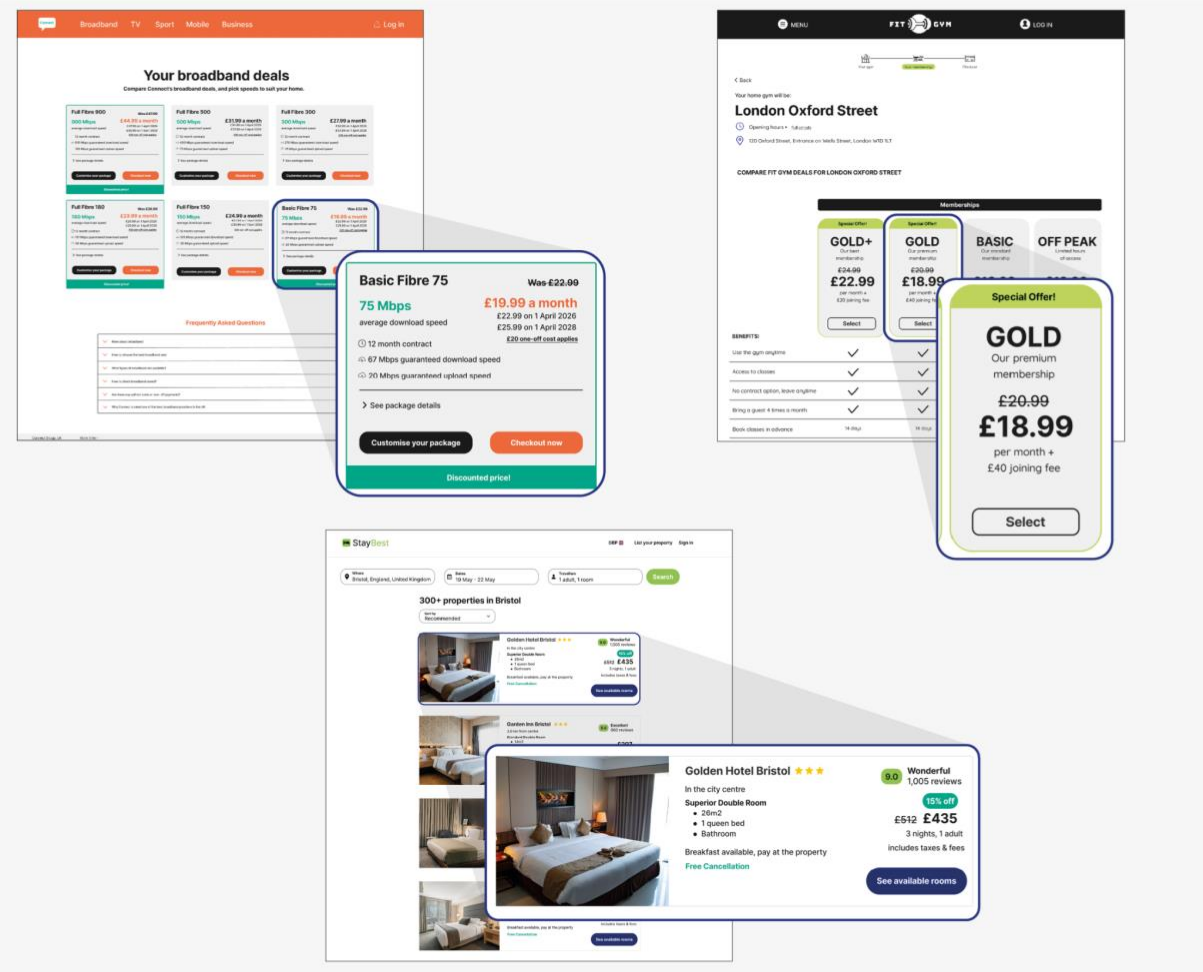
Figure 7. Reference pricing conditions (T1) for the three marketplaces: telecom, gym and hotel booking site



3.1.3.3. Reference Pricing and Sensory Manipulation (Treatment 2)

The reference pricing and sensory manipulation conditions were identical to the previous conditions, except for the addition of visual features that aimed to steer consumers towards the options that included reference pricing.

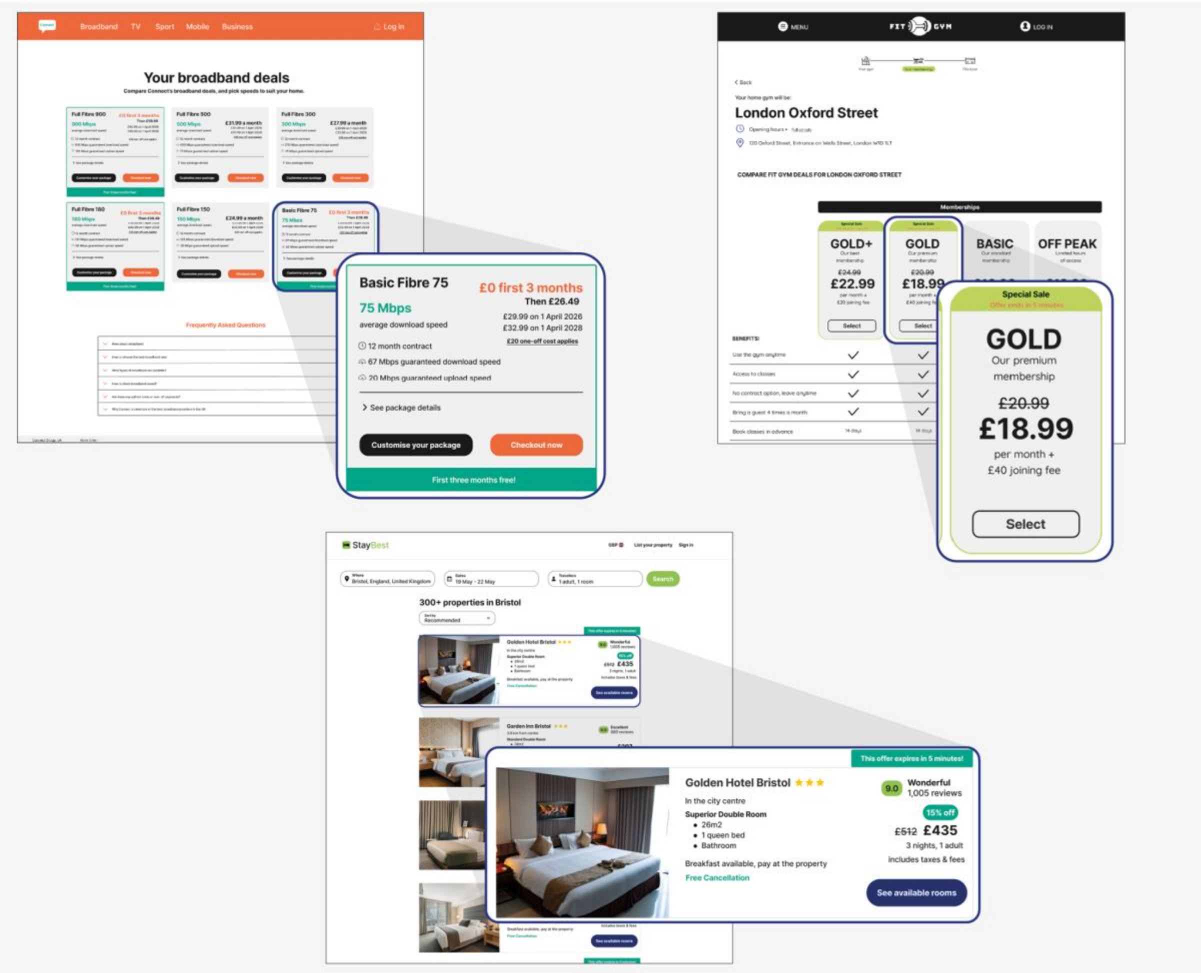
Figure 8. Reference pricing and sensory manipulation conditions (T2) for the three marketplaces: telecom, gym and hotel booking site



3.1.3.4. Reference Pricing, Sensory Manipulation and Time-Bound Elements (Treatment 3)

The reference pricing, sensory manipulation, and time-bound elements conditions were similar to the previous conditions, with the addition of a time-bound element. This element was developed based on desk research, where we gathered examples of how these practices are implemented in real life. For instance, broadband providers offer a time-limited discount that applies for a certain duration within the contract, while gym memberships and hotels utilise countdown timers to indicate that a specific price is available to consumers for a limited time only. Overall, we designed these strategies to create a sense of urgency and pressure in consumers' decision-making.

Figure 9. Reference pricing, sensory manipulation and time-bound discounts conditions (T3) for the three marketplaces: telecom, gym and hotel booking site



3.2 Sampling Approach

3.2.1. Participants were recruited through the panel provider Dynata.

- **Sample:** 8,000 completing respondents, including at least 1,000 UK adults who met fixed financial insecurity criterion (household annual income of less than £15,000) and at least 1,000 UK adults who met characteristics of poor mental health (MHI-5 score of less than or equal to 68).
- **Recruitment:** All participants were recruited using the panel provider Dynata. Respondents received approximately £0.75, which could then be redeemed in the form of gifts, vouchers, etc. Responses from the 8,000 participants were collected by Dynata throughout the course of six weeks. Participants were also entered into a prize draw for £300 for each marketplace task they answered correctly.

- **Screening criteria:** All participants had to be 18+, English-speaking, and live in the United Kingdom. The sample was representative in terms of age, and gender. Participants were excluded if any of the following was true:
    - They did not finish the survey/drop out;
    - They failed both comprehension checks (asking about their understanding of the instructions);
    - They entered an age of less than 18 years.

3.2.2. Measuring Vulnerability

3.2.2.1. Mental Health

Respondents' mental health was measured with the five-item Mental Health Inventory (MHI-5), a brief, internationally validated scale of psychological well-being and distress (EORG, 2003; Rivera-Riquelme et al., 2019). The instrument contains two positively worded and three negatively worded items, each scored on a six-point Likert scale. After reversing the score of positive items, responses are summed and rescaled to a 0–100 index, where higher values denote better mental health. The full wording of the inventory is provided in Appendix A1. To identify mental health vulnerabilities, we used two cutoff scores. Our primary cutoff was a score of 68 or below, which has been validated for screening mood disorders in the general population (Ten Have et al., 2024). We additionally used a cutoff of 52, which is a widely adopted threshold for screening for severe depressive symptoms or major depression, as supported by research by Rugulies et al. (2006) in a Danish work cohort, Holmes (1998) in a clinical sample, and its use in European population health surveys (EORG, 2003).

3.2.2.2. Financial Insecurity

Financial insecurity was assessed via self-reported gross annual household income. Households reporting income of less than £15,000 per year (before tax and including benefits) were classified as financially insecure. According to the Financial Conduct Authority (2024a, 2024b), being in this low-income group is a primary risk factor for low financial resilience—a status the FCA assigns to adults who:

- have missed payments on domestic bills or credit commitments in three or more of the previous six months,
- feel that keeping up with those commitments is a heavy burden, and/or
- hold such limited savings that they could not cover basic living costs for a week if their main income stopped.

In the 2024 Financial Lives survey, 48% of adults in <£15,000 households met this low-resilience definition, twice the national average of 24 %.

3.3 Data Collection

3.3.1. Primary Outcomes

Our primary outcomes examined whether participants made the optimal choice in each shopping task. For each of the three marketplace tasks (broadband, gym, hotel), we defined a binary indicator for an optimal, objectively correct choice, equal to 1 if the participant selected the cheapest product that met all the predefined needs in that marketplace task (and 0 otherwise). We also constructed an overall measure of performance: the proportion of the three tasks in which the participant chose correctly (taking values of 0, 1/3, 2/3, or 1). In addition, we defined a "decoy choice" indicator for each marketplace task, equal to 1 if the participant chose a specific overpriced decoy option (and 0 otherwise), along with the proportion of tasks in which they chose a decoy. Each of these primary outcomes allowed us to assess whether the presence of reference pricing, sensory manipulation, and time-bound elements led to worse decision outcomes—i.e., fewer correct choices and greater selection of decoy options—relative to the control condition.

3.3.2. Secondary Outcomes

Our secondary outcomes examined participants' propensity to "shop around," reflecting potential impacts on market competition. In each task, participants had the option to click a banner to visit an alternative, competitor website offering similar products without the manipulative design features. We defined a binary variable for visiting the competitor site in each marketplace task (1 if the participant clicked the banner to the alternative site, 0 if they stayed only on the main site). We also computed the proportion of tasks with a competitor visit for each participant (values 0, 1/3, 2/3, or 1). A complementary secondary outcome measured completed purchases on the competitor's site. For each task, we recorded a binary indicator equal to 1 if the participant not only visited the alternative site but also selected a product and finished the purchase process there (and 0 if the final choice was made on the main site). However, given that both the main and the competitor sites featured nearly identical offers, the clicks to visit the competitor site remained our main secondary outcome. Taken together, we assessed whether reference pricing and related online choice architecture elements deterred consumers from exploring alternative shopping options, which draws on the broader market competition effects of online choice architecture.

3.3.3. Data Validity Checks

All analyses used the final sample of respondents who completed the survey and passed the pre-specified screening criteria (18+ age, passing at least 1 of 2 comprehension questions regarding the experimental instructions). We obtained 8,046 complete responses - slightly above the target of 8,000. Of those, 4,259 respondents met the poor mental health criterion (MHI-5 score ≤ 68) and 1,779 met the financial insecurity criterion (household income ≤ £15,000). We verified that random assignment achieved balance across the four experimental groups by comparing baseline demographics (age, gender, education, region, and urbanization) between conditions, including within the vulnerable sub-samples. We also conducted an attrition analysis to assess potential biases: examining at which points participants dropped out, whether certain types of participants were more likely to drop out, and whether dropout rates differed systematically by treatment group. These data validity checks ensured that the randomization was successful and that any attrition was unlikely to compromise the internal validity or representativeness of our results. A detailed evaluation of the internal validity, and external validity (i.e., generalizability of our results using the SANS framework) of our experiment can be found in Appendix C.

3.3.4. Analysis Approach

Our analysis approach followed the pre-registered plan. The main specification was a linear probability model (LPM) applied to the pooled task-level data. In the primary analysis, we regressed a binary indicator for correct/decoy choice in a given task on indicators for each treatment group (T1, T2, T3, with the control group as the reference category), while controlling for marketplace effects and task order. In practice, this means we included dummy variables for the marketplace type (gym and hotel tasks, with broadband as the omitted category) and for the position of the task in the sequence (second task and third task indicators, with first task as baseline). Standard errors were clustered at the individual level to account for having three observations per participant. This LPM setup produced coefficient estimates for each treatment condition's effect on the probability of a correct (and on the probability of decoy) choice. We tested the key hypotheses that these coefficients were zero (i.e., no overall treatment effect and no effect of each individual treatment relative to control) and that the incremental effects between treatments were zero (e.g. adding sensory manipulation in T2 vs T1, or adding time-bound element in T3 vs T2). Rejection of these null hypotheses would indicate that the online choice architecture interventions significantly influence the likelihood of making the correct and the decoy choice.

We also pre-specified a range of heterogeneity analyses and robustness checks. For robustness, we re-ran the main LPM on restricted samples (e.g. dropping participants with abnormally fast completion times, or those who failed the first comprehension check although passed a second one) and also estimated a logistic regression for comparison, to ensure that our findings were not sensitive to model specification. We estimated market-specific regressions for each of the three marketplaces to see if treatment effects varied by context. And, most importantly, we explored heterogeneity by interacting the treatment indicators with indicators for vulnerability (financial insecurity and poor mental health) and other respondent characteristics (such as frequent online shopping, high reported stress in online shopping, bargain hunting habits, full financial literacy score, and present-biased time preferences). This allowed us to detect if any treatment effects were significantly larger or smaller for these subgroups of interest. Finally, similar LPM specifications were applied for the secondary outcomes (banner click and competitor purchases). All analyses and hypotheses were defined in our pre-analysis plan, and the study was pre-registered prior to data collection (AEA Social Science Registry, Trial ID 16094).

4 Results

4.1 Summary

Summary of results

→ **Reference pricing significantly harms consumer decision quality.** On average, exposure to any reference pricing intervention made participants about 5% (-1.7 percentage points) less likely to pick the correct choice, and about 14% (+2.0 percentage points) more likely to select the decoy.

→ **There were differences in how the OCA practices affected participants' choices:**
  - Reference pricing on its own (T1), and reference pricing with sensory manipulation (T2) each lowered correct choices and raised decoy choices, with larger effects in T2.
  - The effect of reference pricing, sensory manipulation and time-bound elements (T3) was inconclusive, varying by timer implementation. When the timer was a countdown timer in the gym market, T3 reduced accuracy more than T1 or even T2. But when the timer was a time-limited discount in the broadband market, T3 actually increased accuracy even relative to the control group.

→ **Participants in all experimental conditions (including the control group) found it difficult to correctly choose the right broadband plan.** In the broadband task, respondents had a 90% error rate in the control group, suggesting that buying broadband is either very difficult even without reference pricing, sensory manipulation and time-bound elements, or that the experimental task was particularly hard.

→ **Vulnerable consumers** (those experiencing financial insecurity or poor mental health) **were about 15.5 - 24.9% less likely than the general population to choose the correct product** (a gap of 5.0 - 7.9 percentage points), but their choices were not disproportionately affected by reference pricing.

→ **The evidence suggests that participants experiencing poor mental health spent longer choosing a broadband plan** in the condition with reference pricing, sensory manipulation and time-bound elements (T3) than others.

→ **There was less shopping around in the reference pricing condition (T1) than in the control.** Participants were less likely to explore alternative marketplaces in T1, suggesting that OCAs can have effects on market competition.

→ **Participants experiencing poor mental health and those with high shopping stress levels were significantly less likely to shop around.** Conversely, bargain-hunters and those with high financial literacy were significantly more likely to visit other marketplaces.

→ **Reference pricing made consumers overspend, not underspend.** When reference pricing was present, people spent more on average across the marketplaces, especially in broadband.

→ **In general, there is an inverse relationship between decision quality and decision confidence.** When reference pricing was present, people perceived the shopping tasks as easier and felt more confident in their decisions, despite making objectively worse choices.

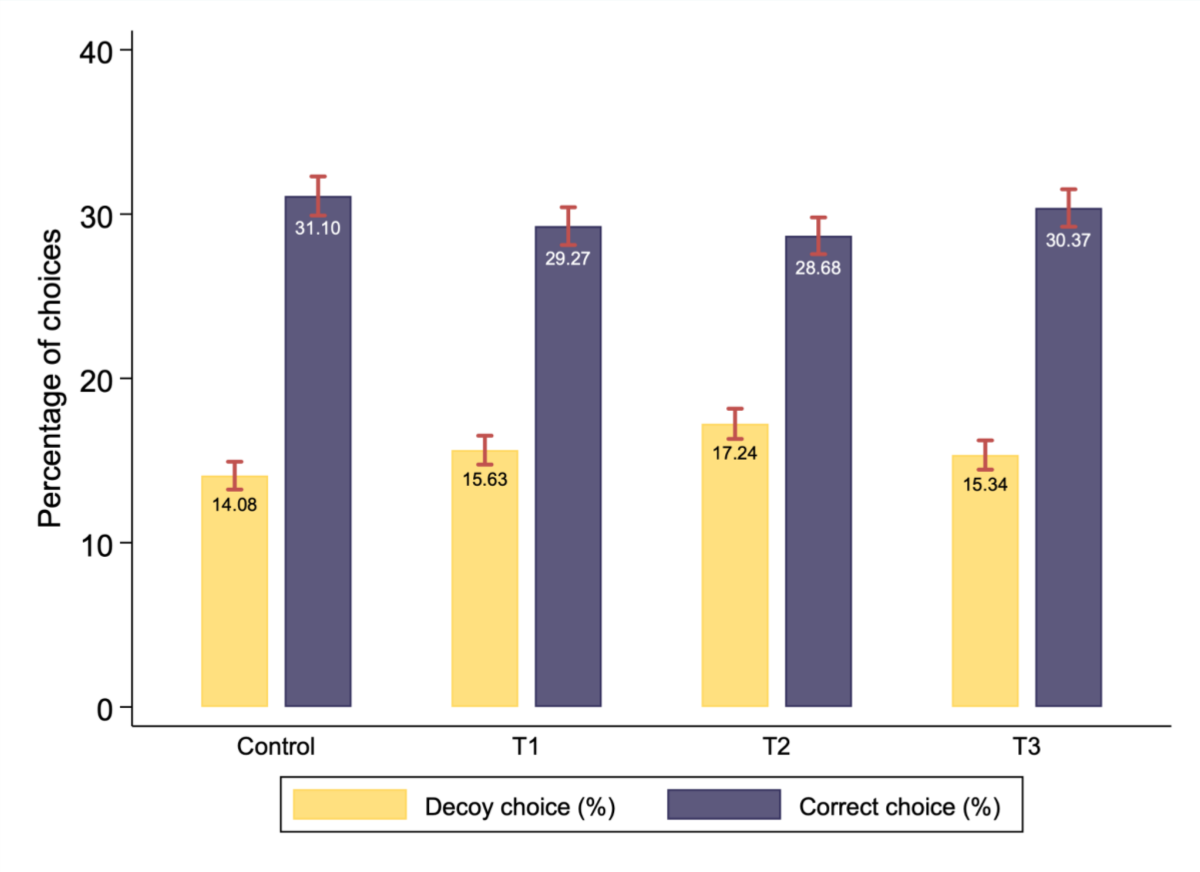
4.2 Primary Outcomes

4.2.1. What is the effect of online choice architecture on customer choices?

We began by examining how the three experimental conditions influenced the quality of choices that participants made across the three online marketplaces (broadband, gym and hotel booking). Figure 10 displays the average effects of treatment (T1, T2 and T3) on selecting the correct and the decoy option. Detailed coefficients and standard errors are reported in Table B1 of the Appendix. Each treatment increased participants' susceptibility to the decoy option, while only the pure reference pricing (T1) and the reference pricing made salient with sensory manipulation (T2) reduced the likelihood of choosing correctly. More specifically:

- T1 lowered the probability of a correct choice by about 6 % (i.e., -1.8 p.p., $p < 0.05$) and simultaneously raised decoy selection by approximately 11 % (i.e., +1.6 p.p., $p < 0.05$).
- T2 had a larger impact in both directions: correct choices fell by 8 % (-2.4 p.p., $p < 0.01$), while decoy choices increased by approximately 23 % (+3.2 p.p., $p < 0.01$) relative to the control condition. A Wald test confirmed that T2's effect on decoy choice was significantly larger than T1's ($p < 0.05$), indicating that sensory manipulation amplified the negative effect of reference pricing.
- T3 did not significantly change the overall probability of making a correct choice relative to the control group, but it increased decoy choices by 9 % (+1.3 p.p., $p < 0.05$).

Figure 10. Treatment effects on purchase decisions across marketplaces.



Notes: The bar chart shows the mean percentage of correct choices and decoy choices for each of the experimental groups (control, T1, T2, T3). Means are labelled on the bars and red whiskers represent the 95% confidence intervals. The pooled values are computed per respondent as average across the three (broadband, gym, hotel) tasks and then averaged within each experimental group. N = 24,138

These findings also held in logistic specifications (Table B2 in the Appendix) and remained across three different robustness checks: (1) completion-time filter, that drops the slowest 5% of respondents, (2) comprehension filter, that keeps only participants who passed the first comprehension question quizzing the understanding of instructions and, finally, (3) attention filter, that keeps only participants who failed at most one attention check. Across all three restricted samples, the signs, magnitudes and significance levels of the T1, T2 and T3 coefficients remain effectively unchanged (full results in Table B3 and Table B4).

To establish the average treatment effect of any reference pricing intervention, we further combined the three experimental conditions (T1, T2 and T3) into a single treatment indicator and re-estimated the primary linear probability model. The pooled specification indicates that exposure to any reference pricing intervention reduced correct choices by 5 % (-1.7 p.p., $p<0.05$) and increased the uptake of decoy choices by 14 % (+2.0 p.p., $p< 0.01$) relative to the control group. Therefore, regardless of additional OCA elements layered onto the reference pricing, we see a small but systematic effect of reference pricing steering participants away from correct choices and toward the dominated, decoy alternatives.

4.2.2. Do the effects of online choice architecture differ by vulnerability status and consumer traits?

Next, we investigated whether vulnerable customers were disproportionately affected by reference pricing. To do so, we extended our regression models to include two indicators of vulnerability status: (1) financial insecurity (household income below £15,000) and (2) poor mental health (MHI-5 ≤ 68), plus their interactions with the treatment indicators (T1, T2, T3). This allowed direct comparison of treatment effects for vulnerable versus non-vulnerable participants while holding task and order effects constant. As evident in Figure 1, we found that vulnerable participants start from a lower baseline rate of correct choice. Detailed coefficients and standard errors are reported in Table B5 of the Appendix. Overall, financially insecure participants were 7.9 p.p. less likely to choose correctly in the control group ($p<0.01$), and those with poor mental health were 5.0 p.p. less likely ($p<0.01$).

In this specification, any disproportionate impact of reference pricing on vulnerable participants would show up as statistically significant coefficients on the interaction terms between the treatment dummies (T1, T2, T3) and the vulnerability indicators—that is, as a differential treatment effect for vulnerable versus non-vulnerable participants over and above the main effects. We did not find consistent evidence of such disproportionate harm: the coefficients for the T1, T2 and T3 dummies closely mirrored the main results, and the corresponding interaction terms were generally small and statistically insignificant. If anything, the combination of reference pricing, sensory manipulation and time-bound elements (T3) seemed to offset some of the disadvantage among the financially insecure group in selecting the correct choice by 4.5 p.p. ($p<0.05$).

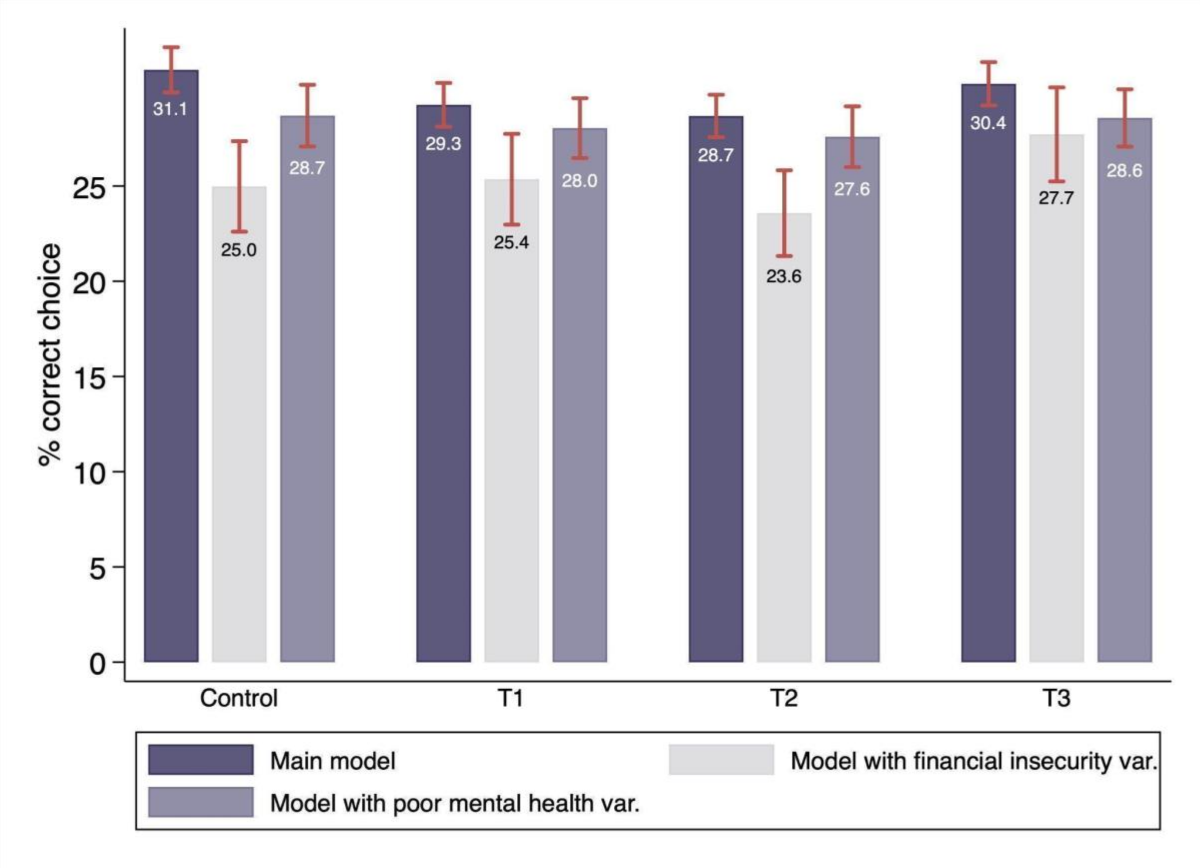
Crucially, the lower baseline rate of correct choices among the vulnerable groups did not translate into a greater tendency to pick the decoy. In the decoy choice models (Appendix Table B6), the main effects of financial insecurity and poor mental health were near zero and the treatment × vulnerability interactions were generally insignificant. In fact, T3 was associated with 2.6 p.p. lower decoy uptake among the financially insecure participants and 2.3 p.p. lower for those with poor mental health in the control condition, although both of these were only marginally significant ($p <0.10$). Overall, reference pricing lowered decision quality on average but not differentially for vulnerable groups. Vulnerability was mainly associated with lower rate of correct choices when shopping online rather than a systematically larger effect of reference pricing, and we did not find treatment x vulnerability interactions for either correct or decoy choice.

These results were robust to (1) combining T1, T2 and T3 into a single "any treatment" indicator,[1] to (2) combining financial insecurity and poor mental health into an "any vulnerability" indicator,[1] and to (3) six alternative vulnerability definitions. These specifications yielded similar conclusions with interaction terms still insignificant at the conventional 5 percent level. Full specifications and estimates are reported in Tables B7 - B10 in the Appendix.

---

[1] This analysis should be considered exploratory as it was not preregistered.

Figure 11. Differential effects of reference pricing on correct choices by vulnerability status



Notes: Bars show predicted probabilities of choosing the correct option for each condition (control, T1, T2, T3), with three bars per condition: (1) main model, (2) main model additionally with financial insecurity indicator and its interaction with T1, T2, T3, (3) main model additionally with poor mental health indicator and its interaction with T1, T2, T3. Predictions come from pooled linear probability models with treatment indicators, marketplace dummies (gym, hotel; broadband omitted) and task-order dummies (order2, order3). FI indicates household income < £15,000; MH indicates MHI-5 ≤ 68. Red whiskers are 95% CIs from delta-method estimates with SEs clustered by participant.

Beyond vulnerability, several consumer traits were associated with differential propensities to choose correctly (and to select decoys) as well as differing susceptibility to online choice architecture interventions (i.e., reference pricing, sensory manipulation, and time-bound elements). Self-identified bargain hunters chose the correct option 7.1 p.p. more often than those who do not frequently look for deals and discounts in the control condition ($p<0.01$), while the online choice architecture interventions were disproportionately harmful for bargain hunters. Correct choice among bargain seekers fell by an additional 3.6 p.p. in T1 ($p<0.05$), by an additional 4.0 p.p. in T2 ($p<0.05$), and by an additional 2.9 p.p. in T3 ($p<0.10$), narrowing their baseline advantage. These "additional" drops in the share that made a correct choice among bargain hunters were interaction effects: they were estimated over and above the pooled main effects of T1 - T3, suggesting that reference pricing was particularly effective at diverting deal-seekers away from optimal products and services. Likewise, high financial literacy[2] was associated with a much higher baseline rate of correct choices. Optimal selection was 13.3 p.p. higher for highly financially literate consumers in the control condition ($p<0.01$), although this advantage was again partially offset when reference pricing was combined with sensory manipulation and a time-bound element. The T3 × financial literacy interaction reduced correctness by an additional 4.9 p.p. ($p<0.01$), suggesting that complex online choice architecture can counter the informational advantages of highly numerate consumers.

Turning to other traits, participants reporting high shopping stress picked the correct choice 6.7 p.p. less often in the control condition ($p<0.05$), and present-biased participants[3] chose correctly 2.9 p.p. less often at baseline ($p<0.05$). For both of these traits, interactions between traits and treatments (T1 - T3) were not statistically significant. When it comes to selecting decoys, the only systematic difference is for present-biased participants in the subtle reference pricing condition (T1), who exhibited 3.5 p.p. higher decoy uptake ($p<0.05$). Otherwise, decoy susceptibility did not differ at the 5% level across bargain hunting, shopping frequency, shopping stress, or financial literacy, indicating that these traits primarily shifted the probability of selecting the correct option rather than pulling participants toward decoys.

4.2.3. Are the effects of online choice architecture market-specific?

In addition to varied rates of correct choice between vulnerable and non-vulnerable groups, there were also vast differences in choices across marketplace tasks. In the absence of reference pricing, only about 1 in 10 participants (11.2 %) selected the correct broadband plan, whereas nearly 1 in 2 participants (i.e., 46.4 %) selected the correct hotel. For context, the rate of correct choices in the gym marketplace lies between the rates for hotels and broadband, with roughly one-third (34.4%) of control participants choosing correctly. However, the gym marketplace only had four choices available, whereas both the broadband and hotel aggregator website presented a six-option menu. This pattern suggests that people may simply be poor at evaluating broadband tariffs, perhaps because these plans are inherently complex or rarely chosen, whereas the hotel task may have been intrinsically easier, or participants were more familiar and comfortable with hotel aggregator interfaces. The main specification, including marketplace coefficients, is in Table B1 of the appendix.

Also looking at marketplace-specific linear-probability models (Table B11 and Table B12), it seems that the reference pricing interventions did not work uniformly across marketplaces. Their impact varied by market and might have also depended on the implementation of online choice architecture. Notably, the timers in T3 were implemented differently across the marketplaces. For broadband, they took the form of “first 3 months free”, whereas for the gym and hotel marketplaces, they were “5 minutes left on this offer” countdown clocks.

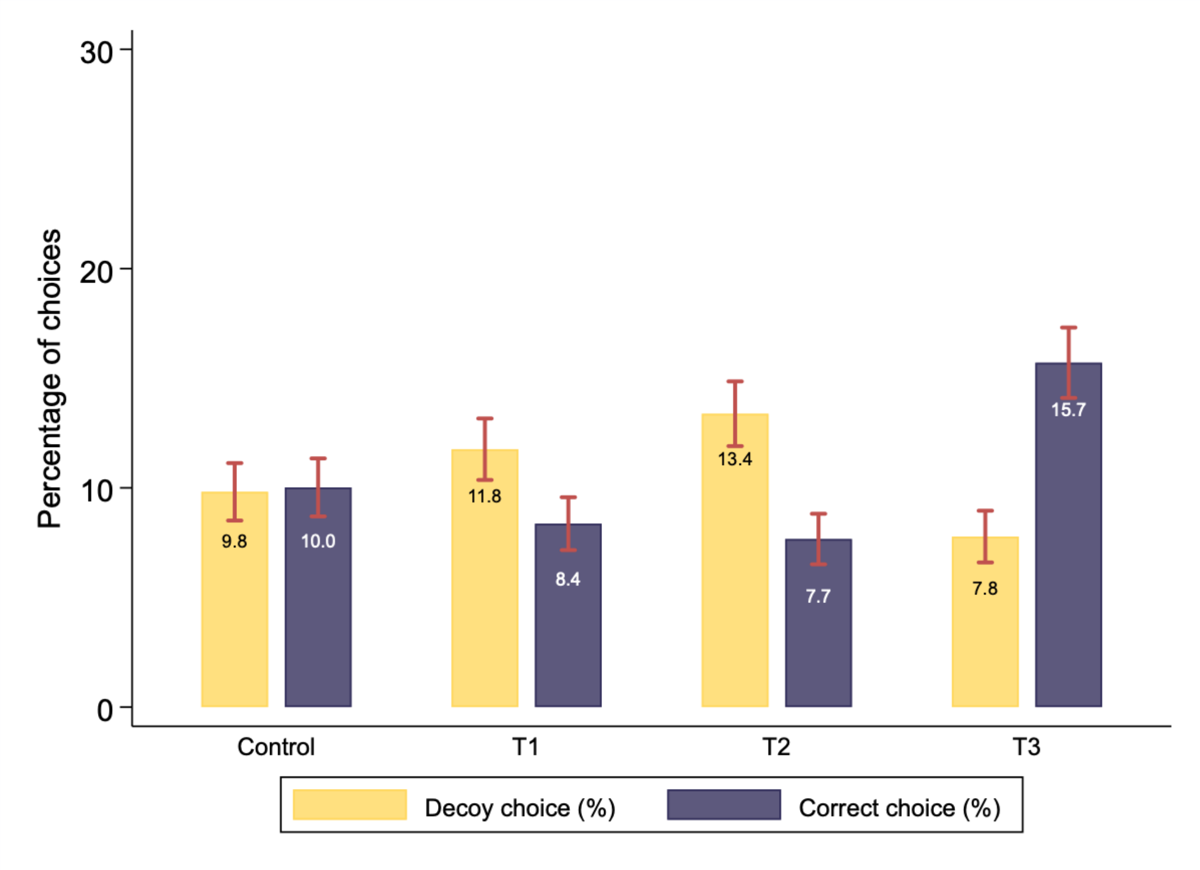
[2] High financial literacy was defined as answering all three standardised financial literacy questions as defined by the Financial Conduct Authority (FCA).

[3] Present bias was assessed using hypothetical scenarios, asking participants whether they preferred receiving a smaller amount of money sooner versus a larger amount later (e.g., £100 now vs. £110 in 1 week; £100 in 4 weeks vs. £110 in 5 weeks). A participant was classified as present biased if they prefer receiving £100 now, rather than £110 in 1 week but also receiving £110 in 5 weeks, rather than £100 in 4 weeks.

4.2.3.1. Broadband

The time-limited discount in T3, framed as “first 3 months free [on this broadband plan]” actually significantly increased the probability of a correct choice by 5.7 p.p. ($p < 0.01$) and cut decoy uptake by 2.0 p.p. ($p < 0.05$) relative to the control group. This suggests that a time-limited discount, placed on a decoy choice, might actually steer customers away from the decoys and in the direction of correct choices. In contrast, plain reference pricing (T1) and overt reference pricing (T2) in the broadband market both significantly reduced correct choices and increased the uptake of decoys. This effect was particularly pronounced in T2, where accuracy was 2.4 p.p. lower ($p < 0.01$) and decoy selection rate 3.6 p.p. higher ($p < 0.01$), relative to the control group.

Figure 12. Treatment effects on choice selection in the broadband marketplace.

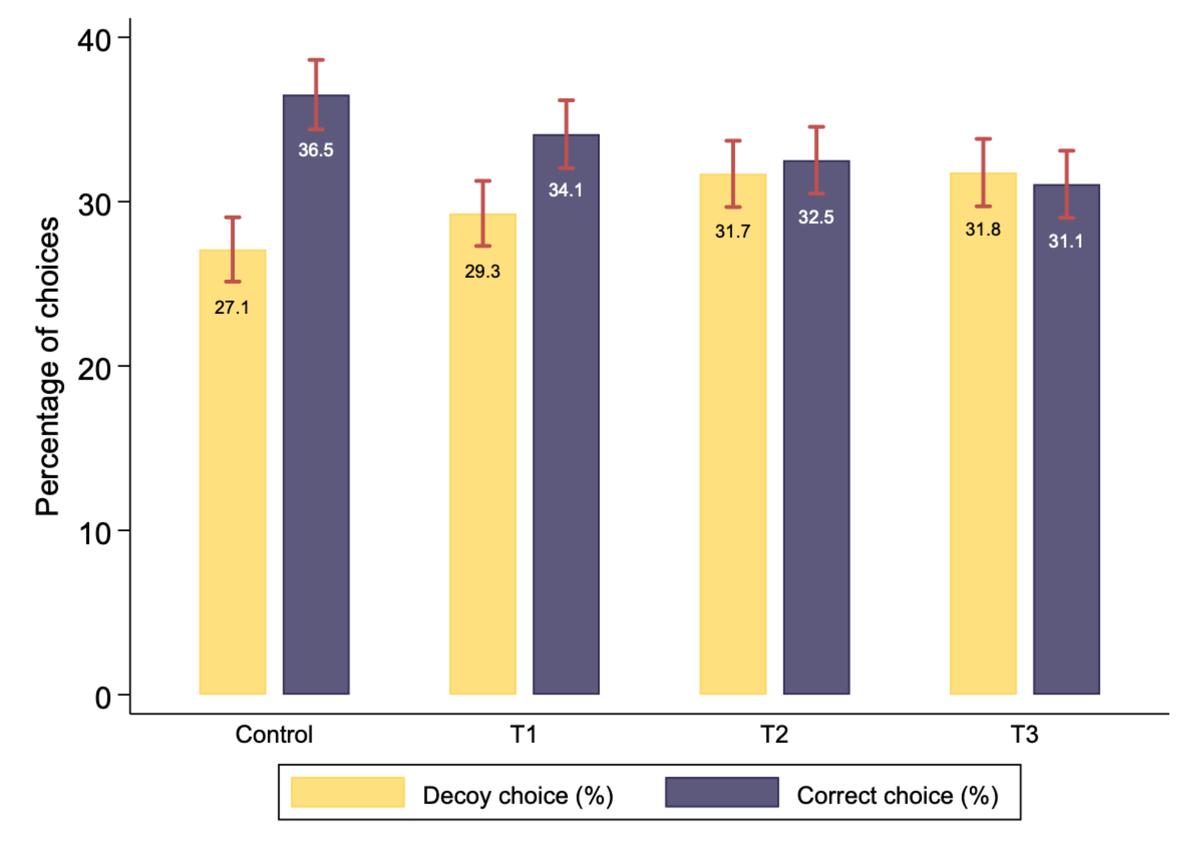


Notes: Percentage of correct and decoy choices by treatment group in the broadband marketplace. The bar chart shows the mean percentage of correct and decoy choices for each of the experimental groups (control, T1, T2, T3). Means are labelled on the bars and red whiskers show the 95% confidence intervals. N = 8,046.

4.2.3.2. Gym

Unlike broadband, every treatment condition in the gym marketplace steered participants away from correct choices and towards decoy choices. There were also clear incremental effects of sensory manipulation in T2, and of countdown timers in T3. The plain reference pricing treatment (T1) lowered the rate of correct choices by 2.4 p.p. and raised decoy uptake by 2.2 p.p., though neither effect was statistically significant at conventional levels. When the reference price was made visually salient (T2) the effect approximately doubled. The rate of correct choices fell by 4.0 p.p. ($p < 0.01$), and the share of decoy choices increased by 4.6 p.p. ($p < 0.01$), relative to the control group. Finally, adding a 5-minute countdown clock to the salient reference pricing (T3) reduced accuracy by 5.4 p.p. ($p < 0.01$) and increased decoy uptake by 4.7 p.p. ($p < 0.01$), relative to the control. Thus, the reference-price implementation in the gym market, especially when combined with sensory manipulation and time pressure, proved the most "effective" in steering consumers toward inferior options.

Figure 13. Treatment effects on choice selection in the gym marketplace.

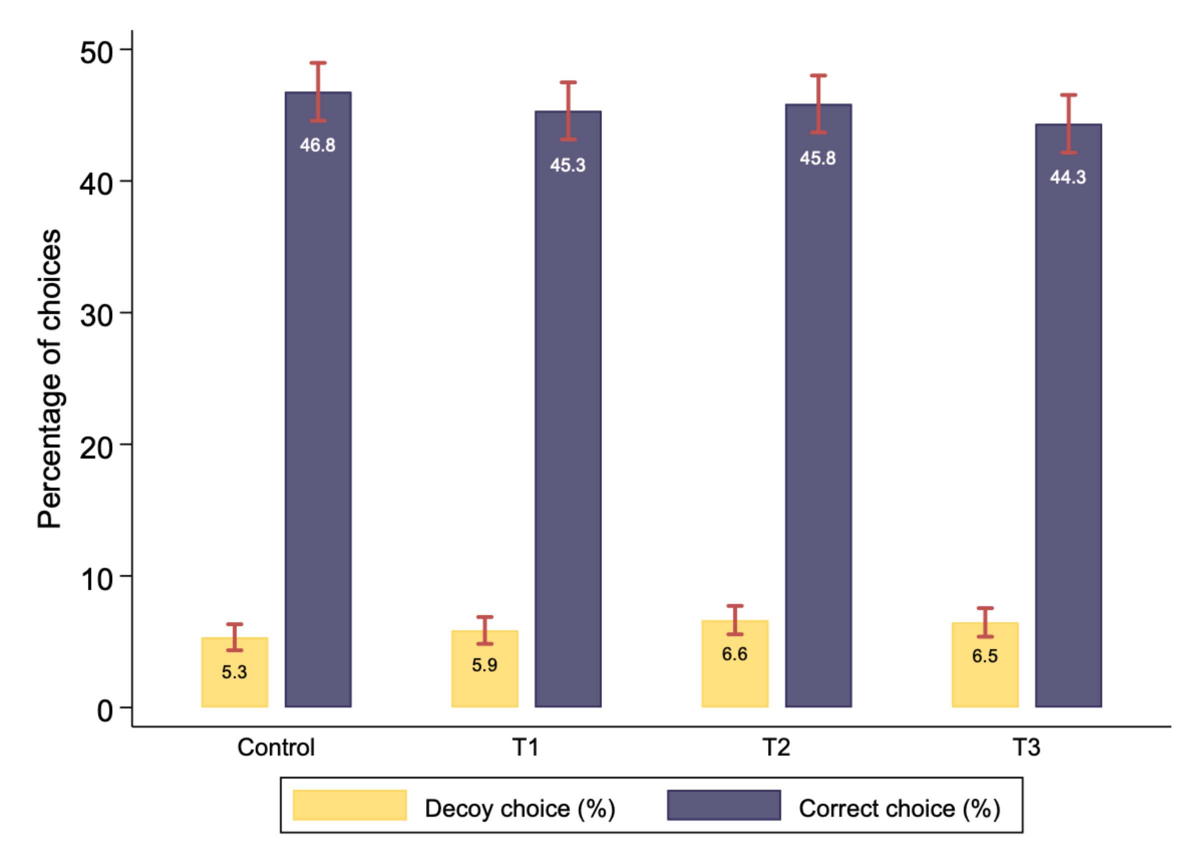


Notes: Percentage of correct and decoy choices by treatment group in the gym marketplace. The bar chart shows the mean percentage of correct and decoy choices for each of the experimental groups (control, T1, T2, T3). Means are labelled on the bars and red whiskers show the 95% confidence intervals. N = 8,046.

4.2.3.3. Hotel

The hotel marketplace exhibited a different pattern. From a baseline accuracy of 46.8 percent in the control group, none of the three treatments had a statistically significant effect on selecting the correct choice. Rate of decoy choices was similarly stable; only the reference pricing with sensory manipulation (T2) produced a modest, marginally significant increase of 1.3 p.p. in decoy selection ($p < 0.1$). The same five-minute countdown clock as in the gym marketplace, failed to move behaviour in the hotels. Factors that might underpin these null findings include (1) the hotel task might have been intrinsically easier than that of selecting a gym membership, with more transparent and familiar price vs. quality trade-offs and (2) participants might have been used to and desensitised to online choice architecture elements on hotel booking platforms (e.g. reference prices, salient elements, countdowns).

Figure 14. Treatment effects on choice selection in the hotel booking marketplace.



Notes: Percentage of correct and decoy choices by treatment group in the hotel marketplace. The bar chart shows the mean percentage of correct and decoy choices for each of the experimental groups (control, T1, T2, T3). Means are labelled on the bars and red whiskers show the 95% confidence intervals. N = 8,046.

In sum, our findings suggest that reference pricing systematically diverted consumers away from correct choices and toward dominated alternatives. The reference pricing effect also seemed to be made stronger with the addition of sensory manipulation (T2). The magnitude of the effects, however, was market-specific—most pronounced for gym memberships, weaker in broadband, and statistically insignificant for hotel bookings. The time-bound elements (T3) add further heterogeneity: a time-limited discount ("first three months free") in broadband actually increased accuracy and reduced decoy uptake, whereas a scarcity-framed countdown clock (“5 minutes left on this offer”) in the gym marketplace reduced the rate of correct choices and left behaviour unchanged on hotel booking websites. Taken together, these findings indicate that the power of online choice architecture is shaped jointly by its exact implementation details, consumers’ familiarity with the marketplace, and the difficulty of the decision task—where broadband tariffs seemed markedly harder to evaluate for customers than hotel stays.

4.3 Secondary Outcomes

4.3.1. How does online choice architecture impact market competition?

Next, we investigated whether the same online choice architecture interventions also changed how much people shopped around—that is, whether they clicked on a competitor website banner and whether they ultimately bought on that competitor website. Using the same modelling approach as in the primary outcomes, we estimated pooled linear probability models with treatment indicators (T1, T2, T3, control group omitted), marketplace indicators for gym and hotel (with broadband omitted), task order indicators for the second and third marketplace (with first marketplace omitted), and standard errors clustered by participant. We treated visits to a competitor site as our primary (binary) operationalisation of shopping around. Purchases on the competitor site were only secondary because the main and the alternative marketplaces featured nearly equivalent offers.

We found that reference pricing (T1) on its own significantly reduced visits to the competitor website relative to the control group, although it did not impact competitor purchases. In the main pooled model, the control-group baseline in the broadband market on the first task was 24.5% for clicking on a banner to see a competitor site and 18.1% for purchasing on the competitor site (Table B13). Relative to this baseline, T1 reduced competitor clicks by 2.3 p.p., statistically significant at the 5% level. In contrast, more overt reference pricing interventions (i.e., T2 with sensory manipulation and T3 with sensory manipulation plus a time-bound element) showed no significant effects on competitor visits, relative to control. And none of the treatments decreased competitor purchases, relative to the control condition.

These patterns are robust in pooled logistic regressions, and in participant-level proportion outcomes across the three tasks (Table B14), and they remain even when applying three robustness filters (Table B15, Table B16). The effect of subtle reference pricing on competitor visits stayed negative and statistically significant, while the effects of other reference pricing conditions on competitor visits—and of all 3 treatment conditions on competitor purchases—remained insignificant. Collapsing T1, T2 and T3 into a single “any treatment” indicator also led to no significant differences versus control for either competitor clicks or purchases (Table B17). Taken together, these results indicate that the OCAs tested did not uniformly reduce shopping around behaviours. The only consistent shift was a small, targeted reduction in competitor visits under the reference pricing on its own (T1).

4.3.2. What is the typical profile of those shopping around?

Additionally, we explored heterogeneity in shopping-around behaviours across participant traits and vulnerabilities[4]. Two factors were associated with lower likelihood to shop around. First, those who reported high stress when online shopping were 12.9 p.p. less likely to click on the alternative marketplace banner (p<0.01), relative to participants low on shopping stress. Second, there was a marginally significant effect of poor mental health on visiting alternative marketplaces, with participants experiencing mental health issues 3.2 p.p. less likely to click around (p<0.10). By contrast, financial insecurity was not significantly related to either clicks or purchases. One interpretation is that stress and mental health difficulties raise cognitive load and perceived hassle, making people more likely to satisfice—i.e., choose a “good enough” option rather than continue searching—and therefore remain on the main marketplace.

In contrast, high financial literacy and frequent bargain hunting were both associated with more shopping around in the experiment. Bargain hunters were 4.4 p.p. more likely to click around (p<0.05) and 2.9 p.p. more likely to purchase on the competitor site (p<0.05), relative to those who do not bargain hunt. And participants with high financial literacy showed even higher propensity to shop around. They were 12.0 p.p. more likely to click on the competitor banner (p<0.01) and 6.7 p.p. more likely to purchase on the competitor site (p<0.01). Notably, present bias and being a frequent online shopper were not significantly associated with more or less shopping around once other factors are controlled, suggesting that specific financial literacy skills and bargain hunting motives matter more than general familiarity with online shopping or time preferences in financial domains. The models with characteristics and vulnerabilities as predictors of shopping around behaviours are detailed in Tables B18 (clicks on alternative marketplace banner[2]), and Table B19 (purchases in the alternative marketplace[3]) in the Appendix.

4.3.2.1. Does the propensity to shop around differ across markets?

Similar to baseline marketplace differences we observe for choice selection, there were also market-by-market differences in shopping around, as evident in Table B13. Relative to broadband, participants in the gym task were 1.1 p.p. less likely to visit ($p < 0.05$) and 5.7 p.p. less likely to purchase ($p < 0.01$) on the competitor website. In the hotel booking journey, the pattern was more mixed: participants were also less likely to purchase in the alternative marketplace (by 1.5 p.p., $p < 0.01$), but they were actually 3.4 p.p. more likely to explore alternatives and click on the competitor marketplace, relative to broadband. In other words, gym membership shopping had the lowest rate of clicking around, followed by shopping for broadband, while shopping for hotel stays was associated with most click-outs to competitor sites. That said, we cannot rule out that some of these cross-market differences were due to differing presentations of the alternative site banner. For example, there were differences in the colour palettes of the competitor banners, and thus its visibility, which could have affected the likelihood of clicking on the alternative independent of underlying intent to shop around.

[4] This analysis should be considered exploratory as it was not preregistered.

Estimating treatment effects separately by marketplace (Table B20 and B21) showed that the hotel context was the only one where pure reference pricing (T1) significantly changed shopping around. In the hotel booking context, T1 reduced click-outs by 2.8 p.p. relative to the control group ($p<0.05$), while we found no comparable shift in clicking around between control and the treatment groups within the contexts of either broadband or gym. This pattern suggests that calibration of reference pricing really matters for market competition and that subtle reference price can reduce the perceived return to looking for alternative deals elsewhere. It is possible that this form of subtle OCAs signals a “fair” deal or anchors expectations, although more research is needed to evaluate different underlying mechanisms.

4.4 Further Outcomes

4.4.1. Does online choice architecture affect how long people spend shopping?

We first asked whether online choice architecture influences the time needed to complete experimental tasks. In the control group, participants spent on average just under four minutes selecting broadband tariffs, just under six minutes on gym memberships, and approximately six minutes and fifteen seconds on hotel offers. Across the three shopping journeys taken together, total time did not change significantly under any of the reference pricing interventions (T1, T2 or T3). The only (marginally) significant difference in time spent appeared in broadband under the most complex online choice architecture (reference pricing with sensory manipulation and a time-bound element). Under T3 in the broadband market, time on task increased by about 1.1 minutes in the full sample ($p < 0.10$) and by about 1.8 minutes among participants with poor mental health, compared to the control group ($p < 0.10$). For the group experiencing mental health difficulties, T1 also added a small amount of time in the broadband marketplace. In contrast, the point estimates for gym and hotel were not significantly impacted by variations in online choice architecture. Taken together, the evidence suggests that complex online choice architecture can slow users during selection of products that might already be difficult (broadband), especially for those with poorer mental health. Detailed results are in Table B22 of the Appendix.

4.4.2. What factors are associated with picking correct products and services?

The most surprising finding was an inverse relationship between participant objective performance and their decision confidence: participants who were more accurate during shopping tasks actually reported lower confidence in their choices. Moving from 0/3 to 3/3 correct choices was associated with -0.542 points on a 1 to 5 confidence scale ($p<0.01$), that is about -0.18 per one additional correct task. Furthermore, the treatments caused a simultaneous increase in confidence and decrease in decision quality. All three treatments produced small but statistically significant increases in self-reported confidence on a 5-point scale (ranging between 0.055 and 0.068 points). These effects might have been coincidental or related (i.e., that the treatments cause an increase in confidence, which then decreases decision quality). One argument for these factors being related (i.e., where confidence decreases decision quality), is that we found a negative correlation between confidence and decision quality in the control group (i.e., more confident participants generally made worse decisions). To investigate whether confidence mediates the treatment effect, we estimated two OLS models with decision quality (the proportion of correct choices) as the outcome: first including only the treatment indicators, and then adding confidence as an additional control variable. The treatment coefficients changed slightly when controlling for confidence—by about 8.7% for T1 and 8.2% for T2, while T3 remained statistically insignificant in both. Thus, while it is possible that changes in confidence can help explain a portion of the treatment effects, it is unlikely to be the main mediating variable. The regression results are in Tables B23 and B30 of the Appendix.

We also found an inverse relationship between correct choices and post-experimental mood. Participants who answered all three tasks correctly rated their mood after shopping 0.363 points lower on average ($p < 0.01$). Put differently, mood fell by about 0.12 points for each additional correct task. This pattern suggests that finding good deals did not necessarily make participants feel better about the shopping experience.

The relationship between perceived value for money and correctness was more complex. Within the control group, participants who selected the optimal offer in all three marketplaces rated value for money 0.197 points lower on a 5-point scale than those with no correct answers. This translates into 0.07 points lower value-for-money rating per additional correct marketplace task. Effects of treatment conditions varied with participant accuracy. At 0/3 correct choices, none of the treatments changed value-for-money ratings at the 5% level. But for participants in T2 and T3 who answered at least one marketplace task correctly, the significant, positive treatment x accuracy interactions applied. This meant that for the same number of correct responses (1/3, 2/3, or 3/3), participants in T2 and T3 rated value for money more favourably than those in the control. The negative relationship between accuracy and value-for-money, observed in the control group, was effectively neutralised under T2. And under T3, the accuracy x value-for-money relationship actually reversed, meaning that more accurate participants rated their selected products as better value for money than less accurate participants. Because the offers were identical across treatment arms, the T2, T3 interactions suggest that the combination of reference pricing with sensory manipulation (and a time-bound element in T3) made the same choices feel like better value.

By contrast, in the control group not featuring the OCA practices, an informed-scepticism mechanism might have explained the findings. Participants who investigated offers enough and, consequently, got experimental tasks right, might have uncovered caveats of the products and services offered, or their hidden fees, and small print. This extra information might have affected the perceived value for money of the products and services offered. Also, a greater understanding of the product and service features might have made participants more aware of remaining uncertainty—e.g., uncertainty about ideal download and upload speeds in the broadband market—and hence lowered their confidence despite objectively more correct choices.

4.4.3. If online choice architecture sways participants away from correct choices, where does it sway towards?

Beyond increasing decoy choices and reducing correct choices, the reference pricing, sensory manipulation and time-bound elements also raised average shopping expenditure, as evidenced in Table B24. In the first task, control condition baseline, 42.8% of participants picked the “too expensive” and 25.1% picked the “too cheap” options, while hotel booking marketplace had only 9.1% “too expensive” and 40.7% “too cheap” in the control condition baseline. In contrast, broadband already showed substantial overspending even without OCAs. 66.6% picked “too expensive” while only 15.1% picked “too cheap.” Treatments then strengthened the tendency to overspend. Reference pricing with sensory manipulation and time-bound elements (T3) significantly reduced “too cheap” picks by 2.7 p.p. ($p < 0.01$) and increased “too expensive” by 3.6 p.p. ($p < 0.01$). Reference pricing on its own (T1) and reference pricing with sensory manipulation (T2) significantly raised “too expensive” choices by 2.3 and 2.5 p.p., respectively (both $p < 0.01$).

In monetary terms, overspending differed by marketplace. Broadband showed the highest average overspending. Already in the control condition, participants shopping for the 1-year broadband plan overspent by about £83.0 on the annual contract cost, which increased to £94.1 in T1 ($p < 0.01$), £96.0 in T2 ($p < 0.01$), and £96.6 in T3 ($p < 0.01$). For gym memberships, the average overspend for the 9-month gym access that participants were purchasing was £8.4 in the control condition, rising to £10.7 in T1 ($p < 0.05$), £11.1 in T2 ($p < 0.01$) and £11.0 in T3 ($p < 0.01$). For detailed regression results, see Table B29 in the Appendix.

Together with the primary outcomes (fewer correct choices and more decoys), these results indicate that OCAs do not redirect choices toward “approximately correct” options. Instead, they steer participants away from cheaper and toward pricier options, raising average online shopping expenditure. And, crucially, this is not a novice mistake that fades with experience. Relative to the first round, participants became significantly more likely to pick “too expensive” options in later rounds, while too cheap choices fell by 2.0 p.p. in the second round, and 3.2 p.p. in the third round, relative to the start.

5 Discussion and recommendations

5.1 Recommendations

Our experiment shows that reference pricing - on its own and in combination with other OCAs - harms consumer decision making, driving people away from correct choices and towards decoy options. Based on our findings, regulators must act to better protect consumers from harm caused by these OCAs. We recommend that the CMA investigate reference pricing further and publish additional guidance for firms based around consumer harm, that Ofcom investigate broadband pricing practices to address the difficulties consumers face making choices in this market, and that all regulators engaged in this process continue conducting research to fill the gap in evidence on OCA practices and their impact on consumer behaviour.

5.1.1. CMA

We recommend that the CMA investigate reference pricing across markets and develop guidance centered on harms experienced by consumers because of the practice. The CMA has investigated reference pricing in the past, specifically in the context of online mattress sales (CMA, 2024). The guidance they generated for firms off the back of this investigation focused on whether or not the advertised reference price was 'genuine' - meaning that the firm had actually sold the product or service for the original 'referenced' price before offering the discount. The 'genuine' nature of a referenced price is determined by how long the product or service was sold at the original price and how many units were sold at that price, compared to the new discounted price. This guidance protects consumers from reference pricing that is misleading, with firms representing an original price that is listed for the sake of making a discounted price appear more appealing rather than genuinely intending to sell the product or service at that price. While we agree that consumers should be protected from such misleading reference pricing, our experiment reveals that reference pricing that displays genuine information can still be harmful by steering consumers away from good deals and towards decoy offers and in doing so causing them to overspend. These results held true in multiple experimental marketplaces. We therefore recommend that the CMA conduct further investigations into reference pricing across marketplaces, with particular attention paid to the harms consumers experience because of this and related OCA practices. The CMA should also publish additional guidance for firms to limit reference pricing that is misleading as well as reference pricing that is genuine but that harms consumers - whether that be financially, in terms of the quality of their decisions, or otherwise.

5.1.2. Ofcom

We recommend that Ofcom investigate pricing structures in the broadband market and the way this information is communicated to consumers. Ofcom has acted to improve pricing transparency in this market, recently banning inflation-linked mid contract price rises (Ofcom, 2024) and requiring firms to list price increases in pounds and pence. But our experiment reveals that broadband consumers face significant difficulty in choosing the right package based on specific instructions. The overwhelming majority of participants in the control did not choose the right option in the experiment, and this was made worse by T1 and T2. Adding to this picture, our previous research (Citizens Advice, 2025) into consumers' experiences negotiating with broadband providers for renewal deals revealed that pricing structures are obscure and difficult for consumers to navigate. Based on our findings, Ofcom should consider the harm of reference pricing and sensory manipulation to consumers shopping for broadband packages, as well as why consumers would find it so difficult to make the right choice in this marketplace independently of OCA practices.

5.1.3. All regulators

We recommend that all regulators we engaged with throughout this research continue to conduct research examining the impact of OCA practices on consumer behaviour, particularly for vulnerable consumers. Citizens Advice's previous research indicates that though there is clear reasoning from behavioural science suggesting vulnerable consumers would be disproportionately impacted by OCA practices, little high-quality evidence exists on this interaction. This experiment makes an important step to address part of this research gap, but much more research is needed to properly understand the impact of OCA practices and protect consumers from harm. This must include research investigating a range of relevant OCA practices, in a range of relevant markets, and the impacts on consumers in a range of vulnerable circumstances. For this to be done effectively, regulators must contribute to commissioning research to close evidence gaps in the markets they oversee. Potential areas for further research are detailed below.

5.2 Design summary

This study used an online experiment to examine how specific online choice architecture (OCA) practices influence consumer decisions. A large UK sample, representative in terms of gender, age and UK regions, together with a sample of participants experiencing financial insecurity or poor mental health, completed three realistic shopping tasks on online marketplace replicas. Each marketplace had a predefined optimal and a predefined decoy (i.e., poor) choice. We randomly assigned participants to one of four conditions with equal probability: a control, a subtle reference pricing treatment (T1), reference pricing plus sensory manipulation (T2), and reference pricing plus sensory manipulation and countdown timers or time-limited discounts (T3). In T1, product pages displayed unsubstantiated reference prices ("was £X, now £Y") on some suboptimal, incorrect deals to inflate perceptions of their value. The magnitude of referenced prices mirrored real-world shopping websites. In T2, the reference priced options were also made visually salient with bold colours or larger text to draw attention, a form of sensory manipulation. In T3, we further added time pressure (via countdown timers) or time-limited discounts on the referenced and salient deals. All other aspects of the shopping interfaces, including the layouts, market journeys and other offer details, were held constant across the four groups. This enabled us to investigate the incremental effect of each employed OCA practice separately. Participants could also click on a banner to shop around on a competitor's site in each marketplace. This enabled us to observe the effects of reference pricing, sensory manipulation and time-bound elements on market competition. The randomisation was successful and balanced in terms of demographics across groups, supporting internal validity. Taken together, our experiment provides a controlled investigation of individual and combined OCA practices on consumer choice, with a focus on whether these tactics cause welfare-reducing decisions and whether vulnerable consumers are disproportionately affected.

5.3 Summary of key results

Overall, exposure to the OCA practices led to worse consumer choices. Participants who were shown any reference pricing were 1.7 percentage points less likely to choose the objectively correct option and 2 percentage points more likely to pick a decoy, than those in the control condition. In particular, reference pricing on its own (T1) significantly reduced propensity to choose correctly, and making the reference price salient with sensory manipulation (T2) magnified this. The addition of time-bound elements (T3) had mixed effects: in the gym scenario, a countdown timer further reduced the propensity to choose correctly, consistent with scarcity claims reinforcing bad choices. But, in the broadband scenario, a 'first 3 months free' tag raised the rate of correct choices even relative to the control group. This anomaly may reflect that biases, such as present bias, might not have been activated in the experiment, since participants were not shopping for the deals out of their pockets and were only incentivised to pick the objectively correct choice (i.e., entered into a lottery for each correct choice).

The broadband task was difficult for everyone. In the control group, only about 10% chose the optimal offer, suggesting that the complexity of broadband offers can overwhelm consumers even in the absence of harmful online choice architecture. Across markets, we found that financially insecure customers and those with poor mental health performed worse on average, about 5.0 - 7.9 percentage points less likely to select the optimal deal. This confirms the hypothesis that these groups face additional challenges while shopping online. However, they were equally likely to be harmed by reference pricing interventions as the general population. The magnitude of the reference price effect was roughly equal across subgroups.

We also observed impacts on market competition: participants in the pure reference pricing group (T1) were significantly less likely to “shop around” by clicking the competitor’s banner than control participants. This conforms to the concerns that these tactics reduce consumers’ willingness to explore alternatives and, consequently, hinder market competition. In our data, individuals with higher stress or mental health struggles were especially unlikely to browse the alternative site, whereas self-identified bargain-hunters and those with strong financial literacy were more likely to seek other offers. Finally, despite the objectively worse outcomes under harmful online choice architecture, participants paradoxically felt more confident and found the tasks easier when reference prices were present. This suggests that these tactics create an illusion of simplicity or deal value while they mislead. Together, these results provide evidence that common online sales tactics like reference prices, sensory manipulation, and scarcity claims can meaningfully impair consumer decision making and may undermine market competition.

5.4 How it fits into the existing literature

We contribute to a body of literature on how reference pricing, and its combination with other ‘dark’ OCAs, impacts consumer decision-making. We explicitly tested whether OCAs were more harmful for vulnerable consumers than for the general population. In this regard, our study builds upon BIT (2024), which lays out theoretical reasons for why vulnerable customers might disproportionately suffer poor outcomes as a result of OCA, and extends the experimental literature on differential effects of harmful online tactics by vulnerability (especially Lupiáñez-Villanueva et al., 2022; Luguri & Strahilevitz, 2021; Zac et al., 2023). We organise the section into four strands—reference pricing, sensory manipulation, time-bound elements, and choice complexity and competition. For each of the first three strands, we first summarise evidence on average effects in the general population and then review evidence on differential impacts for vulnerable consumers, relating both to our findings.

5.4.1. Reference pricing

Classic and recent studies (e.g., Lichtenstein, Burton & O'Hara, 1988; Lichtenstein, Burton & Karson, 1991; Jensen et al., 2003; Krishnan et al., 2013; Kan et al., 2014) report that advertised "was £X/now £Y" prices (even if exaggerated) generally increase perceived value of deals, improve the attitude towards those deals, increase shopping intentions and even reduce intentions to shop elsewhere. However, past research often asked hypothetical questions, and used print ads with small student samples. We move beyond stated intentions to measuring revealed preferences in online shopping settings, by employing realistic online shopping replicas with a large sample representative of the general UK population in gender, age, and region. Consistent with past literature, we find that reference pricing distorts decisions away from the objectively optimal deal and can deter shopping around.

Prior work rarely tested heterogeneity in reference price effects by employing direct vulnerability measures. We add new evidence by directly measuring financial insecurity and poor mental health as part of our research. Although vulnerable groups perform worse on average, we do not find disproportionate harms from reference pricing relative to the general population.

5.4.2. Sensory manipulation

Similarly, a substantial literature exists on sensory manipulation in consumer choice. Past studies (e.g., Milosavljevic et al., 2012; Clement et al., 2015) report that making an option brighter or changing its colours can increase time people gaze at that option which in turn predicts choice, especially for fast decisions or weak priors, typically in simple A/B tasks and frequently in grocery contexts.

In terms of sample and marketplaces, this research is most closely related to Lupiáñez-Villanueva et al. (2022) and Zac et al. (2025). Both studies find that visually manipulative elements increase choices inconsistent with stated preferences or raise willingness to accept a fictitious offer. But the interactive effects of vulnerability and these manipulative elements are mixed: Zac et al. (2025) report no distinctive effects by age, income, or education beyond general-population effects, while Lupiáñez-Villanueva et al. (2022) find some additional effects for older and lower-education participants, typically only at the 10% level. Our study extends this literature by testing three online marketplaces, combining sensory manipulation with reference pricing and time-bound elements to identify incremental and joint effects, and using objective correctness in multi-option menus. We do so with a large UK-representative sample and targeted samples identified by direct measures of financial insecurity and poor mental health rather than demographic proxies. Consistent with prior work, we find that sensory manipulation amplifies the negative effect of reference pricing and that vulnerable consumers are not differentially impacted by salient reference pricing.

5.4.3. Time-bound elements

Evidence on countdown clocks and time-limited discounts suggests that they can alter choices (e.g., Sugden, Wang & Zizzo, 2019; Aggarwal, Jun & Hun, 2011; Godinho, Prada & Garrido, 2016), but much of it comes from different marketplace contexts, is not always experimental, and rarely centres on consumer harm. Literature on time-limited discounts has largely been survey-based (Citizens Advice, 2023). In the countdown-timer work, lab studies often use ultra-short decision deadlines or game-like interfaces, while field studies examine week-long promotions for durable goods (e.g., condiments) where faster purchase may be welfare-neutral. In hotel booking context, Godinho, Prada and Garrido (2016) report no deterioration in average choices with a 3-minute countdown, where choices are evaluated by previous utility ratings of hotel photos from about 40 participants. We built on this by providing a more robust definition of correct choice, by placing countdown clocks only on sub-optimal options and by combining time-bound elements with other OCA practices. In our research, countdowns reduced correct choice in gyms, had no detectable effect in hotels, and a time-limited discount in broadband deters selection of that option, consistent with the mixed pattern in prior evidence.

Evidence on the impact of time-bound elements on choices and wellbeing of vulnerable consumers is limited. The key exception, Luguri and Strahilevitz (2021), finds higher uptake of free-trial identity protection (akin to a time-limited discount) among lower-education participants. In our realistic multi-option marketplaces, where time-bound elements were combined with reference pricing and sensory manipulation, and direct measures of financial insecurity and mental health were used to measure vulnerability, we did not observe that vulnerability gradient. Vulnerable participants were no more influenced by time-bound elements than others.

Beyond OCAs, our results speak to decision-making under complexity. Even without OCAs, many participants struggled. In broadband, only about 10% of the control group chose the objectively optimal deal. That task presented six offers varying on multiple attributes (download/upload speed, monthly price, upfront fee) and a usage scenario (video calls, file transfers, smart-home devices) likely inducing choice overload. This accords with theory on limited attention and sellers' incentives to obfuscate (Persson, 2020) and evidence that too much choice and information can paradoxically reduce consumers' ability to make good choices as they resort to simplification heuristics (Zac et al., 2025). Our evidence is consistent with this: faced with a complex set of broadband offers, many participants defaulted to picking an option that wasn't cost-optimal. Also, many gravitated toward high-speed packages beyond their needs, a 'bias towards faster speed' that has been previously reported by Lovett and McDonald (2024) in the UK context.

5.5 Strengths and limitations

This study offers several notable strengths. First, the design of the present study enabled causal inference. Our online randomised controlled trial assigned participants to a control arm and multiple treatment arms with equal probability, allowing us to make causal claims about the impact of selected online choice architecture practices. Our stepwise treatment structure further isolated additive and joint effects of different online choice architecture interventions: T1 implemented pure reference pricing; T2 bundled reference pricing with sensory manipulation; and T3 further added a time-bound element, enabling identification of the incremental effect of each OCA practice as well as their interactions. Second, the experimental tasks were designed to closely mirror real online marketplaces: participants navigated responsive replica websites on their own devices, browsed multiple offers (including add-ons), and went through a full checkout process for each product or service. The inclusion of three distinct market contexts (broadband plans, gym memberships, and hotel bookings), further added insights into how choices are made for different types of purchases. Third, we employed a large sample of 8,046 UK adults that was nationally representative in terms of age, gender, and region, which enhances the generalizability of the findings. Our sample size also allowed for robust subgroup analyses: the experiment included large numbers of vulnerable participants, enabling precise estimates of how these participants fared. And, crucially, the experiment was co-designed with regulators, ensuring survey questions, harm metrics (objective correctness, decoy selection, and shopping-around behaviour), and population coverage directly map onto the evidence needed for policy; this collaboration also informed our target population definition (UK online consumers exposed to OCAs, plus participants experiencing financial insecurity and poor mental health).

Despite its contributions, the research has limitations. The experimental design required certain compromises in realism for the sake of control and measurement. For example, participants shopped for hypothetical needs and were forced to make a purchase in each scenario (they could not opt to defer or exit without buying), and navigation to a competitor's site was provided via a banner click on the main marketplace site instead of an open web search. These constraints were necessary to keep the tasks within an approx. 15-minute completion time and to ensure well-defined "optimal" and decoy offers for measuring objective choice quality, but they do mean the shopping experience was not entirely natural. Another issue is the attrition and exclusion rate: a large number of initial respondents failed comprehension checks or dropped out during the process. This suggests some participants found the tasks cognitively taxing or difficult, especially in the broadband market where even in the control group only about 10% chose the optimal deal. The study addressed this limitation by simplifying instructions in piloting and, importantly, the attrition was approximately equal across all experimental arms, so it likely did not bias the comparison between conditions.

Another potential limitation is that participants faced incentivised, albeit not fully natural, stakes (with a lottery incentive for correct answers rather than spending their own money), which might have altered certain behaviours. To illustrate, the broadband task's anomaly - where a "3 months free" promotional tag (a time-limited discount) somewhat improved choices in that scenario—could be partly due to the lack of real financial consequences of decision or due to present-bias not being activated in the experimental context. These limitations were recognized and mitigated where possible (through the validity checks and balanced design), and they highlight areas where further research can build upon the findings.

5.6 Future directions

There are several avenues for future research. First, studies could apply similar experimental designs to other market domains. For example, our desk research suggests that flight booking sites, electronics retailers, and mobile phone plan providers frequently apply potentially harmful online choice architecture tactics. In travel sites, a notable issue is drip pricing, where some fees are revealed only at checkout. Investigating drip pricing in an online randomised controlled trial would be especially valuable. Second, future work should explore a broader range of dark patterns and their variants. This includes testing different types of scarcity and urgency cues (e.g. “Only X rooms left” inventory messages or time-limited discount pop-ups), different forms of sensory manipulation, and varying how reference prices are presented. Within reference pricing, experiments might manipulate the size and plausibility of the “Was £X” price to see if implausibly high reference prices differ in impact. Additionally, it would be worthwhile to examine brand-specific dynamics. Consumers might react differently to an exaggerated “Was £100” claim if it is made by a luxury, or well-known brand vs. by an unknown brand that we introduced in our experiment. Third, methodological alterations would add further insight. For example, introducing the option not to buy or to delay a decision (rather than forcing a choice) could help build a more comprehensive understanding of whether harmful OCAs increase the likelihood of unnecessary or impulsive purchases when a no-purchase option exists. Similarly, giving participants more freedom to shop around, such as letting them search external competitor sites (instead of only clicking a provided link), would shed light on how these tactics influence the willingness to explore alternatives. Incorporating real economic incentives or consequences is another crucial extension: if participants were spending their own money or if outcomes had tangible financial payoffs, phenomena like present bias or loss aversion might be activated, possibly changing the impact of tactics like “first 3 months free”. Researchers should also consider the role of choice set complexity. Varying the number of options presented could help determine how choice overload interacts with dark patterns. By testing smaller or larger assortments of offers, one can see if the manipulative tactics become more or less effective. Finally, an important direction is to develop and evaluate interventions to mitigate the harm of exploitative online tactics.

This report aims to address the gaps in evidence regarding how OCAs influence consumer decision-making, particularly among vulnerable consumers. The findings presented in this report offer valuable insights into this issue; however, further work is still needed to achieve a comprehensive understanding. We urge regulators to take action to close this evidence gap and ensure that consumers are properly protected.

6 References

Aggarwal, P., Jun, S.Y., & Huh, J. (2011). Scarcity messages: A consumer competition perspective. *Journal of Advertising, 40*(3), 19–30

Akesson, J. (2025). Understanding the impact of online choice architecture on vulnerable customers [Pre-registration]. AEA RCT Registry. https://www.socialscienceregistry.org/trials/16094 (Accessed 28 August 2025).

Behavioural Insights Team, Citizens Advice (2024). Review of online choice architecture and vulnerability. Report. https://www.bi.team/wp-content/uploads/2024/07/Review-of-Online-Choice-Architecture-and-Vulnerability-July-2024-.pdf (Accessed 18 November 2025).

Citizens Advice (2022). *Tricks of the trade: how online customer journeys create consumer harm and what to do about it.* Policy Research. https://assets.ctfassets.net/mfz4nbgura3g/4UtD4GkI7cmdVrps2Uy2ZG/378374c06e75496974571cfd6a9237bf/OCA_20report_20-_20version_202_20_5_.pdf (Accessed 18 November 2025).

Citizens Advice (2025). *The Real Cost of Hidden Deals.* Policy Research. https://assets.ctfassets.net/mfz4nbgura3g/1P7DsKlsttXtP6CB4CdkpB/f7f57732ffe00df38142605f9568cd89/TheRealCostofHiddenDeals__1_.pdf (Accessed 18 November 2025).

Clement, J., Aastrup, J., & Forsberg, S. C. (2015). Decisive visual saliency and consumers' in-store decisions. Journal of Retailing and Consumer Services, 22, 187-194.

Competition and Markets Authority (2022): *Online Choice Architecture: How Digital Design Can Harm Competition and Consumers*. Discussion Paper.

Competition and Markets Authority (2024). *Discount and Reference Pricing Principles: Selling Mattresses Online*. Guidance. https://assets.publishing.service.gov.uk/media/66ab4347a3c2a28abb50db3c/Discount_and_reference_pricing_principles.pdf (Accessed 18 November 2025).

Di Geronimo, L., Braz, L., Fregnan, E., Palomba, F., & Bacchelli, A. (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* (pp. 1-14).

European Opinion Research Group. (2003). The mental health status of the European population. *Directorate-General Press and Communication ''Opinion Polls, Europe Direct''. SANCO Directorate General, Brussels, 27.*

Financial Conduct Authority. (2024a). Financial Lives 2024 survey: Key findings from the FCA's Financial Lives May 2024 survey. Financial Conduct Authority.

Financial Conduct Authority. (2024b). Financial Lives 2024 survey: Vulnerability & financial resilience—Selected findings. Financial Conduct Authority. https://www.fca.org.uk/publication/financial-lives/fls-2024-vulnerability-financial-resilience.pdf (Accessed 18 November 2025).

Godinho, S., Prada, M., & Garrido, M. V. (2016). Under pressure: An integrative perspective of time pressure impact on consumer decision-making. *Journal of International Consumer Marketing*, *28*(4), 251-273.

Holmes, W. C. (1998). A short, psychiatric, case-finding measure for HIV seropositive outpatients: performance characteristics of the 5-item mental health subscale of the SF-20 in a male, seropositive sample. *Medical care*, *36*(2), 237-243.

Jensen, T., Kees, J., Burton, S., & Turnipseed, F. L. (2003). Advertised reference prices in an internet environment: effects on consumer price perceptions and channel search intentions. *Journal of Interactive Marketing*, *17*(2), 20-33.

Kan, C., Lichtenstein, D. R., Grant, S. J., & Janiszewski, C. (2014). Strengthening the influence of advertised reference prices through information priming. *Journal of Consumer Research*, *40*(6), 1078-1096.

Kelly, M. J., Dunstan, F. D., Lloyd, K., & Fone, D. L. (2008). Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: a comparison of five different methods. *BMC psychiatry*, *8*, 1-9.

Krishnan, Balaji C., Sujay Dutta, and Subhash Jha. "Effectiveness of exaggerated advertised reference prices: the role of decision time pressure." *Journal of retailing* 89, no. 1 (2013): 105-113.

Lichtenstein, D. R., Burton, S., & O'Hara, B. S. (1989). Marketplace attributions and consumer evaluations of discount claims. *Psychology & Marketing*, *6*(3), 163-180.

Lichtenstein, D. R., Burton, S., & Karson, E. J. (1991). The effect of semantic cues on consumer perceptions of reference price ads. *Journal of Consumer research*, *18*(3), 380-391.

List, J. A. (2020). *Non est disputandum de generalizability? A glimpse into the external validity trial* (No. w27535). National Bureau of Economic Research.

Lovett, D., & McDonald, S. (2024, November 21). *Decoy offers and consumer choice*. Which?.Policy Research Paper. https://www.which.co.uk/policy-and-insight/article/decoy-offers-and-consumer-choice-arLiU9B2jXYW (Accessed 18 November 2025).

Luguri, J., & Strahilevitz, L. J. (2021). Shining a light on dark patterns. *Journal of Legal Analysis, 13*(1), 43-109.

Lupiáñez-Villanueva, F., Boluda, A., Bogliacino, F., Liva, G., Lechardoy, L., & de las Heras Ballell, T. R. Behavioural study on unfair commercial practices in the digital environment–dark patterns and manipulative personalisation–Final report. Publications Office of the European Union (2022). data. europa. eu. Available at: data. europa. eu/doi/10.2838/859030.

Lusardi, A., & Mitchell, O. S. (2008). Planning and financial literacy: How do women fare?. *American economic review*, *98*(2), 413-417.

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on human-computer interaction*, *3*(CSCW), 1-32.

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of consumer psychology*, *22*(1), 67-74.

Persson, P. (2018). Attention manipulation and information overload. *Behavioural Public Policy*, *2*(1), 78-106.

Rivera-Riquelme, M., Piqueras, J. A., & Cuijpers, P. (2019). The Revised Mental Health Inventory-5 (MHI-5) as an ultra-brief screening measure of bidimensional mental health in children and adolescents. *Psychiatry research*, *274*, 247-253.

Rugulies, R., Bültmann, U., Aust, B., & Burr, H. (2006). Psychosocial work environment and incidence of severe depressive symptoms: prospective findings from a 5-year follow-up of the Danish work environment cohort study. *American journal of epidemiology*, *163*(10), 877-887.

Sugden, R., Wang, M., & Zizzo, D. J. (2019). Take it or leave it: Experimental evidence on the effect of time-limited offers on consumer behaviour. *Journal of Economic Behavior & Organization*, *168*, 1-23.

Ofcom. (2024, July 19). Telecoms customers must be told upfront in pounds and pence about any price rises their provider includes in their contract, under new consumer protection rules announced today by Ofcom. Ofcom. https://www.ofcom.org.uk/phones-and-broadband/bills-and-charges/ofcom-bans-mid-contract-price-rises-linked-to-inflation (Accessed 18 November 2025).

Ten Have, M., Van Bon-Martens, M. J., Schouten, F., Van Dorsselaer, S., Shields-Zeeman, L., & Luik, A. I. (2024). Validity of the five-item mental health inventory for screening current mood and anxiety disorders in the general population. *International Journal of Methods in Psychiatric Research*, *33*(3), e2030.

Zac, A., Huang, Y. C., von Moltke, A., Decker, C., & Ezrachi, A. (2023). Dark patterns and consumer vulnerability. *Behavioural Public Policy*, 1-50.

7 Appendices

7.1 Appendix A – Mental Health Inventory (MHI-5)

Item A1. 5-Item Mental Health Inventory (MHI-5)

Please indicate how much of the time during the last month you have experienced the following:

1. been a very nervous person?
2. felt calm and peaceful?
3. felt downhearted and blue?
4. been a happy person?
5. felt so down in the dumps that nothing could cheer you up?

For each item, 6-point likert scale

1. All of the time
2. Most of the time
3. A good bit of the time
4. Some of the time
5. A little of the time
6. None of the time

7.2 Appendix B – Experiment Results

Table B1. Main LPM specifications for predicting correct and decoy choice

	(1) Correct choice	(2) Decoy choice
T1 (reference pricing)	-0.018** (0.008)	0.016** (0.006)
T2 (reference pricing + sensory manipulation)	-0.024*** (0.008)	0.032*** (0.006)
T3 (reference pricing + sensory manipulation + time-bound element)	-0.007 (0.008)	0.013** (0.006)
Gym task	0.232*** (0.006)	0.193*** (0.006)
Hotel task	0.352*** (0.007)	-0.046*** (0.004)
Second task (order2)	0.005 (0.007)	0.005 (0.005)
Third task (order3)	0.008 (0.007)	0.013** (0.005)
Constant	0.112*** (0.007)	0.086*** (0.006)
Observations	24,138	24,138

*Notes: Main linear probability model specifications for (1) predicting correct choices, (2) predicting decoy choices. Standard errors clustered by participant in parentheses. Both models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). * p<.10, ** p<.05, *** p<.01.*

Table B2. Main logistic specifications for predicting correct and decoy choice

	(1) Correct choice	(2) Decoy choice
T1 (reference pricing)	-0.097** (0.045)	0.133** (0.053)
T2 (reference pricing + sensory manipulation)	-0.129*** (0.044)	0.262*** (0.053)
T3 (reference pricing + sensory manipulation + time-bound element)	-0.038 (0.044)	0.109** (0.054)
Gym task	1.471*** (0.042)	1.275*** (0.043)
Hotel task	1.978*** (0.043)	-0.615*** (0.059)
Second task (order2)	0.025 (0.035)	0.045 (0.046)
Third task (order3)	0.043 (0.036)	0.109** (0.045)
Constant	-2.111*** (0.050)	-2.304*** (0.057)
Observations	24,138	24,138

*Notes: Logistic regression specifications for (1) predicting correct choices, (2) predicting decoy choices. Standard errors clustered by participant in parentheses. Both models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). * p<.10, ** p<.05, *** p<.01.*

Table B3. Robustness checks for predicting correct choice

	(1) Correct choice robustness check 1: only the bottom 95% of completion times included	(2) Correct choice robustness check 2: only those who passed first comprehension test included	(3) Correct choice robustness check 3: only those who failed at most 1 attention check included
T1 (reference pricing)	-0.020** (0.009)	-0.025** (0.010)	-0.020** (0.009)
T2 (reference pricing + sensory manipulation)	-0.025*** (0.009)	-0.027*** (0.009)	-0.024*** (0.008)
T3 (reference pricing + sensory manipulation + time-bound element)	-0.009 (0.009)	-0.014 (0.009)	-0.008 (0.009)
Gym task	0.241*** (0.006)	0.247*** (0.007)	0.233*** (0.006)
Hotel task	0.373*** (0.007)	0.396*** (0.007)	0.358*** (0.007)
Second task (order2)	0.005 (0.007)	0.004 (0.008)	0.004 (0.007)
Third task (order3)	0.009 (0.007)	0.004 (0.008)	0.007 (0.007)
Constant	0.112*** (0.008)	0.116*** (0.008)	0.113*** (0.008)

*Notes: Linear probability models for predicting correct choice, copying the main specification but with the following amendments: 1) dropping participants with unusually fast completion times (i.e.., below the 5th percentile of total survey completion time), 2) dropping participants who failed the first comprehension check for the instructions but passed the second one (those failing both are already screened out), 3) dropping participants who failed two or more of the three geography trivia attention checks. All 3 models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B4. Robustness checks for predicting decoy choice

	(1) Decoy choice robustness check 1: only the bottom 95% of completion times included	(2) Decoy choice robustness check 2: only those who passed first comprehension test included	(3) Decoy choice robustness check 3: only those who failed at most 1 attention check included
T1 (reference pricing)	0.019*** (0.006)	0.019*** (0.007)	0.017*** (0.006)
T2 (reference pricing + sensory manipulation)	0.031*** (0.007)	0.032*** (0.007)	0.032*** (0.006)
T3 (reference pricing + sensory manipulation + time-bound element)	0.013** (0.007)	0.017** (0.007)	0.014** (0.006)
Gym task	0.191*** (0.006)	0.198*** (0.007)	0.194*** (0.006)
Hotel task	-0.049*** (0.005)	-0.046*** (0.005)	-0.048*** (0.004)
Second task (order2)	0.003 (0.006)	0.004 (0.006)	0.005 (0.005)
Third task (order3)	0.012** (0.006)	0.014** (0.006)	0.014** (0.006)
Constant	0.089*** (0.006)	0.084*** (0.007)	0.086*** (0.006)

*Notes: Linear probability models for predicting decoy choice copying the main specification but with the following amendments: 1) dropping participants with unusually fast completion times (i.e.., below the 5th percentile of total survey completion time), 2) dropping participants who failed the first comprehension check for the instructions but passed the second one (those failing both are already screened out), 3) dropping participants who failed two or more of the three geography trivia attention checks. All 3 models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B5. Heterogeneity analyses for predicting correct choice

	(1) Correct choice, financial insecurity (FI)	(2) Correct choice, Poor mental health (MH)	(3) Correct choice, frequent online shoppers (FSHOP)	(4) Correct choice, stressed shoppers (STR)	(5) Correct choice, bargain hunters (BH)	(6) Correct choice, high financial literacy (LIT)	(7) Correct choice, present bias (PB)
T1	-0.024** (0.010)	-0.030** (0.012)	0.002 (0.017)	-0.019** (0.009)	0.007 (0.014)	-0.010 (0.012)	-0.019** (0.010)
T2	-0.027*** (0.009)	-0.039*** (0.012)	-0.002 (0.017)	-0.023*** (0.008)	0.002 (0.014)	-0.024** (0.012)	-0.026*** (0.009)
T3	-0.018* (0.010)	-0.012 (0.012)	-0.004 (0.016)	-0.006 (0.009)	0.012 (0.014)	0.022* (0.012)	-0.011 (0.009)
FI	-0.079*** (0.014)						
T1 × FI	0.028 (0.020)						
T2 × FI	0.013 (0.019)						
T3 × FI	0.045** (0.020)						
gym_task	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)
hotel_task	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)
order2	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)
order3	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)
MH		-0.050*** (0.012)					
T1 × MH		0.024 (0.017)					
T2 × MH		0.028* (0.017)					
T3 × MH		0.011 (0.017)					
FSHOP			0.026* (0.014)				
T1 × FSHOP			-0.027 (0.020)				
T2 × FSHOP			-0.029 (0.019)				
T3 × FSHOP			-0.005 (0.019)				
STR				-0.067** (0.034)			
T1 × STR				0.013 (0.051)			

T2 × STR				-0.056 (0.050)			
T3 × STR				-0.055 (0.045)			
BH					0.071*** (0.013)		
T1 × BH					-0.036** (0.017)		
T2 × BH					-0.040** (0.017)		
T3 × BH					-0.029* (0.017)		
LIT						0.133*** (0.012)	
T1 × LIT						-0.013 (0.016)	
T2 × LIT						-0.002 (0.016)	
T3 × LIT						-0.049*** (0.016)	
PB							-0.029** (0.014)
T1 × PB							0.001 (0.020)
T2 × PB							0.004 (0.020)
T3 × PB							0.014 (0.020)
Constant	0.130*** (0.008)	0.138*** (0.010)	0.093*** (0.013)	0.114*** (0.007)	0.066*** (0.011)	0.036*** (0.010)	0.119*** (0.008)
Observations	24,138	24,138	24,138	24,138	24,138	24,138	24,138

*Notes: Main linear probability model specifications for predicting correct choices, extended to include indicator variables for vulnerability and personal characteristics, and their interactions with treatment indicators (T1–T3). Financial insecurity (FI) is defined as income < £15,000; poor mental health (MH) is defined by MHI-5 score ≤ 68; frequent online shopping (FSHOP) is conceptualised as shopping online either about once or twice per week or about once or twice a month; stressed shoppers (STR) are defined as those who ticked in the survey that they find online shopping either extremely or very stressful; bargain hunters (BH) are characterised by their stated preference to either always or frequently look for discounts; high financial literacy (LIT) is operationalised as answering all 3 financial literacy questions in the survey correctly; present bias (PB) is defined by a preference reversal in the survey whereby participants prefer receiving £100 now, rather than £110 in 1 week but also receiving £110 in 5 weeks, rather than £100 in 4 weeks. All models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B6. Heterogeneity analyses for predicting decoy choice

	(1) Decoy choice, financial insecurity (FI)	(2) Decoy choice, Poor mental health (MH)	(3) Decoy choice, frequent online shoppers (FSHOP)	(4) Decoy choice, stressed shoppers (STR)	(5) Decoy choice, bargain hunters (BH)	(6) Decoy choice, high financial literacy (LIT)	(7) Decoy choice, present bias (PB)
T1	0.018** (0.007)	0.018** (0.009)	0.026** (0.013)	0.014** (0.006)	0.010 (0.010)	0.009 (0.009)	0.008 (0.007)
T2	0.031*** (0.007)	0.033*** (0.009)	0.028** (0.013)	0.031*** (0.006)	0.023** (0.011)	0.024** (0.009)	0.031*** (0.007)
T3	0.018** (0.007)	0.025*** (0.009)	0.017 (0.012)	0.012* (0.006)	0.011 (0.011)	0.017* (0.010)	0.010 (0.007)
FI	-0.000 (0.010)						
T1 × FI	-0.009 (0.015)						
T2 × FI	0.001 (0.015)						
T3 × FI	-0.026* (0.015)						
gym_task	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)
hotel_task	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)
order2	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)
order3	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)
MH		0.006 (0.009)					
T1 × MH		-0.005 (0.012)					
T2 × MH		-0.003 (0.013)					
T3 × MH		-0.023* (0.013)					
FSHOP			0.013 (0.010)				
T1 × FSHOP			-0.014 (0.015)				
T2 × FSHOP			0.005 (0.015)				
T3 × FSHOP			-0.005 (0.014)				
STR				-0.033 (0.026)			
T1 × STR				0.046 (0.040)			

T2 × STR				-0.002 (0.038)			
T3 × STR				0.028 (0.038)			
BH					0.004 (0.009)		
T1 × BH					0.009 (0.013)		
T2 × BH					0.014 (0.013)		
T3 × BH					0.003 (0.013)		
LIT						0.015* (0.009)	
T1 × LIT						0.011 (0.013)	
T2 × LIT						0.013 (0.013)	
T3 × LIT						-0.007 (0.013)	
PB							-0.009 (0.010)
T1 × PB							0.035** (0.015)
T2 × PB							0.002 (0.015)
T3 × PB							0.013 (0.015)
Constant	0.086*** (0.006)	0.083*** (0.008)	0.076*** (0.010)	0.087*** (0.006)	0.083*** (0.008)	0.077*** (0.008)	0.088*** (0.006)

*Notes: Main linear probability model specifications for predicting decoy choices, extended to include indicator variables for vulnerability and personal characteristics, and their interactions with treatment indicators (T1–T3). Financial insecurity (FI) is defined as income < £15,000; poor mental health (MH) is defined by MHI-5 score ≤ 68; frequent online shopping (FSHOP) is conceptualised as shopping online either about once or twice per week or about once or twice a month; stressed shoppers (STR) are defined as those who ticked in the survey that they find online shopping either extremely or very stressful; bargain hunters (BH) are characterised by their stated preference to either always or frequently look for discounts; high financial literacy (LIT) is operationalised as answering all 3 financial literacy questions in the survey correctly; present bias (PB) is defined by a preference reversal in the survey whereby participants prefer receiving £100 now, rather than £110 in 1 week but also receiving £110 in 5 weeks, rather than £100 in 4 weeks. All models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B7. Further heterogeneity analyses for predicting correct choice

	(1) Correct choice	(2) Correct choice, poor mental health (MH)	(3) Correct choice, financial insecurity (FI)	(4) Correct choice, Any vulnerability
any treatment	-0.017** (0.007)	-0.027*** (0.010)	-0.023*** (0.008)	-0.030*** (0.011)
MH		-0.050*** (0.012)		
any treatment × MH		0.021 (0.014)		
FI			-0.079*** (0.014)	
any treatment × FI			0.028* (0.016)	
any vulnerability				-0.061*** (0.012)
(any vulnerability) × (any treatment)				0.024* (0.014)
Constant	0.311*** (0.006)	0.337*** (0.009)	0.329*** (0.007)	0.347*** (0.010)
Observations	8,046	8,046	8,046	8,046

*Notes: The linear probability models above are predicting correct choices, with an adjusted treatment indicator independent variable; i.e., the treatment assignment is simplified to just one binary variable, coded as 1 if the participant is assigned to any treatment condition (T1, T2, T3), and 0 if they are assigned to the control group. Additionally, specification (4) combines FI and MH into an "any vulnerability" indicator. Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B8. Further heterogeneity analyses for predicting decoy choice

	(1) Decoy choice	(2) Decoy choice, poor mental health (MH)	(3) Decoy choice, financial insecurity (FI)	(4) Decoy choice, Any vulnerability
any treatment	0.020*** (0.005)	0.026*** (0.007)	0.022*** (0.006)	0.031*** (0.008)
MH		0.006 (0.009)		
any treatment × MH		-0.011 (0.010)		
FI			-0.000 (0.010)	
any treatment × FI			-0.011 (0.012)	
any vulnerability				0.010 (0.009)
(any vulnerability) × (any treatment)				-0.018* (0.010)
Constant	0.141*** (0.004)	0.138*** (0.006)	0.141*** (0.005)	0.135*** (0.007)
Observations	8,046	8,046	8,046	8,046

*Notes: The linear probability models above are predicting decoy choices, with an adjusted treatment indicator independent variable; i.e., the treatment assignment is simplified to just one binary variable, coded as 1 if the participant is assigned to any treatment condition (T1, T2, T3), and 0 if they are assigned to the control group. Additionally, specification (4) combines FI and MH into an "any vulnerability" indicator. Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B9. Further heterogeneity analyses with alternative definitions of vulnerability as predictors of correct choice

	(1) Correct choice, mhi52	(2) Correct choice, diag	(3) Correct choice, disab	(4) Correct choice, paymentdiff	(5) Correct choice, benefits	(6) Correct choice, benefits_unemp
T1	-0.015 (0.010)	-0.014 (0.009)	-0.014 (0.009)	-0.015* (0.009)	-0.016 (0.010)	-0.021** (0.009)
T2	-0.024** (0.010)	-0.026*** (0.009)	-0.026*** (0.009)	-0.023** (0.009)	-0.028*** (0.010)	-0.027*** (0.009)
T3	-0.004 (0.010)	-0.005 (0.009)	-0.007 (0.009)	-0.008 (0.009)	-0.014 (0.010)	-0.014 (0.009)
mhi52	-0.028** (0.013)					
T1 × mhi52	-0.012 (0.018)					
T2 × mhi52	-0.002 (0.018)					
T3 × mhi52	-0.013 (0.018)					
gym_task	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)	0.232*** (0.006)
hotel_task	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)	0.352*** (0.007)
order2	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)
order3	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)	0.008 (0.007)
diag		-0.012 (0.015)				
T1 × diag		-0.020 (0.021)				
T2 × diag		0.010 (0.021)				
T3 × diag		-0.011 (0.021)				
disab			-0.012 (0.015)			
T1 × disab			-0.019 (0.021)			
T2 × disab			0.009 (0.021)			
T3 × disab			-0.003 (0.021)			

paymentdiff				-0.091*** (0.015)		
T1 × paymentdiff				-0.021 (0.021)		
T2 × paymentdiff				-0.006 (0.020)		
T3 × paymentdiff				0.007 (0.021)		
benefits					-0.086*** (0.013)	
T1 × benefits					0.006 (0.018)	
T2 × benefits					0.023 (0.018)	
T3 × benefits					0.030 (0.018)	
benefits_unemp						-0.048*** (0.015)
T1 × benefits_unemp						0.019 (0.022)
T2 × benefits_unemp						0.020 (0.022)
T3 × benefits_unemp						0.041* (0.022)
Constant	0.121*** (0.008)	0.115*** (0.008)	0.114*** (0.008)	0.129*** (0.008)	0.135*** (0.008)	0.120*** (0.008)
Observations	24,138	24,138	24,138	24,138	24,138	24,138

*Notes: The heterogeneity analyses for predicting correct choice are rerun using alternative definitions for mental health and financial insecurity. (1) Severe MHI-5 threshold (mhi52): vulnerable if MHI-5 score is ≤ 52; (2) Formal diagnosis (diag): vulnerable if the participant writes down a formal mental health diagnosis, for free text diagnoses, we only include those that correspond to diagnoses mentioned by a recognised manual of mental health illness - i.e., DSM-5; (3) Mental health condition classified as a disability (disab): vulnerable if the participant's experience of mental health condition corresponds to a definition of disability; (4) Payment difficulties (paymentdiff): vulnerable if the participant reports late payments in at least 3 of the past 6 months; (5) Receipt of benefits (benefits): vulnerable if the participant is currently receiving any government benefits; (6) Not employed and receiving benefits (benefits_unemp): vulnerable if the participant is not in full-time employment and is currently receiving government benefits. Notes: Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B10. Further heterogeneity analyses with alternative definitions of vulnerability as predictors of decoy choice

	(1) Decoy choice, mhi52	(2) Decoy choice, diag	(3) Decoy choice, disab	(4) Decoy choice, paymentdiff	(5) Decoy choice, benefits	(6) Decoy choice, benefits_unemp
T1	0.013* (0.008)	0.020*** (0.007)	0.021*** (0.007)	0.015** (0.007)	0.015** (0.007)	0.018*** (0.007)
T2	0.029*** (0.008)	0.032*** (0.007)	0.032*** (0.007)	0.028*** (0.007)	0.025*** (0.008)	0.029*** (0.007)
T3	0.015** (0.007)	0.017** (0.007)	0.017** (0.007)	0.011* (0.007)	0.015** (0.007)	0.018*** (0.007)
mhi52	0.000 (0.009)					
T1 × mhi52	0.008 (0.013)					
T2 × mhi52	0.009 (0.014)					
T3 × mhi52	-0.009 (0.014)					
gym_task	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)	0.193*** (0.006)
hotel_task	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)	-0.046*** (0.004)
order2	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)
order3	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)
diag		0.014 (0.011)				
T1 × diag		-0.022 (0.016)				
T2 × diag		-0.002 (0.016)				
T3 × diag		-0.022 (0.016)				
disab			0.014 (0.011)			
T1 × disab			-0.027* (0.016)			
T2 × disab			-0.005 (0.016)			
T3 × disab			-0.022 (0.016)			

payment				-0.017 (0.011)		
T1 × payment				0.004 (0.016)		
T2 × payment				0.019 (0.017)		
T3 × payment				0.006 (0.016)		
benefits					-0.016* (0.010)	
T1 × benefits					0.004 (0.014)	
T2 × benefits					0.023 (0.014)	
T3 × benefits					-0.008 (0.014)	
benefits_unemp						0.000 (0.012)
T1 × benefits_unemp						-0.016 (0.017)
T2 × benefits_unemp						0.018 (0.018)
T3 × benefits_unemp						-0.035** (0.017)
Constant	0.086*** (0.007)	0.083*** (0.006)	0.083*** (0.006)	0.089*** (0.006)	0.090*** (0.007)	0.086*** (0.006)
Observations	24,138	24,138	24,138	24,138	24,138	24,138

*Notes: The heterogeneity analyses for predicting decoy choice are rerun using alternative definitions for mental health and financial insecurity. (1) Severe MHI-5 threshold (mhi52): vulnerable if MHI-5 score is ≤ 52; (2) Formal diagnosis (diag): vulnerable if the participant writes down a formal mental health diagnosis, for free text diagnoses, we only include those that correspond to diagnoses mentioned by a recognised manual of mental health illness - i.e., DSM-5; (3) Mental health condition classified as a disability (disab): vulnerable if the participant's experience of mental health condition corresponds to a definition of disability; (4) Payment difficulties (paymentdiff): vulnerable if the participant reports late payments in at least 3 of the past 6 months; (5) Receipt of benefits (benefits): vulnerable if the participant is currently receiving any government benefits; (6) Not employed and receiving benefits (benefits_unemp): vulnerable if the participant is not in full-time employment and is currently receiving government benefits. Notes: Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B11. Correct choice by marketplace

	(1) Correct choice, main model	(2) Correct choice, Broadband only	(3) Correct choice, Gym only	(4) Correct choice, Hotel only

T1	-0.018** (0.008)	-0.017* (0.009)	-0.024 (0.015)	-0.015 (0.016)
T2	-0.024*** (0.008)	-0.024*** (0.009)	-0.040*** (0.015)	-0.009 (0.016)
T3	-0.007 (0.008)	0.057*** (0.011)	-0.054*** (0.015)	-0.024 (0.016)
gym_task	0.232*** (0.006)			
hotel_task	0.352*** (0.007)			
order2	0.005 (0.007)			
order3	0.008 (0.007)			
Constant	0.112*** (0.007)	0.100*** (0.007)	0.365*** (0.011)	0.468*** (0.011)
Observations	24,138	8,046	8,046	8,046

*Notes: The left-most model (1) is the main linear probability specification for predicting correct choices, same as in table B1. (2) - (4) are LPM regressions for predicting decoy choices per marketplace, with each regression having the same independent variables (binary variables indicating assignment to treatment groups 1, 2, 3). These alternative specifications enable us to capture differential treatment effects by marketplace. Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B12. Decoy choice by marketplace

	(1) Decoy choice, main model	(2) Decoy choice, Broadband only	(3) Decoy choice, Gym only	(4) Decoy choice, Hotel only
T1	0.016** (0.006)	0.019** (0.010)	0.022 (0.014)	0.005 (0.007)
T2	0.032*** (0.006)	0.036*** (0.010)	0.046*** (0.014)	0.013* (0.007)
T3	0.013** (0.006)	-0.020** (0.009)	0.047*** (0.014)	0.011 (0.007)
gym_task	0.193*** (0.006)			
hotel_task	-0.046*** (0.004)			
order2	0.005 (0.005)			
order3	0.013** (0.005)			
Constant	0.086*** (0.006)	0.098*** (0.007)	0.271*** (0.010)	0.053*** (0.005)
Observations	24,138	8,046	8,046	8,046

*Notes: The left-most model (1) is the main linear probability specification for predicting decoy choices, same as in table B1. (2) - (4) are LPM regressions for predicting decoy choices per marketplace, with each regression having the same independent variables (binary variables indicating assignment to treatment groups 1, 2, 3). These alternative specifications enable us to capture differential treatment effects by marketplace. Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B13. Main LPM and logit specifications for predicting shopping around behaviours

	(1) Click on banner, LPM	(2) Click on banner, logit	(3) Purchase, LPM	(4) Purchase, logit
T1	-0.023** (0.012)	-0.119** (0.059)	-0.008 (0.008)	-0.061 (0.061)
T2	0.005 (0.012)	0.023 (0.059)	-0.004 (0.008)	-0.027 (0.060)
T3	-0.006 (0.012)	-0.029 (0.059)	-0.001 (0.008)	-0.006 (0.061)
gym_task	-0.011** (0.004)	-0.056** (0.023)	-0.057*** (0.005)	-0.423*** (0.036)
hotel_task	0.034*** (0.005)	0.171*** (0.025)	-0.015*** (0.005)	-0.099*** (0.035)
order2	0.032*** (0.005)	0.165*** (0.026)	0.013*** (0.005)	0.098*** (0.038)
order3	0.047*** (0.005)	0.238*** (0.026)	0.024*** (0.005)	0.173*** (0.037)
Constant	0.245*** (0.009)	-1.126*** (0.047)	0.181*** (0.007)	-1.518*** (0.052)

*Notes: Specifications (1) and (2) are linear probability and logistic regression models, respectively, both with a binary dependent variable equal to 1 if the given participant clicks a banner to access the alternative marketplace, 0 otherwise. Specifications (3) and (4) are linear probability and logistic regression models, respectively, both with a binary dependent variable equal to 1if the participant selects and purchases a product in the alternative marketplace, 0 otherwise. All models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant are in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B14. Per participant shopping around behaviours

	(1) Click on banner proportion	(2) Purchase proportion
T1	-0.023** (0.012)	-0.008 (0.008)
T2	0.004 (0.012)	-0.004 (0.008)
T3	-0.005 (0.012)	-0.000 (0.008)
Constant	0.278*** (0.008)	0.169*** (0.006)
Observations	8,046	8,046

*Notes: Two separate LPMs estimated at the participant level. For (1), the dependent variable is the proportion of tasks where the participant clicked the banner to visit the competitor's website. For (2), the dependent variable is the proportion of tasks where the participant purchased on the competitor's website and finished their journey for the given marketplace there. The only independent variables are the treatment group indicators (T1, T2, T3, with the control group as the omitted category). Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B15. Robustness checks for clicks on the alternative marketplace banner

	(1) Click on banner, robustness check 1: only the bottom 95% of completion times included	(2) Click on banner, robustness check 2: only those who passed first comprehension test included	(3) Click on banner, robustness check 3: only those who failed at most 1 attention check included
T1	-0.022* (0.012)	-0.026** (0.013)	-0.023** (0.012)
T2	0.002 (0.012)	-0.001 (0.013)	0.005 (0.012)
T3	-0.008 (0.012)	-0.007 (0.014)	-0.006 (0.012)
gym_task	-0.011** (0.005)	-0.009* (0.005)	-0.011** (0.004)
hotel_task	0.035*** (0.005)	0.037*** (0.006)	0.034*** (0.005)
order2	0.035*** (0.005)	0.036*** (0.006)	0.032*** (0.005)
order3	0.052*** (0.005)	0.055*** (0.006)	0.047*** (0.005)
Constant	0.253*** (0.009)	0.267*** (0.010)	0.245*** (0.009)

*Notes: Linear probability models for predicting clicks on the alternative marketplace banner, copying the main specification but with the following amendments: 1) dropping participants with unusually fast completion times (i.e.., below the 5th percentile of total survey completion time), 2) dropping participants who failed the first comprehension check for the instructions but passed the second one (those failing both are already screened out), 3) dropping participants who failed two or more of the three geography trivia attention checks. All 3 models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B16. Robustness checks for purchases in the alternative marketplace

	(1) Purchase, robustness check 1: only the bottom 95% of completion times included	(2) Purchase, robustness check 2: only those who passed first comprehension test included	(1) Purchase, robustness check 3: only those who failed at most 1 attention check included
T1	-0.008 (0.009)	-0.009 (0.010)	-0.008 (0.008)
T2	-0.004 (0.009)	-0.007 (0.010)	-0.004 (0.008)
T3	-0.001 (0.009)	-0.001 (0.010)	-0.001 (0.008)
gym_task	-0.059*** (0.005)	-0.060*** (0.006)	-0.057*** (0.005)
hotel_task	-0.015*** (0.005)	-0.017*** (0.006)	-0.015*** (0.005)
order2	0.015*** (0.005)	0.015** (0.006)	0.013*** (0.005)
order3	0.027*** (0.005)	0.029*** (0.006)	0.024*** (0.005)
Constant	0.186*** (0.008)	0.196*** (0.008)	0.181*** (0.007)

*Notes: Linear probability models for predicting purchases in the alternative marketplace, copying the main specification but with the following amendments: 1) dropping participants with unusually fast completion times (i.e.., below the 5th percentile of total survey completion time), 2) dropping participants who failed the first comprehension check for the instructions but passed the second one (those failing both are already screened out), 3) dropping participants who failed two or more of the three geography trivia attention checks. All 3 models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B17. Further robustness checks, combining all treatments together to predict shopping around

	(1) Click on banner, proportion of marketplace visits	(2) Purchase, proportion of marketplace visits
any_treatment	-0.008 (0.010)	-0.004 (0.007)
gym_task	-0.011** (0.004)	-0.057*** (0.005)
hotel_task	0.035*** (0.005)	-0.014*** (0.005)
order2	0.035*** (0.005)	0.013*** (0.005)
order3	0.046*** (0.005)	0.023*** (0.005)
Constant	0.244*** (0.009)	0.180*** (0.007)
Observations	24,138	24,138

*Notes: Specification (1) is a linear probability model with a binary dependent variable equal to 1 if the given participant clicks a banner to access the alternative marketplace, 0 otherwise. Specification (2) is again a linear probability model with a binary dependent variable equal to 1if the participant selects and purchases a product in the alternative marketplace, 0 otherwise. For both (1) and (2), there is an independent, treatment indicator variable any_treatment; i.e., the treatment assignment is simplified to just one binary variable, coded as 1 if the participant is assigned to any treatment condition (T1, T2, T3), and 0 if they are assigned to the control group. Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B18. Heterogeneity analyses for predicting clicks on banner

	(1) Click on banner, financial insecurity (FI)	(2) Click on banner, Poor mental health (MH)	(3) Click on banner, frequent online shoppers (FSHOP)	(4) Click on banner, stressed shoppers (STR)	(5) Click on banner, bargain hunters (BH)	(6) Click on banner, high financial literacy (LIT)	(7) Click on banner, present bias (PB)
T1	-0.015 (0.013)	-0.030* (0.017)	-0.035 (0.023)	-0.025** (0.012)	-0.023 (0.019)	-0.029* (0.016)	-0.016 (0.013)
T2	0.010 (0.013)	-0.018 (0.017)	-0.013 (0.023)	0.002 (0.012)	0.000 (0.019)	-0.005 (0.016)	0.012 (0.013)
T3	-0.004 (0.013)	-0.029* (0.017)	-0.004 (0.023)	-0.006 (0.012)	-0.029 (0.019)	0.002 (0.016)	0.003 (0.013)
FI	-0.022 (0.020)						
T1 × FI	-0.035 (0.027)						
T2 × FI	-0.028 (0.027)						
T3 × FI	-0.009 (0.028)						
gym_task	-0.011** (0.004)	-0.011** (0.004)	-0.011** (0.004)	-0.011** (0.004)	-0.011** (0.004)	-0.011** (0.004)	-0.011** (0.004)
hotel_task	0.035*** (0.005)	0.035*** (0.005)	0.035*** (0.005)	0.035*** (0.005)	0.035*** (0.005)	0.035*** (0.005)	0.035*** (0.005)
order2	0.031*** (0.005)	0.031*** (0.005)	0.031*** (0.005)	0.031*** (0.005)	0.031*** (0.005)	0.031*** (0.005)	0.031*** (0.005)
order3	0.046*** (0.005)	0.046*** (0.005)	0.046*** (0.005)	0.046*** (0.005)	0.046*** (0.005)	0.046*** (0.005)	0.046*** (0.005)
MH		-0.032* (0.017)					
T1 × MH		0.014 (0.023)					
T2 × MH		0.042* (0.024)					
T3 × MH		0.045* (0.024)					
FSHOP			0.007 (0.019)				
T1 × FSHOP			0.016 (0.027)				
T2 × FSHOP			0.023 (0.027)				
T3 × FSHOP			-0.002 (0.027)				

STR				-0.129*** (0.034)			
T1 × STR				0.069 (0.052)			
T2 × STR				0.032 (0.057)			
T3 × STR				0.002 (0.052)			
BH					0.044** (0.017)		
T1 × BH					0.003 (0.024)		
T2 × BH					0.007 (0.024)		
T3 × BH					0.039 (0.024)		
LIT						0.120*** (0.016)	
T1 × LIT						0.012 (0.022)	
T2 × LIT						0.014 (0.023)	
T3 × LIT						-0.010 (0.023)	
PB							0.018 (0.020)
T1 × PB							-0.032 (0.028)
T2 × PB							-0.036 (0.028)
T3 × PB							-0.036 (0.028)
Constant	0.249*** (0.010)	0.261*** (0.013)	0.239*** (0.017)	0.248*** (0.009)	0.215*** (0.014)	0.175*** (0.012)	0.240*** (0.010)

*Notes: Main linear probability model specifications for predicting clicks on the alternative marketplace banner, extended to include indicator variables for vulnerability and personal characteristics, and their interactions with treatment indicators (T1–T3). Financial insecurity (FI) is defined as income < £15,000; poor mental health (MH) is defined by MHI-5 score ≤ 68; frequent online shopping (FSHOP) is conceptualised as shopping online either about once or twice per week or about once or twice a month; stressed shoppers (STR) are defined as those who ticked in the survey that they find online shopping either extremely or very stressful; bargain hunters (BH) are characterised by their stated preference to either always or frequently look for discounts; high financial literacy (LIT) is operationalised as answering all 3 financial literacy questions in the survey correctly; present bias (PB) is defined by a preference reversal in the survey whereby participants prefer receiving £100 now, rather than £110 in 1 week but also receiving £110 in 5 weeks, rather than £100 in 4 weeks. All models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B19. Heterogeneity analyses for predicting purchases in the alternative marketplace

	(1) Purchase, financial insecurity (FI)	(2) Purchase, poor mental health (MH)	(3) Purchase, frequent online shoppers (FSHOP)	(4) Purchase, stressed shoppers (STR)	(5) Purchase, bargain hunters (BH)	(6) Purchase, high financial literacy (LIT)	(7) Purchase, present bias (PB)
T1	-0.002 (0.009)	-0.011 (0.012)	-0.024 (0.016)	-0.009 (0.008)	-0.007 (0.013)	-0.013 (0.012)	-0.006 (0.009)
T2	-0.000 (0.009)	-0.016 (0.012)	-0.032* (0.016)	-0.005 (0.008)	-0.003 (0.013)	-0.004 (0.012)	0.000 (0.009)
T3	0.003 (0.010)	-0.015 (0.012)	-0.009 (0.017)	-0.000 (0.009)	-0.022* (0.013)	0.003 (0.012)	0.004 (0.010)
FI	-0.005 (0.014)						
T1 × FI	-0.025 (0.020)						
T2 × FI	-0.018 (0.020)						
T3 × FI	-0.017 (0.020)						
gym_task	-0.057*** (0.005)	-0.057*** (0.005)	-0.057*** (0.005)	-0.057*** (0.005)	-0.057*** (0.005)	-0.057*** (0.005)	-0.057*** (0.005)
hotel_task	-0.014*** (0.005)	-0.014*** (0.005)	-0.014*** (0.005)	-0.014*** (0.005)	-0.014*** (0.005)	-0.014*** (0.005)	-0.014*** (0.005)
order2	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)	0.013** (0.005)
order3	0.023*** (0.005)	0.023*** (0.005)	0.023*** (0.005)	0.023*** (0.005)	0.023*** (0.005)	0.023*** (0.005)	0.023*** (0.005)
MH		-0.022* (0.012)					
T1 × MH		0.005 (0.017)					
T2 × MH		0.022 (0.017)					
T3 × MH		0.027 (0.017)					
FSHOP			-0.006 (0.014)				
T1 × FSHOP			0.021 (0.019)				
T2 × FSHOP			0.037* (0.019)				
T3 × FSHOP			0.012 (0.019)				
STR				-0.056* (0.028)			
T1 × STR				0.020 (0.040)			

T2 × STR				0.036 (0.046)			
T3 × STR				-0.023 (0.043)			
BH					0.029** (0.012)		
T1 × BH					-0.000 (0.017)		
T2 × BH					-0.001 (0.017)		
T3 × BH					0.034** (0.017)		
LIT						0.067*** (0.012)	
T1 × LIT						0.009 (0.016)	
T2 × LIT						-0.002 (0.016)	
T3 × LIT						-0.004 (0.017)	
PB							0.008 (0.014)
T1 × PB							-0.008 (0.020)
T2 × PB							-0.020 (0.020)
T3 × PB							-0.018 (0.020)
Constant	0.182*** (0.008)	0.192*** (0.010)	0.185*** (0.013)	0.182*** (0.007)	0.162*** (0.011)	0.142*** (0.009)	0.179*** (0.008)

*Notes: Main linear probability model specifications for predicting purchases in the alternative marketplace, extended to include indicator variables for vulnerability and personal characteristics, and their interactions with treatment indicators (T1–T3). Financial insecurity (FI) is defined as income < £15,000; poor mental health (MH) is defined by MHI-5 score ≤ 68; frequent online shopping (FSHOP) is conceptualised as shopping online either about once or twice per week or about once or twice a month; stressed shoppers (STR) are defined as those who ticked in the survey that they find online shopping either extremely or very stressful; bargain hunters (BH) are characterised by their stated preference to either always or frequently look for discounts; high financial literacy (LIT) is operationalised as answering all 3 financial literacy questions in the survey correctly; present bias (PB) is defined by a preference reversal in the survey whereby participants prefer receiving £100 now, rather than £110 in 1 week but also receiving £110 in 5 weeks, rather than £100 in 4 weeks. All models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B20. Clicks on alternative marketplace banner by market

	(1) Click on banner, main model	(2) Click on banner, Broadband only	(3) Click on banner, Gym only	(4) Click on banner, Hotel only
T1	-0.023** (0.012)	-0.019 (0.014)	-0.022 (0.014)	-0.028** (0.014)
T2	0.004 (0.012)	-0.002 (0.014)	0.001 (0.014)	0.013 (0.014)
T3	-0.005 (0.012)	-0.017 (0.014)	-0.013 (0.014)	0.015 (0.015)
gym_task	-0.011** (0.004)			
hotel_task	0.035*** (0.005)			
order2	0.031*** (0.005)			
order3	0.046*** (0.005)			
Constant	0.244*** (0.009)	0.274*** (0.010)	0.261*** (0.010)	0.298*** (0.010)
Observations	24,138	8,046	8,046	8,046

*Notes: The left-most model (1) is the main linear probability specification for predicting clicks on the banner to visit the alternative site, same as in table B13. (2) - (4) are LPM regressions for predicting clicks separately per marketplace, with each regression having the same independent variables (binary variables indicating assignment to treatment groups 1, 2, 3). These alternative specifications enable us to capture differential treatment effects by marketplace. Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B21. Purchases in the alternative marketplace by market

	(1) Purchase, main model	(2) Purchase, Broadband only	(3) Purchase, Gym only	(4) Purchase, Hotel only
T1	-0.008 (0.008)	-0.009 (0.012)	-0.013 (0.011)	-0.002 (0.012)
T2	-0.004 (0.008)	-0.004 (0.012)	-0.010 (0.011)	0.001 (0.012)
T3	-0.000 (0.008)	-0.013 (0.012)	-0.003 (0.011)	0.015 (0.012)
gym_task	-0.057*** (0.005)			
hotel_task	-0.014*** (0.005)			
order2	0.013** (0.005)			
order3	0.023*** (0.005)			
Constant	0.180*** (0.007)	0.196*** (0.009)	0.139*** (0.008)	0.171*** (0.008)
Observations	24,138	8,046	8,046	8,046

*Notes: The left-most model (1) is the main linear probability specification for predicting purchases in the alternative marketplace, same as in table B13. (2) - (4) are LPM regressions for predicting purchases separately per marketplace, with each regression having the same independent variables (binary variables indicating assignment to treatment groups 1, 2, 3). These alternative specifications enable us to capture differential treatment effects by marketplace. Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B22. Time needed to finish the shopping tasks, as a function of experimental conditions

	(1) total time, all participants	(2) broadband task time, all participants	(3) gym task time, all participants	(4) hotel booking task time, all participants	(5) total time, poor mental health	(6) broadband task time, poor mental health	(7) gym task time, poor mental health	(8) hotel booking task time, poor mental health	(9) total time, financial insecurity	(10) broadband task time, financial insecurity	(11) gym task time, financial insecurity	(12) hotel booking task time, financial insecurity
T1	2.111 (3.564)	2.185 (2.007)	0.807 (2.355)	-0.881 (1.755)	-3.187 (3.267)	0.575* (0.329)	-2.650 (2.438)	-1.113 (2.124)	-2.056 (1.605)	-0.153 (0.215)	-0.133 (0.430)	-1.770 (1.484)
T2	-2.942 (1.959)	0.722 (0.585)	-1.570 (1.290)	-2.093 (1.334)	-4.226 (3.123)	0.953 (0.705)	-2.774 (2.426)	-2.405 (1.813)	-0.593 (1.675)	0.295 (0.293)	0.057 (0.449)	-0.945 (1.528)
T3	-0.824 (2.185)	1.072* (0.629)	-0.059 (1.555)	-1.837 (1.378)	-1.961 (3.339)	1.797* (1.078)	-2.434 (2.466)	-1.324 (1.952)	-0.461 (1.805)	0.021 (0.221)	0.794 (0.831)	-1.276 (1.512)
Constant	15.751*** (1.840)	3.638*** (0.146)	5.863*** (1.270)	6.250*** (1.311)	15.590*** (3.027)	3.142*** (0.146)	6.677*** (2.416)	5.771*** (1.806)	12.139*** (1.548)	3.197*** (0.139)	3.967*** (0.365)	4.975*** (1.474)
Observations	8,046	8,046	8,046	8,046	4,259	4,259	4,259	4,259	1,779	1,779	1,779	1,779

*Notes: OLS linear regression models with treatment indicators (T1, T2, T3) as independent variables and time spent on the tasks, in minutes, as the dependent variable. Specification (1) measures the total time across all tasks, (2), (3), (4) for each marketplace task individually. (5) - (12) assess the time needed to complete the tasks specifically for vulnerable groups. Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B23. Mediator regressions

	Reported task stress			Post-experiment mood			Decision confidence			Perceived value for money				Noticed reference pricing	
	(1) all partici.	(2) poor m. health	(3) fin. insec.	(4) all partic.	(5) poor m. health	(6) fin. insec.	(7) all partic.	(8) poor m. health	(9) fin. insec.	(10) all partic.	(11) all partic. (spec. 2)	(12) poor m. health	(13) fin. insc.	(14) all partic. logit	(15) all partic. LPM
prop_correct	0.090	0.226*	0.125	-0.363***	-0.401***	-0.329*	-0.542***	-0.637***	-0.426**	-0.197***		-0.232**	-0.244		
	(0.080)	(0.117)	(0.190)	(0.075)	(0.106)	(0.174)	(0.088)	(0.129)	(0.202)	(0.069)		(0.100)	(0.166)		
T1	0.022	0.032	-0.029	0.015	0.008	-0.086	0.028	-0.004	0.028	-0.004	0.034	-0.045	-0.033	0.138**	0.028**
	(0.046)	(0.063)	(0.094)	(0.044)	(0.058)	(0.084)	(0.052)	(0.071)	(0.103)	(0.041)	(0.026)	(0.056)	(0.082)	(0.070)	(0.014)
T2	0.012	0.111*	-0.017	-0.026	-0.030	0.037	0.028	-0.024	0.099	-0.036	0.037	-0.121**	-0.097	0.185***	0.038***
	(0.046)	(0.064)	(0.093)	(0.043)	(0.057)	(0.082)	(0.051)	(0.069)	(0.097)	(0.040)	(0.026)	(0.055)	(0.082)	(0.069)	(0.014)
T3	0.023	0.035	0.073	-0.021	-0.000	-0.072	0.001	0.016	0.025	-0.080*	0.021	-0.105*	-0.093	0.135*	0.028*
	(0.048)	(0.066)	(0.103)	(0.044)	(0.059)	(0.089)	(0.054)	(0.072)	(0.109)	(0.043)	(0.026)	(0.058)	(0.089)	(0.070)	(0.014)
T1 x prop_correct	-0.103	-0.161	-0.109	0.015	0.111	0.242	0.058	0.220	0.041	0.117		0.287**	0.259		
	(0.112)	(0.161)	(0.257)	(0.105)	(0.147)	(0.233)	(0.125)	(0.180)	(0.273)	(0.096)		(0.139)	(0.220)		
T2 x prop_correct	-0.132	-0.331**	-0.053	0.104	0.154	0.259	0.093	0.240	0.149	0.236**		0.462***	0.639***		
	(0.111)	(0.159)	(0.260)	(0.104)	(0.144)	(0.242)	(0.124)	(0.173)	(0.270)	(0.096)		(0.135)	(0.222)		
T3 x prop_correct	-0.142	-0.140	-0.463*	0.193*	0.239	0.573**	0.195	0.179	0.395	0.327***		0.416***	0.616***		
	(0.114)	(0.165)	(0.266)	(0.107)	(0.147)	(0.241)	(0.128)	(0.180)	(0.289)	(0.100)		(0.144)	(0.229)		
Constant	2.041***	2.238***	2.210***	3.679***	3.511***	3.520***	3.044***	2.894***	2.820***	3.748***	3.687***	3.623***	3.601***	-0.982***	0.272***
	(0.033)	(0.046)	(0.069)	(0.031)	(0.042)	(0.060)	(0.038)	(0.052)	(0.074)	(0.029)	(0.018)	(0.040)	(0.060)	(0.050)	(0.010)
Observations	8,046	4,259	1,779	8,046	4,259	1,779	8,046	4,259	1,779	8,046	8,046	4,259	1,779	8,046	8,046

*Notes: OLS linear regression models exploring potential mechanisms through which the online choice architecture might operate. The independent variables count proportion of correct responses, treatment indicators (T1, T2, T3) and their interactions. Specifications (1) - (3) have reported task stress as the ordinal dependent variable. It is measured on a 5-point likert scale, with 5 being extreme stress. Specifications (4) - (6) have the post-experimental mood as the dependent variable. It is again measured on a 5-point likert scale, with 5 being very positive. Specifications (7) - (9) have the decision confidence in choices across tasks as the dependent variable. It is again measured on a 5-point likert scale with 5 meaning extremely confident. Specifications (10) - (13) have an ordinal dependent variable - perceived value for money. It is measured on a 5-point likert scale with 1 being definitely not good value for money and 5 being definitely a good value for money. Specifications (14) - (15) measure the noticing of reference pricing. This variable is equal to 1 if the participant reported noticing reference pricing. Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B24. Selecting too cheap and too expensive options as a function of experimental conditions

	(1) Too cheap choice	(2) Too cheap choice	(3) Too expensive choice	(4) Too expensive choice
T1	-0.015 (0.009)		0.023*** (0.009)	
T2	-0.013 (0.009)		0.025*** (0.009)	
T3	-0.027*** (0.009)		0.036*** (0.009)	
gym_task	0.100*** (0.005)	0.100*** (0.005)	-0.238*** (0.007)	-0.238*** (0.007)
hotel_task	0.256*** (0.006)	0.256*** (0.006)	-0.575*** (0.006)	-0.575*** (0.006)
order2	-0.020*** (0.006)	-0.020*** (0.006)	0.015** (0.006)	0.015** (0.006)
order3	-0.032*** (0.006)	-0.032*** (0.006)	0.019*** (0.006)	0.019*** (0.006)
treat_any		-0.018** (0.007)		0.028*** (0.007)
Constant	0.151*** (0.008)	0.151*** (0.008)	0.666*** (0.008)	0.666*** (0.008)
Observations	24,138	24,138	24,138	24,138

*Notes: Main linear probability model specifications for (1) - (2) predicting too cheap choices, (3) - (4) predicting too expensive choices. Models (1) and (3) feature separate indicators for each treatment group, while models (2) and (4) group all treatments into a single treat_any indicator, equal to 0 for the control group only. Both models include marketplace dummies (gym, hotel, broadband omitted) and task-order dummies (order2, order3, first task omitted). Standard errors clustered by participant in parentheses. * p<.10, ** p<.05, *** p<.01.*

Table B25. Participant characteristics by survey outcome

Statistic	Completes	Screenouts	Dropouts
Female (%)	52.35	50.46	58.04
Male (%)	47.33	49.14	41.67
Non-binary (%)	0.29	0.30	0.23
Other gender (%)	0.04	0.10	0.06
Primary education (%)	0.32	2.12	0.71
Secondary education (%)	17.38	20.89	23.92
Further education (%)	22.96	23.01	23.32
University education (%)	40.93	35.36	38.78
Postgraduate education (%)	18.42	18.61	13.27
Region East England (%)	8.72	8.18	9.36
Region East Midlands (%)	7.23	7.05	7.36
Region Greater London (%)	13.42	16.66	9.41
Region Northeast England (%)	4.11	6.49	4.58
Region Northwest England (%)	11.12	10.76	11.42
Region Northern Ireland (%)	2.67	2.98	2.11
Region Scotland (%)	8.13	8.54	8.15
Region Southeast England (%)	15.16	11.39	15.85
Region Southwest England (%)	7.94	6.85	9.04
Region Wales (%)	4.61	4.80	4.92
Region West Midlands (%)	8.60	8.64	8.88
Region Yorkshire (%)	8.26	7.65	8.92
Urban area (%)	34.65	41.99	29.31
Suburban area (%)	47.24	41.95	49.10
Rural area (%)	18.03	15.73	21.35
Other area (%)	0.07	0.33	0.24
Mean age (years)	48.02	43.52	54.14
N	8,046	6,318	8,843

Table B26. Dropouts by survey block

Block	Number of dropouts	Percentage
Instructions	926	10.47%
Pre-experimental questionnaire	877	9.92%
Broadband journey	2,478	28.02%
Gym journey	2,305	26.07%
Hotel journey	2,208	24.97%
Post-experimental questionnaire	49	0.55%
Total	8,843	100.00%

Table B27. Dropouts by treatment assignment

	(1) Dropout - LPM	(2) Dropout - Logit
group assignment		
T1	-0.001	-0.003
	(0.009)	(0.038)
T2	0.004	0.015
	(0.009)	(0.038)
T3	0.004	0.019
	(0.009)	(0.038)
Constant	0.379***	-0.493***
	(0.006)	(0.027)
Observations	23,207	23,207

*Notes: In both specifications, the control group assignment serves as a reference category. The observations include all completes (N = 8046), screenouts (N = 6318) and dropouts (N = 8843). Standard errors in parentheses. * p<.1; ** p<.05; *** p<.01.*

Table B28. Balance in demographics across treatment groups

Statistic	control group	T1	T2	T3	(1) vs. (2), p-value	(1) vs. (3), p-value	(1) vs. (4), p-value	(2) vs. (3), p-value	(2) vs. (4), p-value	(3) vs. (4), p-value	p-value from joint orthogonality test of treatment arms
Mean age (years)	48.302	47.671	47.804	48.320	0.259	0.368	0.974	0.809	0.245	0.350	0.534
Female (%)	52.80%	51.30%	52.80%	52.50%	0.361	0.968	0.853	0.337	0.468	0.820	0.755
Male (%)	46.90%	48.50%	46.90%	47.00%	0.297	0.998	0.929	0.292	0.341	0.926	0.666
Non-binary (%)	0.40%	0.10%	0.20%	0.50%	0.089	0.528	0.612	0.261	0.031	0.257	0.178
Other gender (%)	0.00%	0.00%	0.00%	0.10%	0.323	0.325	0.317	0.996	0.985	0.981	0.805
Primary education (%)	0.20%	0.30%	0.40%	0.40%	0.551	0.270	0.245	0.603	0.559	0.946	0.647
Secondary education (%)	17.70%	17.30%	16.50%	18.00%	0.738	0.284	0.801	0.460	0.557	0.186	0.578
Further education (%)	23.40%	21.70%	23.20%	23.60%	0.194	0.927	0.869	0.224	0.144	0.797	0.445
University education (%)	39.50%	42.00%	42.50%	39.60%	0.114	0.053	0.964	0.722	0.125	0.059	0.104
Postgraduate education (%)	19.20%	18.80%	17.40%	18.40%	0.726	0.139	0.519	0.256	0.766	0.407	0.496
Region East England (%)	9.10%	7.60%	9.20%	9.00%	0.089	0.942	0.892	0.074	0.118	0.834	0.245

Region East Midlands (%)	7.70%	7.90%	7.00%	6.40%	0.841	0.380	0.099	0.278	0.063	0.432	0.229
Region Greater London (%)	12.80%	13.60%	12.50%	14.80%	0.458	0.782	0.060	0.305	0.251	0.030	0.128
Region Northeast England (%)	4.50%	4.20%	3.30%	4.40%	0.643	0.047	0.834	0.127	0.801	0.077	0.202
Region Northwest England (%)	11.70%	11.00%	10.70%	11.10%	0.511	0.344	0.538	0.773	0.970	0.746	0.812
Region Northern Ireland (%)	2.50%	2.40%	2.80%	3.10%	0.828	0.531	0.240	0.396	0.162	0.576	0.486
Region Scotland (%)	7.20%	9.50%	8.30%	7.50%	0.007	0.171	0.651	0.179	0.025	0.362	0.034
Region Southeast England (%)	13.70%	15.50%	15.80%	15.60%	0.115	0.070	0.098	0.813	0.928	0.885	0.244
Region Southwest England (%)	8.50%	7.90%	8.00%	7.40%	0.463	0.525	0.207	0.920	0.592	0.525	0.656
Region Wales (%)	4.70%	4.50%	5.00%	4.20%	0.757	0.715	0.409	0.497	0.603	0.232	0.676
Region West Midlands (%)	8.50%	8.00%	9.20%	8.70%	0.574	0.454	0.842	0.187	0.446	0.584	0.619
Region Yorkshire (%)	9.10%	7.90%	8.30%	7.80%	0.176	0.390	0.162	0.617	0.957	0.582	0.461
Urban area (%)	34.10%	34.50%	33.30%	36.70%	0.785	0.620	0.084	0.439	0.142	0.025	0.135

Suburban area (%)	47.60%	47.40%	47.30%	46.60%	0.903	0.815	0.521	0.911	0.601	0.680	0.926
Rural area (%)	18.10%	18.10%	19.30%	16.60%	0.957	0.325	0.194	0.296	0.211	0.022	0.154
Other area (%)	0.20%	0.00%	0.00%	0.10%	0.080	0.302	0.657	0.319	0.152	0.544	0.328
Sample size (N)	1,986	2,032	2,048	1,980							

Table B29. Average underspending vs. average overspending by marketplace

	Broadband plans				Gym memberships				Hotel bookings			
	Average overpaid	Average underpaid	Average signed error	Average absolute error	Average overpaid	Average underpaid	Average signed error	Average absolute error	Average overpaid	Average underpaid	Average signed error	Average absolute error
T1	10.228***	-0.865	11.093***	9.364***	1.827***	-0.489	2.316**	1.339**	2.896*	-0.100	2.996	2.796
	(2.994)	(0.649)	(3.305)	(2.801)	(0.585)	(0.499)	(0.902)	(0.607)	(1.610)	(1.118)	(2.173)	(1.721)
T2	12.710***	-0.330	13.039***	12.380***	1.979***	-0.739	2.717***	1.240**	3.568**	0.061	3.507	3.629**
	(3.015)	(0.658)	(3.336)	(2.813)	(0.585)	(0.486)	(0.893)	(0.599)	(1.619)	(1.126)	(2.187)	(1.730)
T3	9.449***	-4.117***	13.566***	5.332*	2.032***	-0.597	2.628***	1.435**	3.603**	0.491	3.112	4.022**
	(3.005)	(0.590)	(3.256)	(2.856)	(0.591)	(0.490)	(0.903)	(0.603)	(1.635)	(1.135)	(2.210)	(1.740)
Constant	91.504***	8.460***	83.044***	99.963***	15.484***	7.100***	8.384***	22.584***	18.483***	22.247***	-3.764**	40.731***
	(2.095)	(0.472)	(2.322)	(1.958)	(0.413)	(0.356)	(0.639)	(0.432)	(1.085)	(0.796)	(1.492)	(1.182)
Observations	8046	8046	8046	8046	8046	8046	8046	8046	8046	8046	8046	8046

*Notes: These outcomes measure deviations from each market's "correct" cost. For each participant and marketplace (main vs. alternative), we compute the difference between amount paid and the price of the correct offer. Average overpaid sums positive differences, Average underpaid sums the absolute value of negative differences, both in £. Average signed error is the mean of the raw differences across the choices made (positive = net overpay, negative = net underpay), while Average absolute error is the mean of |diff| and captures the size of the mistake regardless of direction. In the OLS rows, T1–T3 coefficients are differences relative to the control. Thus, a negative coefficient in "Average underpaid" means the treatment reduced underpayment (participants were less below the correct choice price, moving towards the correct choice), whereas positive coefficients in "Overpaid" or "Avg signed error" indicate greater net overpaying. Standard errors in parentheses. * p<.10, ** p<.05, *** p<.01.*

Table B30. Proportion of correct responses with/without confidence as a predictor

	(1) Proportion of correct responses	(2) Proportion of correct responses
T1	-0.018** (0.008)	-0.017** (0.008)
T2	-0.024*** (0.008)	-0.022*** (0.008)
T3	-0.007 (0.008)	-0.005 (0.008)
Confidence (numeric)		-0.029*** (0.003)
Constant	0.311*** (0.006)	0.395*** (0.010)
Observations	8,046	8,046

*Notes: Model (1) is an OLS of decision quality (proportion of correct choices) on treatment indicators T1–T3 (control is the reference), with robust standard errors. Model (2) adds a continuous confidence measure as a control. We estimated Model (2) to test whether confidence mediates the treatment effect. The treatment coefficients change only slightly when controlling for confidence, by about 8.7% for T1 and 8.2% for T2, while T3 remains statistically insignificant in both, suggesting confidence is unlikely to be the main mediator. Standard errors in parentheses. * p<.10; ** p<.05; *** p<.01.*

7.3 Appendix C - Data Validity Checks

7.3.1. External Validity

We evaluated the generalizability of our results using the SANS framework (Selection, Attrition, Naturalness, and Scaling) proposed by List (2020):

Selection: Our final sample of 8,046 UK adults was nationally representative in terms of UK regions, gender and age. By hosting the experiment on an online survey platform (Qualtrics), we deliberately excluded non-internet users who are not exposed to online reference pricing in real life, hence improving the relevance to our target population. We note, however, that many respondents did not reach completion: 6,318 individuals were screened out for failing both comprehension checks, or stating less than 18 years of age, and further 8,843 dropped out during the survey (more detailed statistics in Appendix Table B25).

Attrition: To minimize attrition, we pilot-tested and refined the study design before full experimental launch. We simplified tasks (e.g. removing an initial requirement to enter a pre-set postcode in the broadband and gym tasks), and added task instructions to the top of each marketplace page (e.g., broadband instructions on top of each broadband page), all aimed at reducing confusion that could lead to dropout. Nevertheless, attrition in the experiment was non-trivial. Appendix Table B26 details the dropout by experimental stage: roughly 10% of participants dropped during the instructions, about 10% in the pre-experimental questionnaire, and the majority during the marketplace tasks (about 28% of all starters quit in the broadband-shopping task, 26% in the gym-membership task, and 25% in the hotel-booking task). Only 0.55% left during the post-experimental questionnaire. The high proportion of attrition in the marketplaces suggests that some respondents found the shopping tasks taxing or confusing. Importantly, however, because attrition was distributed equally across conditions (see internal validity below), it most likely did not introduce bias in treatment effects.

Naturalness: We designed the experimental tasks to closely mirror real online marketplaces. Participants used their own devices to navigate through the marketplace replicas and these replicas featured responsive web design (i.e., optimised screen sizes). Furthermore, participants were able to browse and select varied offers, and add-on products and services, and they went through the payment and confirmation processes in each market. Only three compromises were made for experimental control and practicality. For one, to keep the survey length to 15 minutes, we introduced a clickable banner at the top of the main marketplace sites for accessing a competitor's sites instead of letting participants navigate to the competitor via a search engine. Secondly, participants shopped for hypothetical needs. We provided scenarios, which might not perfectly match participants' own needs. Finally, participants were required to choose an offer in each marketplace. They could not leave a purchase or defer the decision, which might happen in real life. These design choices may reduce naturalness to some extent, but were necessary to - within agreed survey time - ensure well-defined optimal and decoy choices, to be able to measure harm of online choice architecture and its effect on shopping around. Although we acknowledge that the experimental setting isn't identical to a real-life shopping session, we believe the marketplaces were realistic enough that the observed choices reflect real-world online decision-making.

Scaling: The studied online choice architecture (OCA) practices are widespread across online marketplaces, enhancing confidence in our findings' broader applicability. Also, our experiment included three distinct types of offers: a subscription (gym memberships), a service contract (broadband), and a one-time purchase (hotel booking). By covering different offer types and marketplaces, we provide evidence on how stable the effects of online choice architecture are across different settings. We expect that our results can generalize to other online markets where such OCA tactics are present.

7.3.2. Internal Validity

Finally, we explore the internal validity of the experiment. The main threat to the internal validity of the type of experiment that we conducted is differential attrition between experimental groups. We do not find any significant differences in the share that dropped out of the experiment across experimental groups (Appendix Table B27). In both linear probability and logistic models for predicting attrition, the coefficients on each treatment (T1, T2, T3) were not statistically different from the control group. Also, there is baseline balance in demographic characteristics across the treatment groups. Appendix Table B28 presents demographics across groups (control, T1, T2, T3). We see that the groups are virtually identical in terms of age (mean 48 years in each), gender split (cca. 52% female in all groups), education levels, and regional distribution. Only one difference is statistically significant at the conventional 5 % level. That is the share of participants from Scotland. However, this one discrepancy is likely due to chance given the many demographic variables examined. A further point in favour of internal validity is the subgroup representation in the experiment. In our sample of 8046 completers, 1779 participants met the financial insecurity criterion, and 4,259 met the mental health vulnerability criterion. This large number of vulnerable respondents allowed precise estimation of treatment effects within these subgroups.

7.4 Appendix D – Desk Research and User Flows

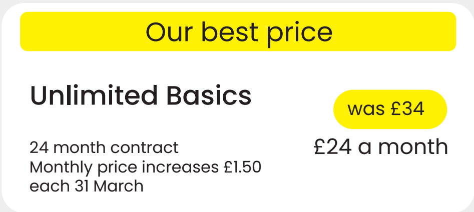
In the following pages, we include two figures. The first figure (Figure D1) consists of a three-page infographic including some illustrative examples of OCA practices found during desk research. The second figure (Figure D2) is a four-page infographic that outlines the user flows and interventions for three different online marketplaces: a telecom company website, a gym company website, and a hotel booking aggregator website.

Figure D1. Desk Research Examples

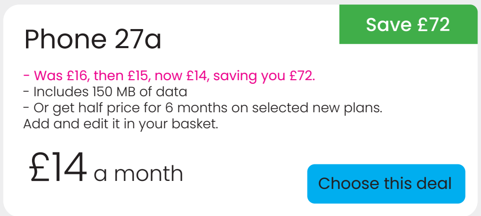
In the following pages, three tables present a series of examples illustrating OCA practices identified during the desk research. The examples reported in these tables are mock-ups of real-world cases observed in online marketplaces. The first table provides illustrative instances of reference pricing practices, the second focuses on sensory manipulation strategies, and the third documents the use of time-bound elements.

Reference Pricing Examples

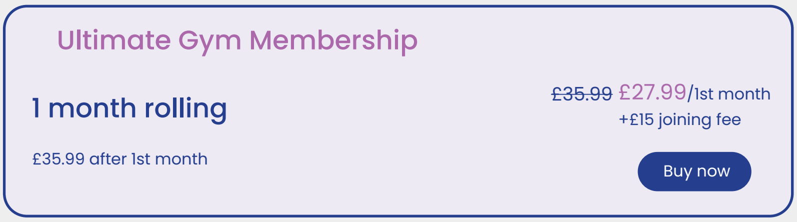
1



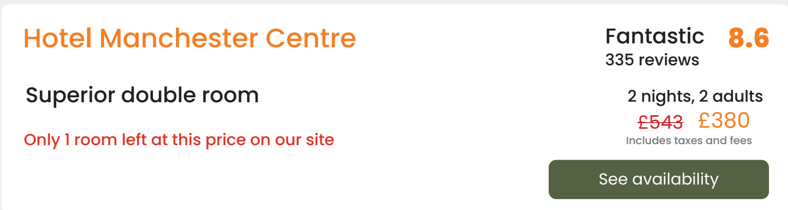
2



3



4



Desk research findings revealed that reference pricing was presented in various formats, as illustrated in the examples above.

In the first example, a phone plan deal called "Unlimited basics" included reference pricing (e.g., current and previous monthly rate) and featured salient yellow labels.

In the second example, the previous price was presented in bright pink text. A time-bound discount was also featured, offering customers a half-price deal for 6 months. Green salient labels in the top right corner highlighted the amount customers could save if they purchased the plan.

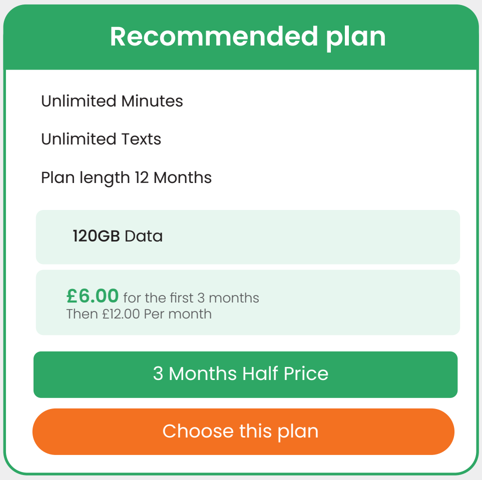
The third example illustrates a gym membership deal that incorporated reference pricing, along with a time-bound element.

In the fourth example, an accommodation offer featured reference pricing with the original price shown in red and the discounted price highlighted in orange. Additionally, scarcity claims such as "Only 1 room left at this price" were used. The offer also used rating, highlighting the number of reviews and average rating at the top right corner.

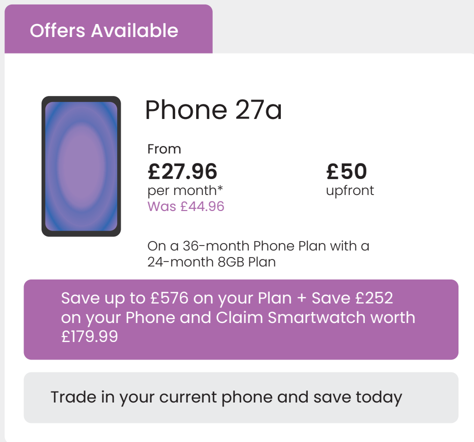
Figure D1. Desk Research Examples

Sensory Manipulation Examples

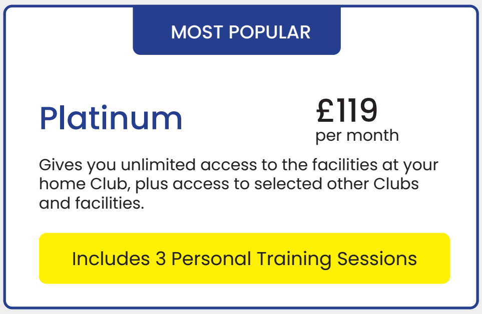
1



2



3



4

During our desk research, we found that sensory manipulation is a widespread OCA practice among choice architects. In the first example, a mobile phone plan was highlighted as a recommended option by using a green label and frame. Additionally, there was a time-limited discount that offered the first three months of a 12-month contract at half price. All plans with time-bound discounts, like the one mentioned, were prominently displayed at the top of the page with eye-catching elements.

In the second example, a discounted phone plan was highlighted with a coloured label that stated “Offers available”. The choice architect employed reference pricing, with the referenced price highlighted in a different colour. A bundle deal was also offered and showcased in a coloured box.

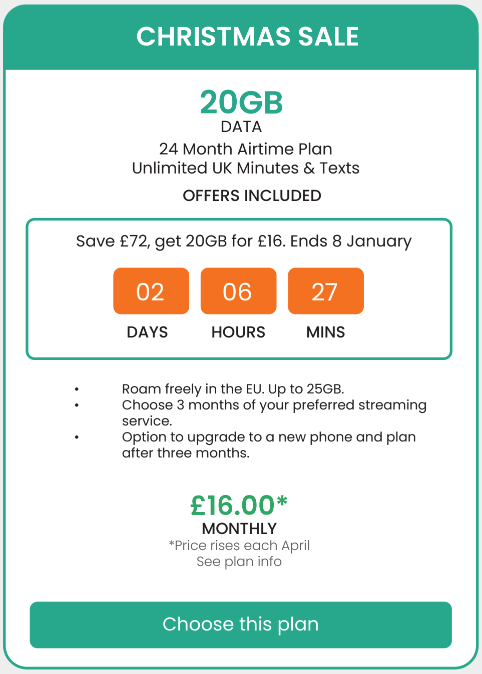
The third example features a gym membership plan. The choice architect employed sensory manipulation by framing this option in blue and highlighting it as the “Most popular” choice with a salient label. This practice, known as a popularity claim, was designed to guide customers to specific selections.

In the fourth example, a hotel offer utilised reference pricing, with the discounted price displayed in large, bold black font. At the same time, the savings were shown in a bright pink font to capture consumer attention. The choice architect also implemented a ranking system, rating the hotel offer at 8.8 out of 10, with the rating highlighted in green to draw attention.

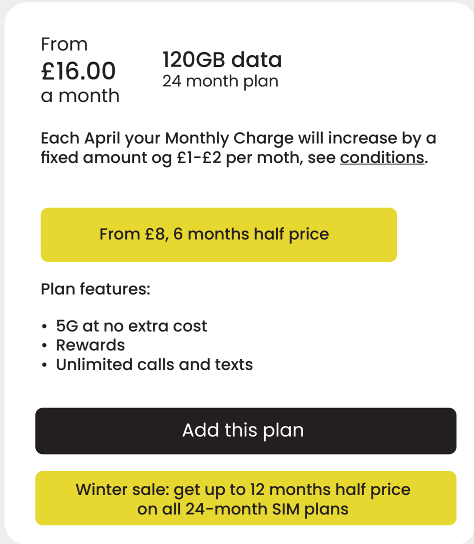
Figure D1. Desk Research Examples

Time-Bound Elements Examples

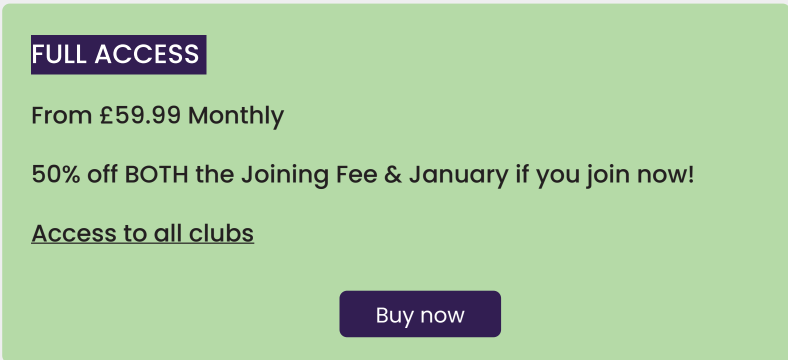
1



2



3



Desk research findings demonstrated that time-bound elements are often combined with multiple other OCA practices. In the first example, a phone plan was featured with a salient label "Christmas sale." The offer was only available for a limited time and included a countdown timer, adding time pressure to consumers. Additionally, the choice architect included information on how much the deal saved over the 24-month contract period.

In the second example, the advertised plan featured a time-limited discount, with the first six months offered at half the standard rate. Additionally, sensory manipulation was also employed by using yellow-coloured boxes to highlight the time-limited discounts.

The third example shows a gym membership deal that included a time-limited discount. The "Full Access" membership option was coloured in green to draw users' attention. Additionally, all deals found on this website included time-limited discounts.

Figure D2. Experimental User Flows and Interventions

User Flow | Telecom Marketplace: Purchase a Broadband Deal Task

This diagram illustrates the steps participants had to take to complete the shopping task in the telecom marketplace. They were asked to purchase a broadband deal that met the requirements outlined in the task instructions.

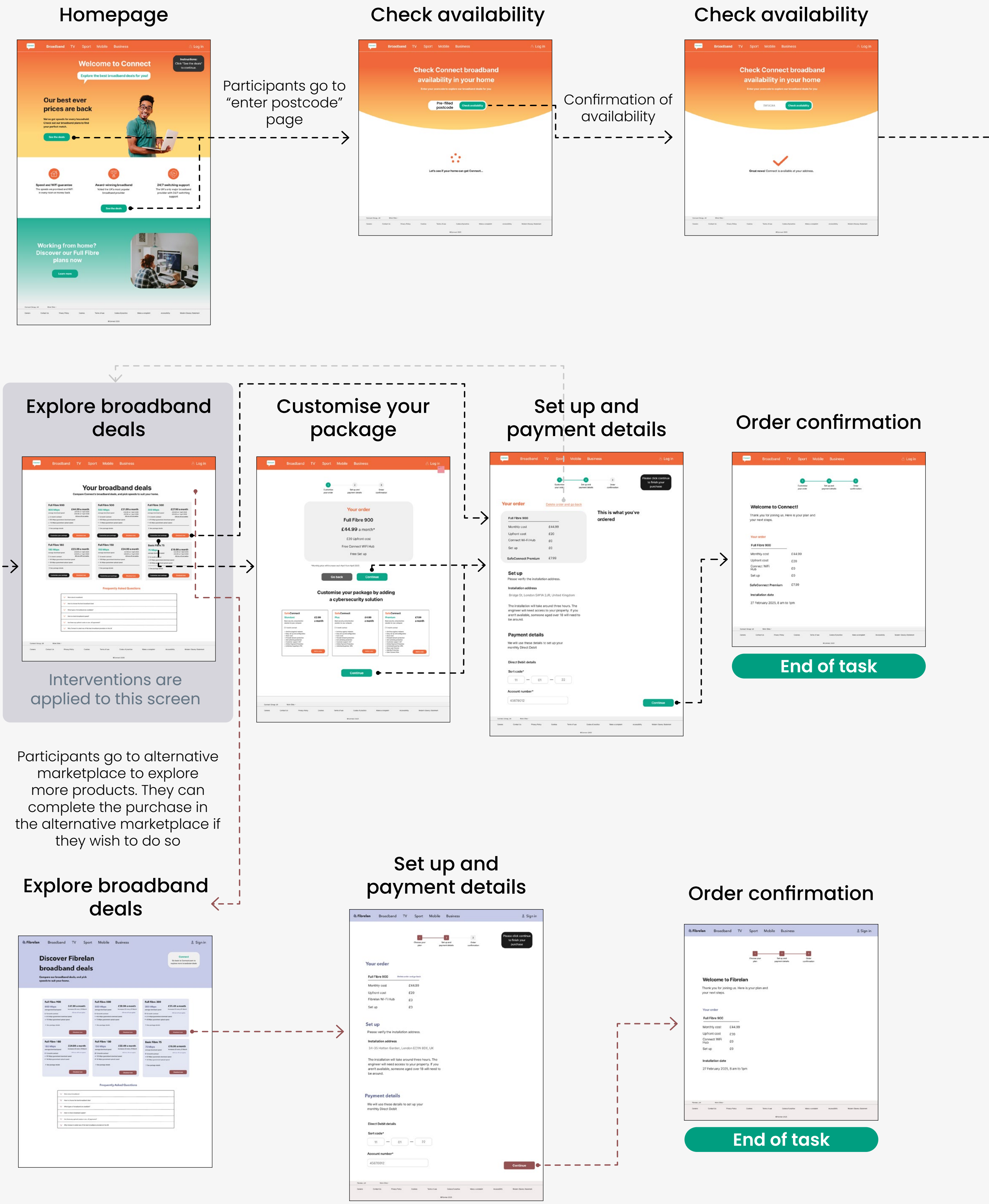


Figure D2. Experimental User Flows and Interventions

User Flow | Gym Marketplace: Purchase a Gym Membership Task

This diagram shows the steps participants had to take to complete the gym shopping task. The selected gym membership had to met the requirements listed in the task instructions.

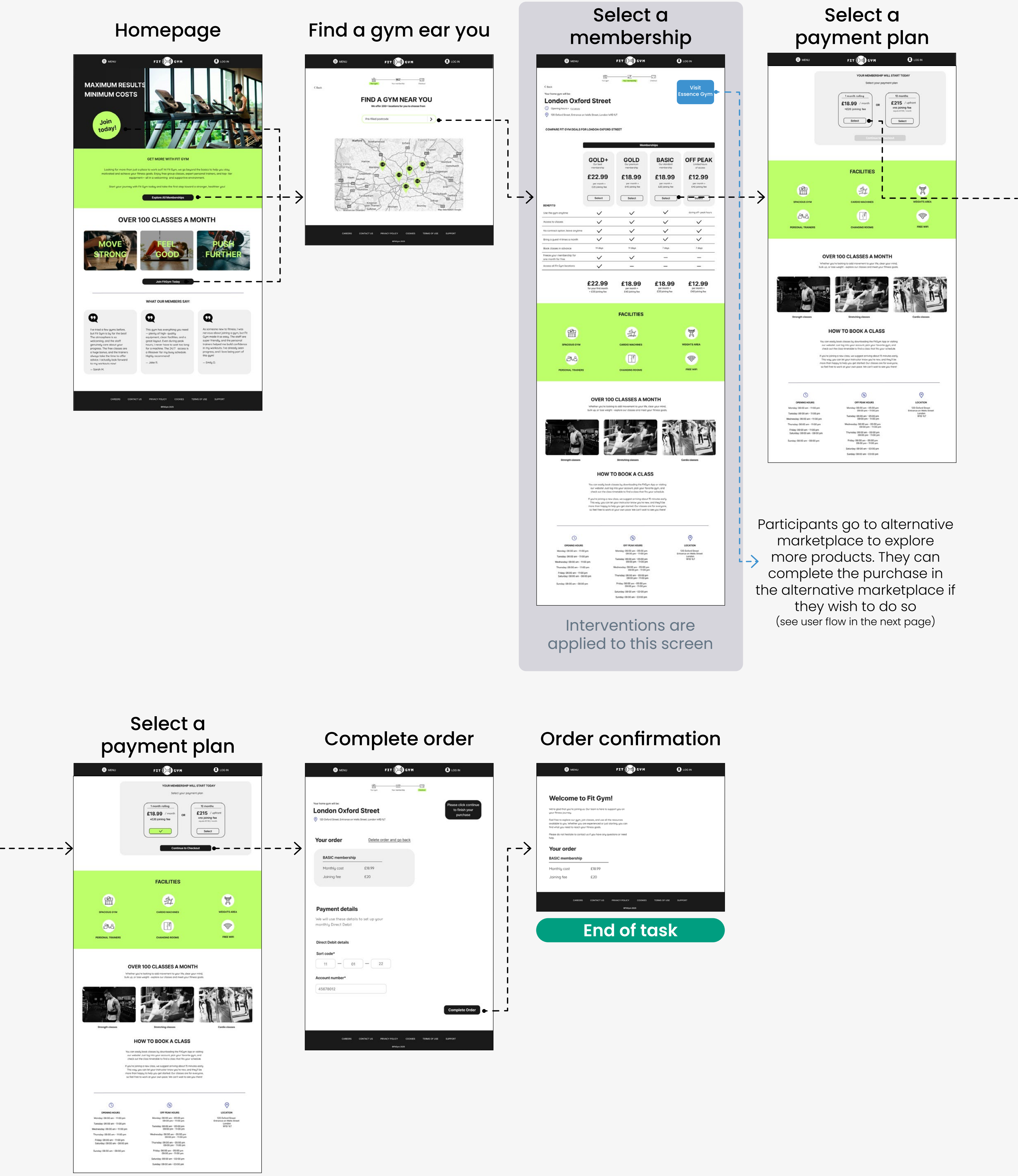
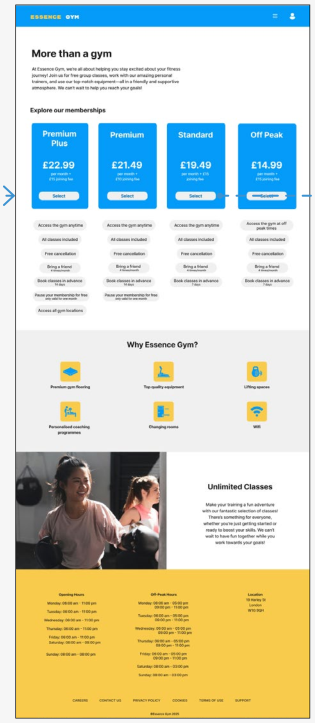


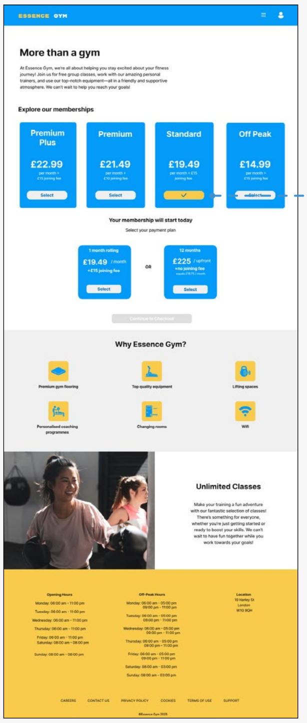
Figure D2. Experimental User Flows and Interventions

User Flow | Gym Marketplace: Purchase a Gym Membership Task

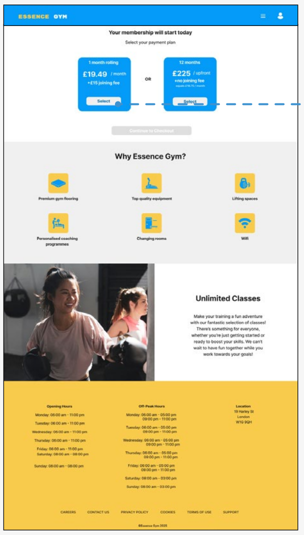
Select a membership



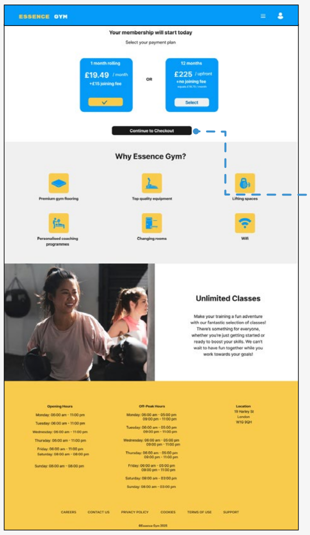
Select a payment plan



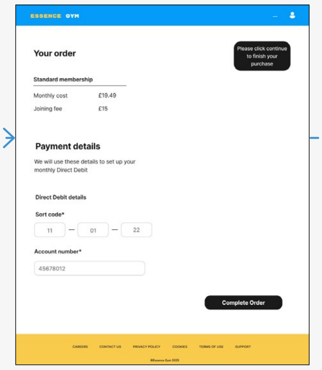
Select a payment plan



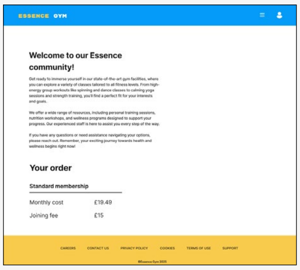
Select a payment plan



Complete order



Order confirmation



End of task

Figure D2. Experimental User Flows and Interventions

User Flow | Hotel Booking Marketplace: Book a Hotel Stay

This diagram illustrates the steps participants had to take to complete the hotel booking task. Participants were showed a hotel booking site and asked to book a hotel stay that fullfilled the needs described in the task instructions.

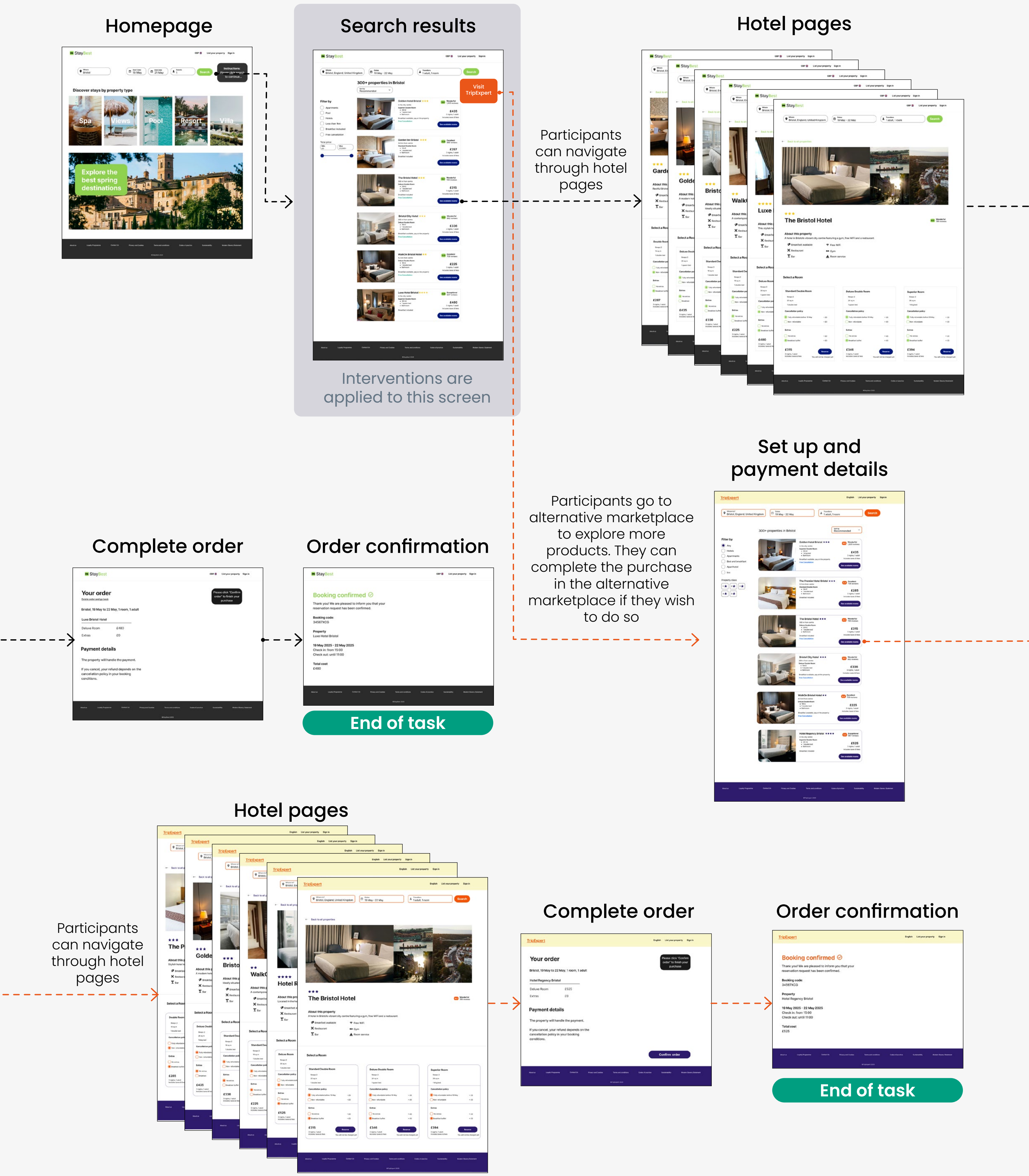
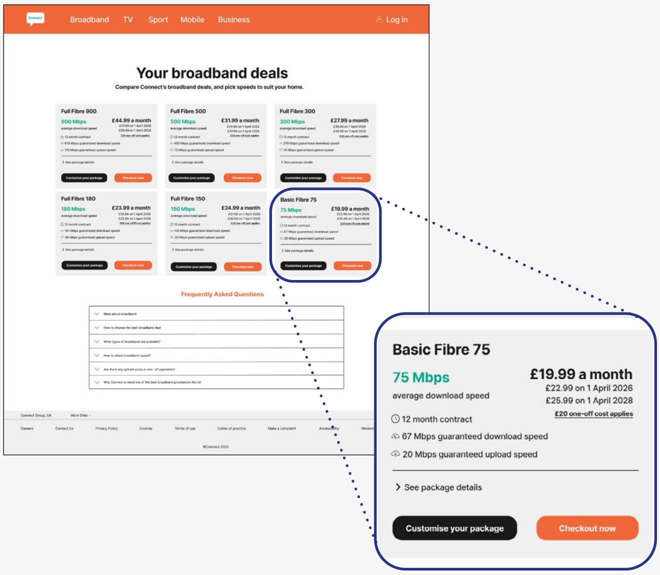


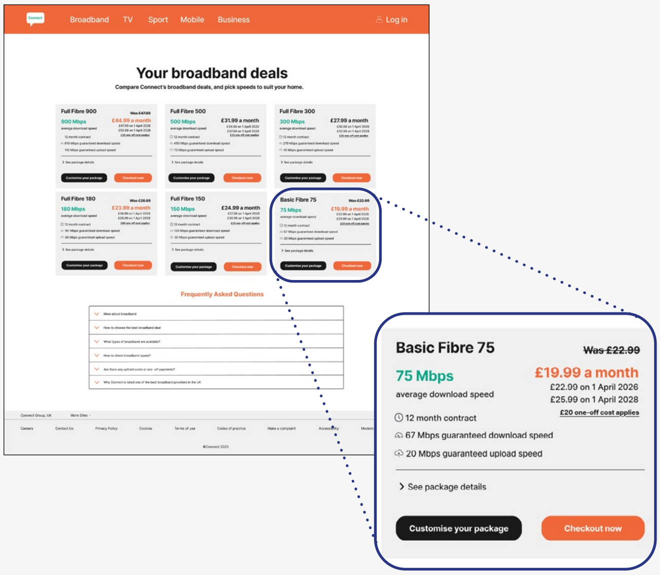
Figure D2. Experimental User Flows and Interventions

Interventions | Telecom Marketplace

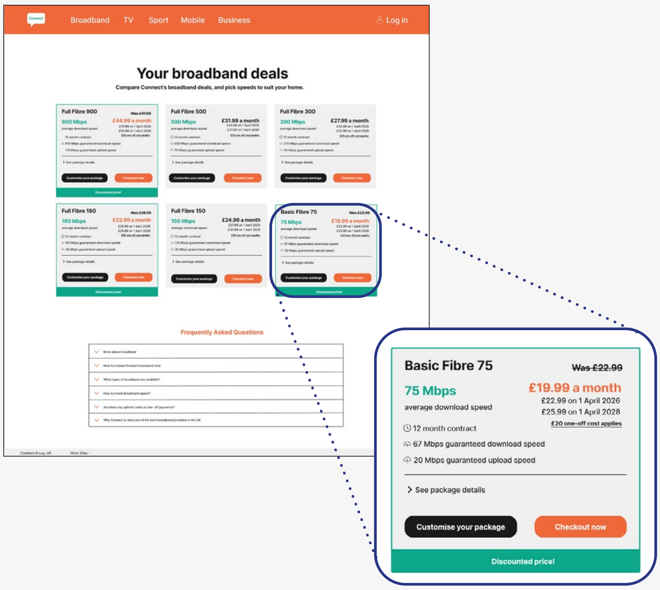
Control



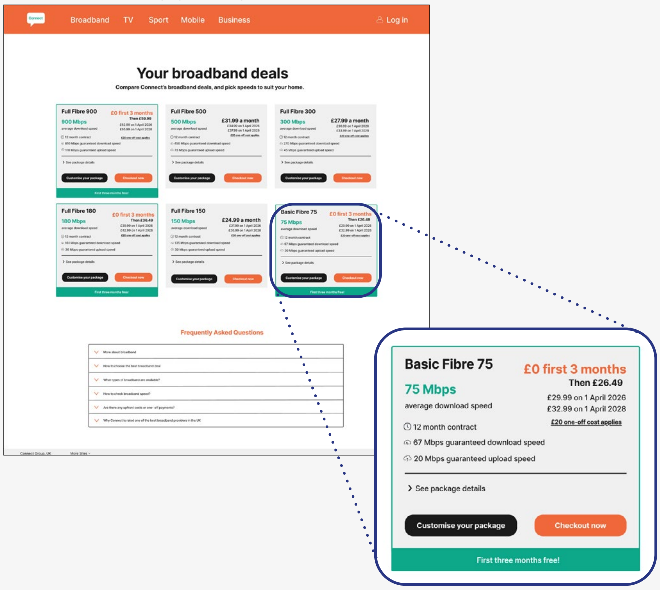
Treatment 1



Treatment 2

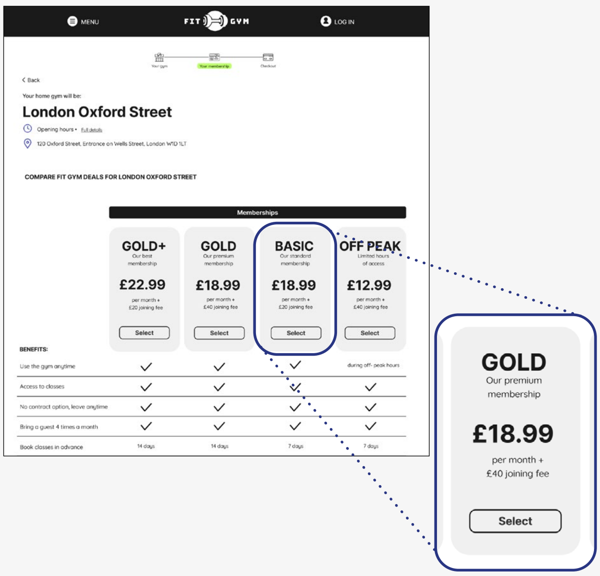


Treatment 3



Interventions | Gym Marketplace

Control



Treatment 1

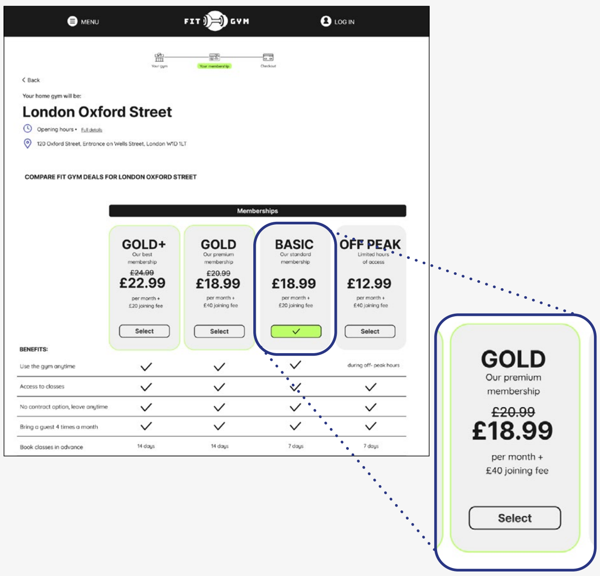
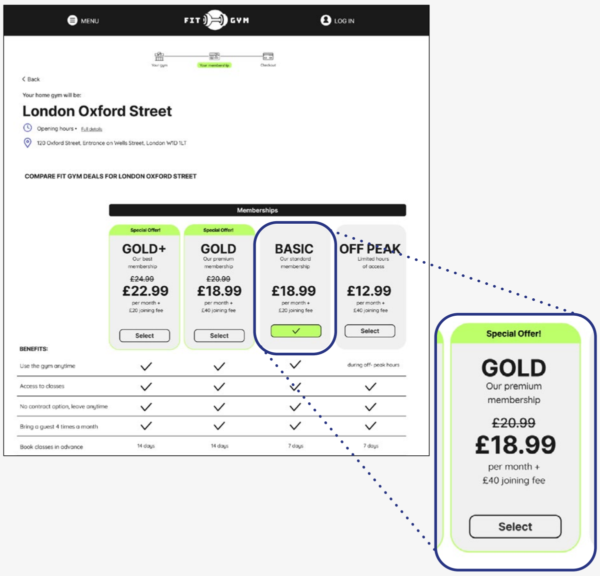


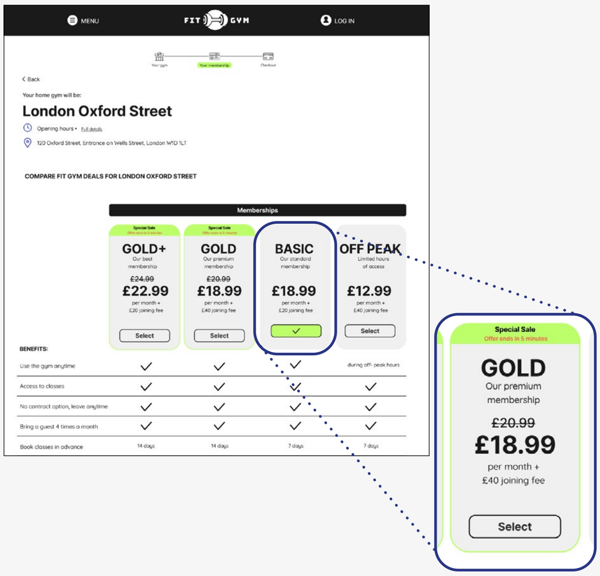
Figure D2. Experimental User Flows and Interventions

Interventions | Gym Marketplace

Treatment 2

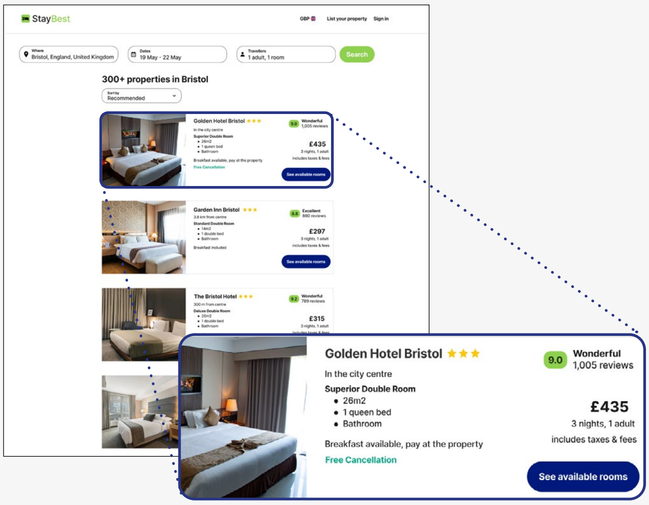


Treatment 3

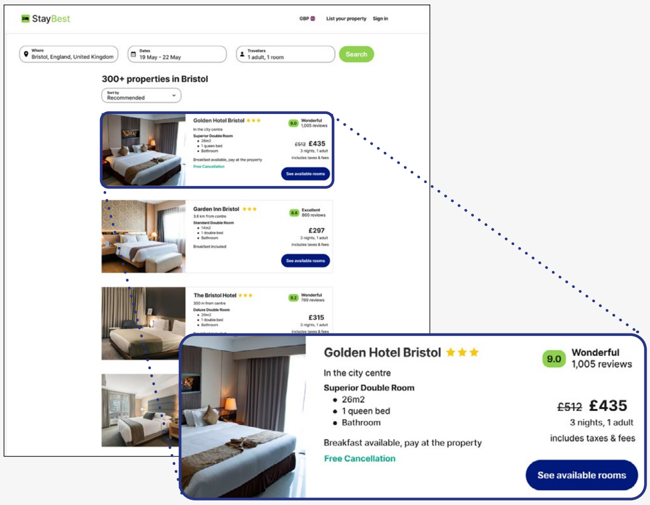


Interventions | Hotel Booking Marketplace

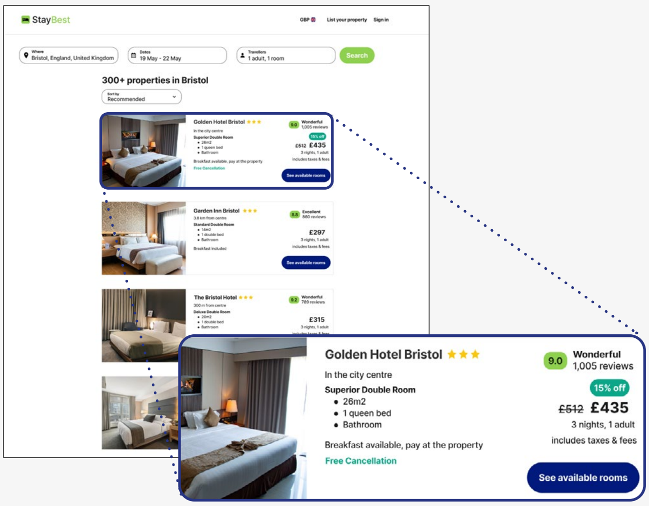
Control



Treatment 1



Treatment 2



Treatment 3