Health
Economics

Keeping an eye on cost: what can eye tracking tell us about attention to cost information in discrete choice experiments?

Journal:	*Health Economics*
Manuscript ID	HEC-22-0398.R2
Wiley - Manuscript type:	Research Article (Direct Via EEO)
Keywords:	Cost information processing, Discrete choice experiment, Eye-tracking, Multi-attribute choices

Keeping an eye on cost: what can eye tracking tell us about attention to cost information in discrete choice experiments?

ABSTRACT

Concern has been expressed about including a cost attribute within discrete choice experiments (DCEs) when individuals do not have to pay at the point of consumption. We use eye tracking to investigate attention to cost when valuing publicly financed health care. 104 individuals completed a DCE concerned with preferences for UK general practitioner (GP) appointments: 51 responded to a DCE with cost included and 53 to the same DCE without cost. Eye-movements were tracked whilst respondents completed the DCE. We assessed if respondents pay attention to cost. We then compare fixation time on attributes, eye movement patterns and mental effort across the experimental groups. Results are encouraging for the inclusion of cost in DCEs valuing publicly provided healthcare. Most respondents gave visual attention to the cost attribute most of the time. Average fixation time on multi-attribute tasks increased by 44% in the cost DCE, with attention to non-monetary attributes increasing by 22%. Including cost led to more structured decision-making and did not increase mental effort. Acceptability of the cost attribute and difficulty of choice tasks were predictors of cost information processing, highlighting the importance of both motivating the cost attribute and considering difficulty of the tasks when developing DCEs.

Keywords: Cost information processing; Discrete choice experiment; Eye-tracking; Multi-attribute choices

JEL codes: D01, D80, C35, D90, I12

1. INTRODUCTION

Discrete choice experiments (DCEs) are commonly used in health economics to investigate individuals' preferences for multi-attribute services (Clark et al., 2014; de Bekker-Grob et al., 2012; Lancsar & Louviere, 2008; Soekhai et al., 2019). DCEs are grounded in microeconomic theory (Lancaster, 1966; Manski, 1977), thus allowing welfare measures to inform policy decisions. When a cost attribute is included (e.g., out-of-pocket expense for medical services), willingness-to-pay (WTP) for changes in services can be estimated. These WTP values can be used within a cost-benefit analysis to inform health policy (McIntosh, 2006). However, questions have been raised about the credibility of including a cost attribute when individuals are not used to paying for health care at the point of consumption[1] (Genie et al., 2021). Lack of credibility may lead participants to change their choice behaviour, for example, by ignoring the cost attribute (Genie et al., 2021; Pedersen et al., 2011; Ratcliffe, 2000; Sever et al., 2019). Given marginal WTP is estimated as the ratio of any given attribute to the cost attribute, this leads to inflated monetary valuations (Balcombe et al., 2015; Scarpa et al., 2009). This limitation has been attributed to the hypothetical nature of the DCE, with choices not related to a budget constraint.

Research investigating the effect of a cost attribute in DCEs is limited. Five studies have addressed the issue in health economics. Whilst this literature provides mixed evidence regarding the impact of cost on preference ranking for non-monetary attributes (Bryan et al., 1998; Essers et al., 2010; Genie et al., 2021; Pedersen et al., 2011; Sever et al., 2019), there is emerging consensus of a negative impact of its inclusion on choice consistency (or response error variance) (Genie et al., 2021; Pedersen et al., 2011; Sever et al., 2019). The literature attributes this "cost effect" to increasing the volume of multi-attribute information to process, thus raising the cognitive difficulty of the choice tasks. However, these studies did not show how including a cost attribute increases the cognitive burden.

This paper contributes to the literature by using an eye-tracker alongside a DCE to investigate how individuals process the cost attribute. Studies in psychology have used eye movements to understand how information is processed (Rayner, 1998). The eye-mind hypothesis underpins most psychological analyses of eye-tracking data and suggests that visual search (i.e., looking at something) and attention (i.e., considering something) are tightly related (Just

---

[1] Studies have also used waiting time (Coast & Horrocks, 2007; Genie et al., 2020), risk (Harrison et al., 2014) and utility/benefit scores (Devlin et al., 2018; Murchie et al., 2016) to estimate value.

& Carpenter, 1993). Eye-tracking technology can thus provide a powerful tool for understanding economic behaviour (Harrison & Swarthout, 2019; Knoepfle et al., 2009; Lahey & Oxley, 2016).

Six studies have combined eye-tracking and DCE health research (see Online Supplementary Material **(OSM A)** for a summary). Spinks & Mortimer (2016) investigated the relationship between complexity and visual attribute non-attendance (ANA) when making choices between complementary and conventional medicine for different health conditions. They found complexity to be the strongest predictor of ANA when other possible influences, such as time pressure, ordering effects, survey specific effects and socio-demographic variables (including proxies for prior experience with the decision problem) were considered. Further, most respondents did not apply a consistent information processing strategy across choice sets. Within the context of preferences for lifestyle interventions, Krucien et al. (2017) show that treating information processing as a latent process outperforms models assuming full information processing. Further, the relationship between visual attention and individuals' preferences depends on the type of attribute: preferences for "easier to process" attributes are less influenced by changes in visual attention than "harder to process" attributes. Using the same eye-tracking data, Ryan et al. (2018) identified a range of visual biases, including a left-to-right, top-to-bottom, and first-to-last, and note these should be considered in the design of the DCE. Experimental factors (whether attributes are defined as "best" or "worst," choice task complexity, and attribute ordering) were also found to influence information processing and choice. Selivanova & Krabbe (2018) also found that respondents fixate slightly longer on the left-sided health-state descriptions. Within a study looking at preferences for breast cancer screening, Vass et al. (2018) investigated presentation/communication of risk (percentages or icon arrays and percentages) and decision-making strategies. They found no statistically significant difference in attention to attributes between communication formats. Respondents completing either version made more horizontal (left-right) saccades than vertical (up-down). Eye-tracking data confirmed self-reported ANA to the risk attributes. Sillero-Rejon et al. (2022) explored how cigarette packaging (standardised or branded) and health warning size affect VA and preferences among smokers and non-smokers (though they did not link their VA and preference data). They observed greater VA to warning labels on standardised packages; as warning size increased the difference in VA between standardised and branded packaging decreased. Standardised cigarette packaging and larger health warnings reduced preferences and have the potential to reduce the demand for cigarette products in Colombia.

Using an eye-tracker alongside a DCE we provide new evidence on how individuals process the cost attribute. Half the sample received a DCE with a cost attribute and the other half the same DCE without the cost attribute. We first assess how often respondents visually attend the cost attribute. We then compare three eye-tracking metrics across the experimental arms: (i) total fixation time spent looking at the attributes; (ii) information processing, measuring the dispersion of eye movements (scan path); and (iii) mental effort (proxied through pupil size). The latter two eye-tracking metrics have not been previously used in DCEs. We consider the effects of acceptability of the cost attribute and difficulty of the choice tasks on all eye-tracking metrics. We use a novel measure of entropy to measure difficulty, incorporating our eye-tracking data.

The rest of this article is organised as follows. Section 2 describes the experimental design and sample. Section 3 describes the methods to address our research questions. Section 4 presents the results and Section 5 discusses these results and considers their implications for the design of DCEs. Section 6 provides concluding comments.

2. EXPERIMENTAL DESIGN AND SAMPLE

2.1. Discrete choice experiment survey

We elicited preferences for an appointment with the general practitioner (GP) in the UK. We purposively chose this healthcare service because GP appointments are provided free at the point of delivery and the cost attribute is hypothetical. Further, the health care context is familiar to most people, making the multi-attribute information relatively easy to understand.

Based on the available literature concerning preferences for GPs (Hole, 2008; Longo et al., 2006; Rubin et al., 2006; Tinelli et al., 2016; Whitaker et al., 2017), the attributes and levels included in the DCE are shown in Table 1. Cost attribute levels were derived from a systematic review of the literature in a similar health care context (Hjelmgren & Anell, 2007; Hole, 2008).

[Table 1 about here]

We used NGENE software (ChoiceMetrics) to generate a D-efficient design with 12 choice tasks (Bliemer & Rose, 2005). The design was based on null priors[2] and optimised for the estimation of a multinomial logit (MNL) model. Given the relatively limited number of attributes' levels, it would have been possible to include fewer choice tasks (technically, the minimum required was six). However, we were interested in how information processing evolves over the sequence of choice tasks (see below). In addition, using an eye-tracker during the experiment led to a relatively small sample of respondents; we increased statistical power by increasing the number of observations per respondent.

In addition to the 12 choice tasks, a warm-up task was included. Choice tasks were unforced pairwise choices among generic options, with an opt-out option. Respondents were told: *Imagine you have had a cough for more than 3 weeks. This is keeping you awake at night. You have tried several home treatments to remedy this such as taking rest, drinking plenty of fluid, drinking hot lemon with honey. However, your cough is not improving, and you have decided that it is now time to consult a GP*. Respondents were also told that if they choose the "neither" appointment this would mean they have decided not to see a GP.

To separate eye-tracking recordings during the actual decision making and using the mouse to respond, participants were asked to press a key to indicate they were ready to respond (Part I), after which they used the mouse to indicate their preference (Part II). Figures 1 and 2 show example choice tasks for the COST DCE and NOCOST DCE. To minimise ordering effects the order of the choice tasks and options within the tasks were randomised across participants (Craig et al., 2015; Janssen et al., 2018; Kjær et al., 2006). The order of attributes within options was fixed and presented in the order shown in Table 1.

Information was also collected on respondent's experience of paying for the cost attribute, how acceptable they found the cost attribute and perceived difficulty of the choice tasks. Question formats are shown in Table 2. At the end of the choice tasks respondents were asked which features of GP appointments they never considered in their choices.

---

[2] A D-efficient design with non-informative priors is equivalent to an orthogonal design (Szinay et al., 2021). In the absence of pilot studies, Bliemer and Collins (2016) suggest using expert judgement to inform priors. We thank an anonymous reviewer for this point.

2.2. Experimental manipulation

The two DCEs were identical other than one included the cost attribute (COST DCE) and the other did not have a cost attribute (NOCOST DCE). Respondents were randomly allocated across the two conditions using the "biased coin" procedure (Smith, 1984): for every new participant, the probability to be assigned to one condition depended on the number of participants already allocated to the two conditions. This procedure allows preserving the randomness of allocation, ensures a perfect balance (i.e., the same number of participants in the two conditions) and easily handles non-participation (i.e., individuals who do not turn up).

[Figure 1 about here]

[Figure 2 about here]

OSM B shows the MNL regression results for the COST DCE and NOCOST DCE. All coefficients have the expected signs, confirming the theoretical validity of the models. Respondents preferred higher flexibility, continuity of care, a longer length of consultation and lower waiting time. In the COST DCE a lower cost was preferred, and individuals were willing to pay: £15.60 for flexibility; £3.75 for a one-day reduction in waiting time; £15.80 for continuity and £2.95 for a one-minute increase in the length of consultation. These results have face validity, with comparable costs of a private GP consultation in the UK (e.g., see https://www.bupa.co.uk/health/payg/gp-services; https://www.mytribeinsurance.co.uk/treatment/cost-to-see-a-private-consultant-uk.

2.3. Eye-tracking

We used an EyeLink 1000 system to record respondents' eye movements while completing the DCE. Eye movements were recorded at a 1,000 Hz frequency (i.e., one observation every millisecond). Participants were seated at an approximate distance of 77 cm from the display monitor. The eye-tracker was calibrated individually with the default nine-point calibration method, done at the beginning of the experiment. Calibration allows for the reverse mapping of the location of the pupil and corneal inflection in the image of the participant's eye to the gaze position on the screen. A calibration that is considered 'good' by the Eyelink 1000 system ensures that the recorded gaze position is within 0.5 degrees of visual angle from the actual gaze position (Balcombe et al., 2015; Gibaldi & Sabatini, 2021). To avoid biases towards particular areas on the screen at stimulus onset, each choice task started with a fixation point presented in the middle of the screen (Krucien et al., 2017; Vass et al., 2018). This procedure (i.e., a between-choice task calibration) also served to correct for any movement of the respondent's head (known as `drift'), thereby improving the accuracy of the collected data (Vass et al., 2018). Respondents were asked to fixate on a fixation point, after which the experimenter initiated the experiment, so that recorded gaze was not influenced by small head movements that could happen if participants would press the key themselves.

Respondents were told that the study was about investigating preferences for GP appointments and that eye-tracking was used to understand how they made their decisions. They were not informed about our focus on the cost attribute. The experiment took place in a dark, windowless room with minimal luminosity to avoid infrared from sunlight.

The eye-tracking data were automatically divided into fixations (i.e., periods where the eyes remain relatively still) and saccades (i.e., fast eyes' movements during which information processing is suppressed) using the default algorithm and saccade detection settings of the eye-tracking system. It was assumed that information extraction only took place during the fixations and that a minimum of 50 milliseconds (ms) was needed for meaningful extraction of information (Balcombe et al., 2015; Krucien et al., 2017; Ryan et al., 2018; Tatler et al., 2006). Fixations were automatically assigned to the 17 and 14 areas of interest (AOI) for the COST DCE and NOCOST DCE, respectively (Figures 1 and 2).

The choice tasks for the ET were presented in picture format on a white background. The AOIs shown in Figures 1 and 2 are displayed in terms of rectangular areas. Whilst it is up to the researcher to define these areas, which typically include some space around the text or picture of focus to account for issues with accuracy and precision (Holmqvist, 2011), it has been indicated that a 3.2 cm AOI will provide 80% capture rate (Orquin & Holmqvist, 2018). All AOIs were consistent in terms of size (width and height) and shape (rectangular) and of sufficient size and space to distinguish between AOIs. Movement from one AOI to another is known as a transition (or 'gaze shift').

2.4. Sample and recruitment

We used the Louviere et al. (2000) formula to determine the minimum sample size. Based on a choice probability of 50%, an accuracy level of 90%, a confidence level of 95% and 12 choice tasks per participant, we needed to recruit a minimum of 44 respondents per condition (Louviere et al., 2000). We anticipated a 25% maximum attrition rate due to technical difficulties with eye-tracking. Thus, we recruited 60 respondents for each experimental condition. To be eligible, participants needed to be: (i) older than 16 years; (ii) able to complete the consent form, (iii) able to answer the questionnaire in the English language, and (iv) not suffer from severe visual impairments (e.g., blindness). Although our sample is not large by DCE standards, it is comparable with other eye-tracking DCE studies in the health literature (see **OSM A**). We recruited participants on the University of Aberdeen (UK) campus using flyers. One-to-one appointments were arranged with the participants who had to attend an experimental laboratory. Respondents received a £20 voucher as compensation for their time. The study was approved by the University of Aberdeen's College Ethics Board (Reference: CERB/2018/2/1538).

The two samples did not differ in terms of socio-demographic characteristics (Table 2). Participants ranged in age from 19 to 69, with an average age of 35 in the COST DCE and 37 in the NOCOST DCE. Female participants made up 65% (36/55) in the COST DCE and 70% (42/60) in the NOCOST DCE (p=0.748). As expected, most respondents had no experience of paying for GP appointments. Information about the cost of a GP appointment was deemed acceptable by the majority of respondents in both arms. Despite an extra attribute in the COST DCE, respondents did not perceive this experiment to be more difficult than the NOCOST DCE (p=0.14).

[Table 2 about here]

3. METHODS

We first assess whether respondents visually attend the cost attribute. We then assess the effect of the EXPERIMENT (COST DCE or NOCOST DCE) on: (ii) fixation time on the monetary and non-monetary attributes; (iii) information search behaviour and (iv) mental effort.

3.1 Do individuals visually attend the cost attribute?

As noted by Just & Carpenter (1993), visual search (i.e., looking at something) and attention (i.e., considering something) are tightly related, such that visual fixation on an attribute is an indicator of attention given to that attribute. Given most respondents (n=47) looked at the cost attribute most of the time, the probability of being visually ignored was low. We thus modelled the share of visual attention on the cost attribute, computed as the total time spent looking at the cost attribute divided by the total time spent looking at all attributes. We used share of visual attention (rather than actual amount of time) to avoid strong effects of long fixations on an attribute and to reduce the skew of the distribution. Further, the share of visual attention on cost provides more information about "attention capture" (e.g., whether search behaviour is biased towards the cost attribute) whilst the total time corresponds to the "depth of information processing" (i.e., whether the cost attribute was superficially processed or not).

3.2 Impact of the cost attribute on fixation time

While we did not impose a time limit, individuals may have a self-imposed time limit when completing the survey. Whilst we expect longer fixation times in the COST DCE (as individuals have more information to process), if the respondent has a time limit the extra attribute may come at the expense of information processing on other attributes. The cost attribute might then act as a reference point, speeding up decision-making. Alternatively, the cost attribute may increase engagement in the DCEs, focusing respondents on the opportunity cost, and then they may spend more time on all attributes. We compare average fixation time both across the 12 choice tasks and on each attribute.

3.3 Does inclusion of the cost attribute influence information search behaviour?

Bogomolova et al. (2020) noted that when individuals are motivated to search for lower prices, they fixate more on cost information, and hence make a more focused information search. We explore if including a cost attribute influences the information search behaviour using the dispersion of transitions (Holmqvist, 2011). If respondents follow a structured information search strategy, the dispersion of transitions will be limited, with most transitions on adjacent AOIs. For example, in Figure 1, a move from 7 to 8; 8 to 9; 9 to 10; 10 to 11 in the case of option-wise search or 7 to 13; 8 to 14; 9 to 15; 10 to 16; 11 to 17 in case of attribute-wise search. In contrast, a less structured information search is associated with larger transitions between non-adjacent AOI. For example, a move from 7 to 11; 11 to 14; 13 to 17; 8 to 10; 10 to 13). See Online Supplementary Material **(OSM C)** for a depiction of structured and unstructured information search patterns. Following Bogomolova et al. (2020), we hypothesise that including a cost attribute results in a lower average distance or shorter transitions across different AOI i.e., a more focused/structured information search pattern.

The collection of all transitions between AOIs is known as the *scan path*. We measured the scan path length as the total distance[3] covered by the eyes during the transitions. Instead of using the actual (X, Y) coordinates of the fixations, we reduced noise in the data by normalising the distances as follows: AOI-7 to AOI-11 took the coordinates (1;1) to (5;1) and the AOI-13 to AOI-17 the coordinates (1;2) to (5;2). The longest transition was thus between AOI-1 (i.e., the attribute 'flexibility' of option A) and AOI-17 (i.e., the attribute 'cost' of option B), and the shortest transitions were made between adjacent AOIs. As the dependent measure, we then computed the length of a line segment between two consecutive fixations (A; B) using the Euclidean distance (D) formula[4] computing for each participant (n) and choice task (t).

3.4 Does the inclusion of the cost attribute require a higher level of mental effort?

We approximated mental effort based on the size (or dilation) of the pupil. Using pupil size as an indication of mental effort can be traced back to Hess & Polt (1964), who demonstrated that pupil size increases with problem difficulty within the context of solving multiplication problems: pupil dilation increased about twice as much (22 versus 11 percent) when participants calculated 16 times 23 compared to 7 times 8. Kahneman & Beatty (1966) suggested that pupil size provides a "*very effective index of the momentary load on a subject as they perform a mental task*." They found larger pupil size when participants memorised more digits (0.1 mm versus 0.55 mm for 3 versus 7 digits). Kahneman (1973) argued that pupillometry (pupil size and reactivity) is "*the best single index*" of effort, capturing within-task, between-task, and between-individual variation. Following this early work, pupil size has been reported in many contexts related to mental workload (or cognitive demand) (Eckstein et al., 2017; Hartmann & Fischer, 2014; Just & Carpenter, 1993; van der Wel & van Steenbergen, 2018) with difficult tasks that require significant mental effort (memory load) leading to the pupils dilating (Korn & Bach, 2016; Laeng et al., 2012). For extensive literature reviews, see Beatty (1982) and van der Wel & van Steenbergen (2018).

---

[3] We then obtained the average dispersion of transitions by dividing the total distance across different AOI in each task by the number of transitions made in each choice task.

[4] The formula for the Euclidean distance is $D_{(A,B)} = \sqrt{(x'_A - x'_A)^2 + (y'_B - y'_B)^2}$

We measured pupil size by counting black pixels on the camera image of the eye to measure pupil diameter. We estimated average pupil size per participant fixating to an attribute in a choice task and pupil dilation as the change in pupil size while fixating (largest-smallest). Pupil size may be influenced by factors such as fatigue, the brightness of the stimuli and the brightness of the environment. We controlled for these factors by: (i) recruiting participants during both morning and afternoon sessions (65%-35% split for the COST condition and 63%-37% split for the NOCOST condition); (ii) running the experiment in a room without a window; and (iii) ensuring stimuli brightness did not change across choice tasks. We used pupil size measured on the warm-up task as a baseline[5] measure and subtracted this from the average pupil size recorded for the 12 experimental tasks. Analysing changes in pupil size rather than absolute pupil size helped to attenuate the level of noise in the data. Given the pupil takes 200 to 800 ms to respond (Korn & Bach, 2016), we conducted the analysis at the task level by averaging all the observations (i.e., pupil size recorded at each fixation).

3.5 Econometric analysis

For all eye-metrics we controlled for cost ACCEPTABILITY and DIFFICULTY of the choice tasks**.** We converted ACCEPTABILITY responses from the survey into a binary variable indicating whether respondents found the cost of GP appointments acceptable ('*completely acceptable*' or '*acceptable*') or not ('*not acceptable*' or '*not acceptable at all*'). There is evidence that information processing of cost depends on price consciousness with high price-conscious consumers seeking the lowest price (Burton et al., 1998; SPROTLES & KENDALL, 1986) and low price-conscious consumers driven by non-price product attributes (Hwang & Lorenzen, 2008; Youn & Kim, 2017). Further, individuals who find cost more acceptable are more cost-conscious, impacting on their visual attention (Ngan et al., 2022). We thus hypothesise that individuals who find cost more acceptable are more likely to pay visual attention to it.

We converted perceived DIFFICULTY collected in the survey into a binary variable indicating whether respondents found making choices difficult or not ('*very easy*' or '*easy*'). Perceived difficulty may be subject to the same biases found with attribute non-attendance de-briefing questions (Kragt, 2013; Mørkbak et al., 2014) and eye-tracking data where they often do not

---

[5] We also included the pupil size on the warm-up task as an additional control (instead of differencing); results are robust and available from the authors.

correlate. We also used two objective measures of difficulty: entropy of transitions and deviation of standard deviations (DSD). The entropy[6] of transitions measure, an estimate of fixation sequence randomness, has been used in previous studies to estimate workload (e.g., Monfort et al., 2016). The higher the entropy, the more random the transition processes across different AOIs in a choice task, and the higher the choice-task difficulty. Shugan (1980) argued that difficulty is inversely related to perceptual similarity - highly different options are more difficult. As alternatives become less similar, the variance in the values on the attributes across alternatives increases. This can be captured by the dispersion of the standard deviation (DSD) among attribute levels across alternatives (DeShazo & Fermo, 2002). Ryan et al. (2018) found a positive relationship between task difficulty and visual attention. Difficult tasks may however reduce processing time as individuals adopt decision heuristics e.g., use cost as a reference point (Lockshin et al., 2006).

Previous studies have reported the impact of task order on the consistency of respondents' choices (Bateman et al., 2008; Day et al., 2012; Mantonakis et al., 2009; Ryan et al., 2018), suggesting learning and fatigue effects. We divided the 12 choice tasks into three BLOCKS (BLOCK 1, Tasks 1-4; BLOCK 2, Tasks 5-8; BLOCK 3, Tasks 9-12). We split the choice tasks into three blocks to attempt to capture the effect of learning and fatigue effects: Block 1 (learning), Block 2 (optimum), and Block 3 (fatigue). BLOCK 2 was the reference.

---

[6] The concept of entropy (H), as originally defined by Shannon (1948), is a measure that calculates the uncertainty in a random variable. The entropy of a transition matrix (R) is:

$$H(R) = -\sum_r P_r log_2(P_r)$$

where r are the off-diagonal elements of R, and P_r is the cell probability. For the COST DCE each task included 10 areas of interest (AOI) and then the transition matrix had 90 off-diagonal elements. For the NOCOST DCE, each choice task consists of 8 AOI and the transition matrix had 56 off-diagonal elements. Entropy reaches its maximum value when all the transitions (cells) are equally likely to happen; in our case this value is $-log_2(0.01) \simeq 6.5$ for the COST DCE and $-log_2(0.02) \simeq 5.8$ for the NONCOST DCE. Alternatively, if the participant focused on one AOI, all the off-diagonal elements would be zero, and entropy would be zero. The entropy is initially measured in information units, known as *bits*. 'Bits' is not a very intuitive unit. To facilitate comparison across individuals and conditions, we normalised this measure by dividing each entropy score by the maximum score possible (6.5 for our COST DCE and 5.8 for our NOCOST DCE). Entropy is an indicator of the randomness of fixation distributions between AOIs (Acartürk & Habel, 2012; di Nocera et al., 2006). Higher entropy could be associated with a higher mental workload as well (Kruizinga et al., 2006). Shic et al. (2008) argued that a high entropy value would indicate a preference for exploration (more random transition processes), while low values indicate data with transitions mainly between a few AOI.

Finally, we controlled for AGE and GENDER. Spooner et al. (1980) highlight the importance of considering age when evaluating eye movements; eyes have been shown to exhibit an age-related decline in performance (Cabeza et al., 2004; Curran et al., 2001; Hahn et al., 2011; Pesce et al., 2005), resulting in difficulties in processing information (Salthouse, 1996). Pupil size has been shown to decrease linearly with age (Rio et al., 2016; Winn et al., 1994) and to depend on gender, with males demonstrating greater pupil size (Iyamu & Osuobeni, 2012; Murray et al., 2017).

A Beta regression model was used to estimate factors determining share of visual attention on the cost attribute (Cribari-Neto & Zeileis, 2010):

$$Y'_{nt} = \beta_0 + \beta_1\, ACCEPTABILITY_n + \beta_2 DIFFICULTY_n + \beta_{3:4}\, BLOCK_{nt} + \beta_5 AGE_n + \beta_6 GENDER_n \quad (1)$$

where Y'_{nt} refers to the share of visual attention on cost attribute by participant (n) at task (t) corrected for visual cost attribute non-attendance[7].

We then estimate linear mixed-effect regression models to address research questions (3.2) to (3.4):

$$ET\ METRIC_{nt} = \beta_0 + \beta_1 EXPERIMENT_{nt} + \beta_2 ACCEPTABILITY_n + \beta_3 DIFFICULTY_n + \beta_{4:5}\, BLOCK_{nt} + \beta_6 AGE_n + \beta_7 GENDER_n + \alpha_n + \varepsilon_{nt} \quad (2)$$

where $ET\ METRIC_{nt}$ refers to the relevant eye-tracking outcome (i.e., fixation time, visual information search (i.e., dispersion of transition) and change in pupil size) by participant (n) at task (t)). The errors (α and ε) are assumed to be normally distributed and uncorrelated.

---

[7] For cases where the cost attribute was visually ignored, and the corresponding share of visual attention is null, we applied the correction (Smithson & Verkuilen, 2006):

$$Y' = \frac{Y(n-1) + 0.5}{n} \quad (3)$$

where Y corresponds to the initial share of visual attention, Y' to the corrected measure, and n to the total number of observations (i.e., 50 participants x 12 tasks + 1 participant x 11 tasks = 611). The average proportion of time spent looking at the cost attribute was 0.1507 (SD = 0.1052) and 0.1513 (SD = 0.1052) before and after correction, respectively.

All four regression models were estimated in R (Brown, 2021; Cribari-Neto & Zeileis, 2010).

4. RESULTS

Due to technical difficulties with eye-tracking, only 51 respondents for the COST DCE and 53 respondents for the NOCOST DCE were used in the final analysis. Fixations in the white spaces outside the AOIs were assumed to indicate "daydreaming" or disinterest (Vass et al., 2018) and excluded from analysis. After excluding fixations from outside of the 17 AOIs in the COST DCE and 14 AOIs in the NOCOST DCE, 73,092 fixations remained. After removing fixations on the descriptive column (AOI-1 to AOI-5 in Figure 1 and AOI-1 to AOI-4 in Figure 2) and labels of alternatives (AOI-6 and AOI-12, Figure 1; AOI-5 and AOI-10 in Figure 2), 37,517 fixations remained. After removing individuals with poor data quality due to eye-tracking problems resulting in no fixation data (n=16; 13%), 35,200 fixations remained. Combining consecutive fixations on the same AOI and removing duplicated fixations, 26,255 fixations (observations) remained for analysis.

Below we report the results for the four research questions. Our three measures of difficulty, self-perceived, entropy transition and DSD consistently gave the same results. We discuss the results with entropy of transitions. **OSM D** shows results with the self-perceived and DSD difficulty measures.

4.1 Do individuals visually attend the cost attribute?

Of 3,055 observations,[8] we observed 179 (5.9%) cases of VANA across the five attributes: flexibility, 42 (24%); waiting time, 21 (11.7%); continuity, 31 (17.3%); length of consultation, 35 (19.6%); cost, 49 (27.4%). Consistent with (Balcombe et al., 2015), most respondents paid attention to most of the attributes (94.1%). VANA was not uniformly distributed across the attributes (Chi-2 = 13.09; P < 0.011)[9]. The 49 cases of VANA for the cost attribute constituted 8% of observations (49/611)[10]. However, two participants accounted for nearly half of the cases (i.e., 22/49).[11]

---

[8] 50 participants x 12 tasks x 5 attributes + 1 participant x 11 tasks x 5 attributes = 3,055.

[9] A similar analysis in the NOCOST DCE indicated 135 (5.3%) cases of VANA that were distributed across the four attributes: flexibility, 30 (22.2%); waiting time, 19 (14.1%); continuity, 32 (23.7%); and length of consultation, 54 (40%). As in the COST DCE, most respondents paid reasonable attention to most attributes (94.7%). VANA was not uniformly distributed across the attributes (Chi-2 = 19.1; p = 0.0003).

[10] 50 participants x 12 tasks + 1 participant x 11 tasks = 611.

Beta regression results are presented in Table 3. Respondents who considered information about the cost of a GP appointment to be acceptable paid more attention to cost. Increased difficulty also increased visual attention. Males gave relatively less attention to cost whilst older people gave more attention to cost.

[Table 3 about here]

4.2 Impact of the cost attribute on fixation time

There was a higher fixation time (FT) in the COST DCE for 11 of the 12 choice task compared to the NOCOST DCE (Figure 3). Time spent looking at the multi-attribute content increased by 44% in the COST DCE (average FT per task, μ_{NOCOST} = 3,697 ms [95% CI: 3,497; 3,896]; μ_{COST} = 5,345 ms [95% CI: 5,065; 5,626]).

[Figure 3 about here]

This increase in FT may hide a re-allocation of cognitive resources across the attributes. For example, if the cost attribute is difficult to process, respondents may both increase their level of visual attention (i.e., allocating more cognitive resources to the completion of the choice tasks) and transfer resources from the other attributes to the cost attribute. Notably, there was a 22% increase in average fixation time on non-monetary attributes, with fixation times increasing for all the non-monetary attributes, in the COST DCE compared to the NOCOST DCE (Figure 4).

[Figure 4 about here]

Linear mixed effects regression results are presented in Table 4, column 2.

[Table 4 about here]

EXPERIMENT (i.e., including a cost attribute) has a positive and statistically significant effect on fixation time. The DIFFICULTY coefficient suggests that higher task difficulty increases fixation time. BLOCK-3 (compared to BLOCK-2) contributed to a reduction in fixation time;

---

[11] Further, five participants accounted for 5 of the 49 cases, two participants accounted for 4 of the 49 cases, two participants accounted for 6 of the 49 cases, one participant accounted for 5 of the 49 cases, and one participant accounted for 7 of the 49 cases.

as participants progress through the later positioned choice tasks, they spend less time looking at the different AOIs.

4.3 Does inclusion of the cost attribute influence information search behaviour?

On average, participants made shorter transitions in the COST DCE for each choice task compared to the NOCOST DCE (Figure 5). The linear mixed effects regression results (Table 4, Column 3) confirmed this relationship, with a negative and significant effect of the EXPERIMENT on the average dispersion of transitions. This confirms the hypothesis of a more structured/focused information search in the COST DCE. We again found the task order to have an effect with the first positioned choice tasks (BLOCK-1, Tasks #1-#4) having a significant negative effect on the average distance (dispersion).

[Figure 5 about here]

4.4 Does the inclusion of the cost attribute require a higher level of mental effort?

The change in average pupil size is slightly higher in the COST DCE for each choice task (Figure 6).

[Figure 6 about here]

The linear mixed effects regression results (Table 4, Column 4) indicate that this difference is not significant, suggesting that the inclusion of a cost attribute did not increase mental effort. The first positioned choice tasks (BLOCK-1) had a positive and significant effect on the changes in pupil size, indicating that the first block of experimental tasks (Task #1-#4) required a higher mental effort (was more cognitively demanding) whilst later positioned tasks (BLOCK-3, Task #9 - #12) are relatively less cognitively demanding.

5 DISCUSSION

Using an eye-tracker, we explored individuals' processing of the cost attribute in a DCE conducted within a publicly provided health care system, where services are free at the point of consumption. Despite concerns often expressed about inclusion of the cost attribute in health DCE surveys, our results are encouraging with most respondents attending to the cost attribute most of the time. The cost attribute engaged individuals in the experiment, with fixation time on non-monetary attributes higher in the COST DCE and responses following a more structured information search. Including a cost attribute did not make tasks more cognitively demanding. Previous studies found that the cost attribute led to a significantly higher response error variance (Genie et al., 2021; Pedersen et al., 2011; Sever et al., 2019). These studies attributed the higher error variance to the increased cognitive burden. Our findings do not support this hypothesis. Further, our finding suggests that moving from four to five attributes does not increase mental effort. Acceptability of the cost attribute of choice tasks was a predictor of cost information processing and increased difficulty consistently led to increased visual attention.

Our findings have a number of practical implications for DCE practitioners. Firstly, whilst cost is known to be important in consumers' decision-making, encouraging engagement (Chandon et al., 2000), it may act as a reference when comparing options (Meißner & Decker, 2010) and be a primary cue when information overload occurs (Grebitus et al., 2015). Thus, a poorly designed DCE (i.e., too many attributes; complex information) may lead to a focus on cost. Researchers should give attention to this issue when developing and piloting their DCE.

Secondly, our finding that respondents who considered the cost attribute acceptable gave it more attention highlights the importance of motivating the cost attribute (Genie et al., 2021). Gafni (1991) highlighted the importance of using payment vehicles that resemble reality and Smith (2003) noted that the most suitable payment format will depend upon the study context and differ across cultures, countries, and products. Limited guidance is provided on how the payment vehicle (cost attribute) should be defined in DCEs, with most studies providing limited information (Rowen et al., 2018). Whilst payment vehicles are context-dependent, DCE practitioners should give consideration to wording, format, and frequency. Future research could explore how different payment vehicles (e.g., monthly versus annually; taxation, charity donation, or out-of-pocket) affect choice behaviour; this will help identify best-practice for incorporating the cost attribute in health care DCEs.

Thirdly, our finding that difficulty increases visual attention raises questions about the trade-off between statistical efficiency and respondent efficiency. Research suggests statistical efficiency of a DCE, which increases difficulty, is negatively correlated with respondents' efficiency (i.e., the ability of participants to make informed decisions; Flynn et al., 2016; Viney et al., 2005). Our study suggests that increased statistical efficiency improves respondents' attention. Whether and when this positive benefit on attention breaks down (e.g., after how many choices) is an important avenue for future research.

Fourthly, our finding that later choices (task order) resulted in a reduction in fixation time may indicate participants learn how to respond as they process through the choice tasks and become more efficient in their information search (Fraser et al., 2021; Ryan et al., 2018). This is supported by our finding that mental effort was greater for earlier choice tasks and less for later choice task. This suggests warm-up choices (e.g., 2 or 3 choice tasks) may help respondents become efficient when answering the experimental tasks and that the order of choice tasks should be randomised across individuals. Finally, the scan path length was lower in the earlier tasks, suggesting a more focused information search for the first positioned tasks. Whether this occurs due to earlier choice tasks being difficult is not clear; we suggest future research explores the link between dispersion transitions and task order.

As well as providing guidance to DCE practitioners on the design of DCEs, we hope our paper stimulates discussion of the use of eye tracking in applied economic research. As Lahey & Oxley (2016) commented, research with an eye-tracker is limited only by our imagination. Possible areas for future research using eye-tracking include ANA, identifying attributes for inclusion in DCEs and hypothetical bias. Two studies in the food choice DCE literature have use eye-tracking to investigate the link between stated ANA and visual ANA: while Balcombe et al. (2015) found inconsistency between visual ANA and stated measures, Dudinskaya et al. (2020) found a more robust association. Whilst not the focus of this paper, we also explored the link between stated and visual ANA. Our results, presented in **OSM E**, are consistent with Balcombe et al. (2015), with visual ANA weakly associated with stated ANA. They raise concerns about the increasing use of debriefing questions in DCEs (Pearce et al., 2020). However, we note here that the results for our self-reported difficulty measure were consistent with our two objective measures of difficulty (entropy of transitions and DSD). Dudinskaya et al. (2020) noted that the study of eye movements could provide additional information in identifying relevant attributes for a DCE; this would be a fruitful area for future research, with eye-tracking used in the development of the DCE survey instrument. Perhaps the greatest methodological challenge facing health economists is whether and to what extent choices made in the DCE, and subsequent WTP estimates, translate to real-world settings, and how any hypothetical bias can be mitigated (Haghani et al., 2021b, 2021a). In the only study employing eye-tracking to look at hypothetical and real choices, Imai et al. (2019) showed that the more people looked at prices, and the longer they took to transition from looking to making a choice, the more likely they were to switch a hypothetical "buy" to a real "don't buy". This suggests that visual attention measured during hypothetical choices could improve prediction in real purchase decisions. An interesting area for future research is whether visual attention could be used to mitigate hypothetical bias.

Whilst offering exciting areas for future research, DCE practitioners should be aware that the environment in which an eye tracking experiment is conducted is crucial. (Nevalainen & Sajaniemi, 2004; Pernice et al., 2009) noted that environmental changes (e.g., light conditions) may result in drift and inaccurate data. Further, measures of pupil size may be influenced by the brightness of the environment and external factors (e.g., drinking a coffee before the experiment, fatigue, etc.). To control for these factors, we recruited participants during both morning and afternoon sessions (to control for fatigue) and ran the experiment in a windowless room (to control for brightness). We used a non-invasive eye-tracker such that we could ask participants to complete the DCE as normally as possible (given that most DCEs are now completed online). However, a problem we encountered was that study participants moved their head (leaned back, forward, or sideways), resulting in eyes moving out of the tracked zone. To address this issue, we re-ran the tracking calibration after each choice task. An alternative approach would have been to use a stationary seat and eye-tracker with a headrest (Krucien et al., 2017; Ryan et al., 2018). However, a more invasive eye-tracking could also place participants in a less "natural" situation and lead to changes in their choice behaviour.

Our study is not exempt from limitations. First, the act of eye tracking may influence visual attention i.e., a Hawthorne effect (Adair, 1984; McCambridge et al., 2014). This, however, is unlikely to influence our results. Although studies in social attention suggest that awareness of the recording of eye movements affects the direction of visual attention (Risko & Kingstone, 2011), these results are for objects that are socially less acceptable to be gazed at (e.g., a swimsuit calendar on the wall). No eye-tracker bias is found for neutral objects. Further, any intervention effect of an eye-tracker would be unlikely to be different across the COST and NOCOST DCE. Second, given our university-based recruitment, our sample may not be representative of the UK population, and the generalisability of our findings might be limited. With the development of more portable eye-tracking equipment (e.g., EyeTribe, Eyelink Portable Duo, Tobii Nano, Pupil labs, and Positive Science eye-tracker), future research should aim to move eye tracking research to a broader population-based sample and move from the laboratory into clinical and community settings. A previous study indicates that being familiar with an environment can affect eye movements and visual attention (Kerstin Gidlöf, Martin Lingonblad, 2015). With experience, we learn to attend to important things and ignore less relevant information (Droll et al., 2007; Meißner & Decker, 2010). Thirdly, our study focused on the effects of including a cost attribute in a health care context where whilst people have limited experience of paying for GP services, they did find this acceptable. In other DCE applications, such as preferences for new cancer treatment, a cost attribute may be 'more' unacceptable, and therefore its inclusion in the DCE may become more problematic. In developing countries, where there is limited ability to pay, the cost attribute may be more challenging. It has been suggested that the payment vehicles used in developed countries should be reconsidered for suitability when conducting DCEs in developing country contexts (Gibson et al., 2016; Hassan et al., 2018). Future research should explore the impact of the cost attribute on the fixation time, pupil dilation, and dispersion of transitions in different health care contexts and different country settings. Finally, there are only two treatments in our experiment, with a focus on the cost attribute. More treatments with less or more attributes and with and without a cost attribute would be useful to understand whether our result is specific to cost or any other attributes (e.g., risk). Given the time and cost involved in implementing ET alongside a DCE, a mouse tracker (Kieslich et al., 2019) might be a useful alternative to scale up such experiments (by including more treatments). We leave this for future research.

6 CONCLUDING REMARKS

We provide encouraging evidence for the inclusion of a cost attribute in a DCE conducted within a publicly provided health care system. Most respondents gave visual attention to the cost attribute most of the time. Average fixation time on multi-attribute tasks increased by 44% in the COST DCE, with attention given to non-monetary attributes increasing by 22%. Including cost led to more structured decision making and did not increase mental effort. Acceptability of the cost attribute of choice tasks was a predictor of cost information processing, highlighting the importance of motivating the cost attribute and including a realistic payment vehicle. Increased difficulty consistently led to increased visual attention, raising the question of when a task is too difficult.

REFERENCES

Acartürk, C., & Habel, C. (2012). Eye tracking in multimodal comprehension of graphs. *CEUR Workshop Proceedings*, *887*.

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, *69*(2). https://doi.org/10.1037/0021-9010.69.2.334

Balcombe, K., Fraser, I., & Mcsorley, E. (2015). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments. *Journal of Applied Econometrics*, *30*(3). https://doi.org/10.1002/jae.2383

Bateman, I. J., Burgess, D., Hutchinson, W. G., & Matthews, D. I. (2008). Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. *Journal of Environmental Economics and Management*, *55*(2). https://doi.org/10.1016/j.jeem.2007.08.003

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2). https://doi.org/10.1037/0033-2909.91.2.276

Bliemer, M. C. J., & Rose, J. M. (2005). Efficiency and Sample Size Requirements for Stated. *Working Paper - Institude of Transport and Logistics Studies*, *ITLS-WP-05*.

Bliemer, M. C. J., & Collins, A. T. (2016). On determining priors for the generation of efficient stated choice experimental designs. *Journal of Choice Modelling*, *21*. https://doi.org/10.1016/j.jocm.2016.03.001

Bogomolova, S., Oppewal, H., Cohen, J., & Yao, J. (2020). How the layout of a unit price label affects eye-movements and product choice: An eye-tracking investigation. *Journal of Business Research*, *111*. https://doi.org/10.1016/j.jbusres.2018.10.049

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1). https://doi.org/10.1177/2515245920960351

Bryan, S., Buxton, M., Sheldon, R., & Grant, A. (1998). Magnetic resonance imaging for the investigation of knee injuries: An investigation of preferences. *Health Economics*, *7*(7). https://doi.org/10.1002/(SICI)1099-1050(1998110)7:7<595::AID-HEC381>3.0.CO;2-E

Burton, S., Lichtenstein, D. R., Netemeyer, R. G., & Garretson, J. A. (1998). A scale for measuring attitude toward private label products and an examination of its psychological and behavioral correlates. *Journal of the Academy of Marketing Science*, *26*(4). https://doi.org/10.1177/0092070398264003

Cabeza, R., Daselaar, S. M., Dolcos, F., Prince, S. E., Budde, M., & Nyberg, L. (2004). Task-independent and Task-specific Age Effects on Brain Activity during Working Memory, Visual Attention and Episodic Retrieval. *Cerebral Cortex*, *14*(4). https://doi.org/10.1093/cercor/bhg133

Chandon, P., Wansink, B., & Laurent, G. (2000). A benefit congruency framework of sales promotion effectiveness. *Journal of Marketing*, *64*(4). https://doi.org/10.1509/jmkg.64.4.65.18071

Clark, M. D., Determann, D., Petrou, S., Moro, D., & de Bekker-Grob, E. W. (2014). Discrete Choice Experiments in Health Economics: A Review of the Literature. In *PharmacoEconomics* (Vol. 32, Issue 9). https://doi.org/10.1007/s40273-014-0170-x

Coast, J., & Horrocks, S. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of Health Services Research and Policy*, *12*(1). https://doi.org/10.1258/135581907779497602

Craig, B. M., Runge, S. K., Rand-Hendriksen, K., Ramos-Goñi, J. M., & Oppe, M. (2015). Learning and satisficing: An analysis of sequence effects in health valuation. *Value in Health*, *18*(2). https://doi.org/10.1016/j.jval.2014.11.005

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, *34*(2). https://doi.org/10.18637/jss.v034.i02

Curran, T., Hills, A., Patterson, M. B., & Strauss, M. E. (2001). Effects of aging on visuospatial attention: An ERP study. *Neuropsychologia*, *39*(3). https://doi.org/10.1016/S0028-3932(00)00112-3

Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., Scarpa, R., & Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management*, *63*(1). https://doi.org/10.1016/j.jeem.2011.09.001

de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics*, *21*(2). https://doi.org/10.1002/hec.1697

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*(1). https://doi.org/10.1006/jeem.2001.1199

Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2018). Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics (United Kingdom)*, *27*(1). https://doi.org/10.1002/hec.3564

di Nocera, F., Camilli, M., & Terenzi, M. (2006). Using the distribution of eye fixations to assess pilots' mental workload. *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/154193120605000114

Droll, J. A., Gigone, K., & Hayhoe, M. M. (2007). Learning where to direct gaze during change detection. *Journal of Vision*, *7*(14). https://doi.org/10.1167/7.14.6

Dudinskaya, E. C., Naspetti, S., & Zanoli, R. (2020). Using eye-tracking as an aid to design on-screen choice experiments. *Journal of Choice Modelling*, *36*. https://doi.org/10.1016/j.jocm.2020.100232

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? In

Developmental Cognitive Neuroscience (Vol. 25). https://doi.org/10.1016/j.dcn.2016.11.001

Essers, B. A. B., van Helvoort-Postulart, D., Prins, M. H., Neumann, M., & Dirksen, C. D. (2010). Does the inclusion of a cost attribute result in different preferences for the surgical treatment of primary basal cell carcinoma?: A comparison of two discrete-choice experiments. *PharmacoEconomics*, *28*(6). https://doi.org/10.2165/11532240-000000000-00000

Flynn, T. N., Bilger, M., Malhotra, C., & Finkelstein, E. A. (2016). Are Efficient Designs Used in Discrete Choice Experiments Too Difficult for Some Respondents? A Case Study Eliciting Preferences for End-of-Life Care. *PharmacoEconomics*, *34*(3). https://doi.org/10.1007/s40273-015-0338-z

Fraser, I., Balcombe, K., Williams, L., & McSorley, E. (2021). Preference stability in discrete choice experiments. Some evidence using eye-tracking. *Journal of Behavioral and Experimental Economics* , *94*. https://doi.org/10.1016/j.socec.2021.101753

Gafni, A. (1991). Willingness-to-pay as a measure of benefits relevant questions in the context of public decision making about health care programs. *Medical Care*, *29*(12). https://doi.org/10.1097/00005650-199112000-00007

Genie, M. G., Nicoló, A., & Pasini, G. (2020). The role of heterogeneity of patients' preferences in kidney transplantation. *Journal of Health Economics*, *72*, 102331. https://doi.org/10.1016/j.jhealeco.2020.102331

Genie, M. G., Ryan, M., & Krucien, N. (2021). To pay or not to pay? Cost information processing in the valuation of publicly funded healthcare. *Social Science and Medicine*, *276*. https://doi.org/10.1016/j.socscimed.2021.113822

Gibaldi, A., & Sabatini, S. P. (2021). The saccade main sequence revised: A fast and repeatable tool for oculomotor analysis. *Behavior Research Methods*, *53*(1). https://doi.org/10.3758/s13428-020-01388-2

Gibson, J. M., Rigby, D., Polya, D. A., & Russell, N. (2016). Discrete Choice Experiments in Developing Countries: Willingness to Pay Versus Willingness to Work. *Environmental and Resource Economics*, *65*(4). https://doi.org/10.1007/s10640-015-9919-8

Grebitus, C., Roosen, J., & Seitz, C. C. (2015). Visual attention and choice: A behavioral economics perspective on food decisions. In *Journal of Agricultural and Food Industrial Organization* (Vol. 13, Issue 1). https://doi.org/10.1515/jafio-2015-0017

Haghani, M., Bliemer, M. C. J., Rose, J. M., Oppewal, H., & Lancsar, E. (2021a). Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. In *Journal of Choice Modelling* (Vol. 41). https://doi.org/10.1016/j.jocm.2021.100309

Haghani, M., Bliemer, M. C. J., Rose, J. M., Oppewal, H., & Lancsar, E. (2021b). Hypothetical bias in stated choice experiments: Part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. In *Journal of Choice Modelling* (Vol. 41). https://doi.org/10.1016/j.jocm.2021.100322

Hahn, M., Wild-Wall, N., & Falkenstein, M. (2011). Age-related differences in performance and stimulus processing in dual task situation. *Brain Research*, *1414*. https://doi.org/10.1016/j.brainres.2011.07.051

Harrison, G. W., & Swarthout, J. T. (2019). Eye-tracking and economic theories of choice under risk. *Journal of the Economic Science Association*, *5*(1). https://doi.org/10.1007/s40881-019-00063-3

Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J., & Payne, K. (2014). Risk as an attribute in discrete choice experiments: A systematic review of the literature. In *Patient* (Vol. 7, Issue 2). https://doi.org/10.1007/s40271-014-0048-1

Hartmann, M., & Fischer, M. H. (2014). Pupillometry: The eyes shed fresh light on the mind. In *Current Biology* (Vol. 24, Issue 7). https://doi.org/10.1016/j.cub.2014.02.028

Hassan, S., Olsen, S. B., & Thorsen, B. J. (2018). Appropriate Payment Vehicles in Stated Preference Studies in Developing Economies. *Environmental and Resource Economics*, *71*(4). https://doi.org/10.1007/s10640-017-0196-6

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611). https://doi.org/10.1126/science.143.3611.1190

Hjelmgren, J., & Anell, A. (2007). Population preferences and choice of primary care models: A discrete choice experiment in Sweden. *Health Policy*, *83*(2–3). https://doi.org/10.1016/j.healthpol.2007.02.006

Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*, *27*(4). https://doi.org/10.1016/j.jhealeco.2007.11.006

Holmqvist, K. (2011). Eye tracking: A comprehensive guide to methods and measures / Kenneth Holmqvist ... [et al.]. In *Oxford University Press* (Issue January).

Hwang, J., & Lorenzen, C. L. (2008). Effective nutrition labeling of restaurant menu and pricing of healthy menu. *Journal of Foodservice*, *19*(5). https://doi.org/10.1111/j.1748-0159.2008.00108.x

Imai, T., Kang, M. J., & Camerer, C. F. (2019). When the eyes say buy: visual fixations during hypothetical consumer choice improve prediction of actual purchases. *Journal of the Economic Science Association*, *5*(1). https://doi.org/10.1007/s40881-019-00071-3

Iyamu, E., & Osuobeni, E. (2012). Age, gender, corneal diameter, corneal curvature and central corneal thickness in Nigerians with normal intra ocular pressure. *Journal of Optometry*, *5*(2). https://doi.org/10.1016/j.optom.2012.02.001

Janssen, E. M., Hauber, A. B., & Bridges, J. F. P. (2018). Conducting a Discrete-Choice Experiment Study Following Recommendations for Good Research Practices: An Application for Eliciting Patient Preferences for Diabetes Treatments. *Value in Health*, *21*(1). https://doi.org/10.1016/j.jval.2017.07.001

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, *47*(2). https://doi.org/10.1037/h0078820

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ. *Prentice-Hall, Inc., Englewood Cliffs, New Jersey.*

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756). https://doi.org/10.1126/science.154.3756.1583

Kerstin Gidlöf, Martin Lingonblad, and A. W. (2015). Effect of Familiarity on Visual Attention and Choice at the Point of Purchase. *SAGA 2015: 2nd International Workshop on Solutions for Automatic Gaze Data Analysis*, *2013*.

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-Tracking: A Practical Guide to Implementation and Analysis. In *A Handbook of Process Tracing Methods: Second Edition*.

Kjær, T., Bech, M., Gyrd-Hansen, D., & Hart-Hansen, K. (2006). Ordering effect and price sensitivity in discrete choice experiments: Need we worry? *Health Economics*, *15*(11). https://doi.org/10.1002/hec.1117

Knoepfle, D. T., Tao-yi Wang, J., & Camerer, C. F. (2009). Studying learning in games using eye-tracking. *Journal of the European Economic Association*, *7*(2–3). https://doi.org/10.1162/JEEA.2009.7.2-3.388

Korn, C. W., & Bach, D. R. (2016). A solid frame for the window on cognition: Modeling eventrelated pupil responses. *Journal of Vision*, *16*(3). https://doi.org/10.1167/16.3.28

Kragt, M. E. (2013). Stated and inferred attribute attendance models: A comparison with environmental choice experiments. *Journal of Agricultural Economics*, *64*(3). https://doi.org/10.1111/1477-9552.12032

Krucien, N., Ryan, M., & Hermens, F. (2017). Visual attention in multi-attributes choices: What can eye-tracking tell us? *Journal of Economic Behavior and Organization*, *135*. https://doi.org/10.1016/j.jebo.2017.01.018

Kruizinga, A., Mulder, B., & Waard, D. de. (2006). Eye scan patterns in a simulated ambulance dispatcher ' s task. *Human Factors*.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1). https://doi.org/10.1177/1745691611427305

Lahey, J. N., & Oxley, D. (2016). The power of eye tracking in economics experiments. *American Economic Review*, *106*(5). https://doi.org/10.1257/aer.p20161009

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2). https://doi.org/10.1086/259131

Lancsar, E., & Louviere, J. (2008). Conducting Discrete Choice Experiments to inform healthcare decision making: an overview of fundamental principles. *Pharmacoeconomics*, *26*(8).

Lockshin, L., Jarvis, W., d'Hauteville, F., & Perrouty, J. P. (2006). Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice. *Food Quality and Preference*, *17*(3–4). https://doi.org/10.1016/j.foodqual.2005.03.009

Longo, M. F., Cohen, D. R., Hood, K., Edwards, A., Robling, M., Elwyn, G., & Russell, I. T. (2006). Involving patients in primary care consultations: Assessing preferences using discrete choice experiments. *British Journal of General Practice*, *56*(522).

Louviere, J. J., Hensher, D. A., Swait, J. D., & Adamowicz, W. (2000). Stated Choice methods: analysis and application. In *Cambridge University Press* (Issue January).

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3). https://doi.org/10.1007/BF00133443

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, *20*(11). https://doi.org/10.1111/j.1467-9280.2009.02453.x

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. In *Journal of Clinical Epidemiology* (Vol. 67, Issue 3). https://doi.org/10.1016/j.jclinepi.2013.08.015

McIntosh, E. (2006). Using discrete choice experiments within a cost-benefit analysis framework: Some considerations. In *PharmacoEconomics* (Vol. 24, Issue 9). https://doi.org/10.2165/00019053-200624090-00004

Meißner, M., & Decker, R. (2010). Eye-tracking information processing in choice-based conjoint analysis. *International Journal of Market Research*, *52*(5). https://doi.org/10.2501/S147078531020151X

Monfort, S. S., Sibley, C. M., & Coyne, J. T. (2016). Using machine learning and real-time workload assessment in a high-fidelity UAV simulation environment. *Next-Generation Analyst IV*, *9851*. https://doi.org/10.1117/12.2219703

Mørkbak, M. R., Olsen, S. B., & Campbell, D. (2014). Behavioral implications of providing real incentives in stated choice experiments. *Journal of Economic Psychology*, *45*. https://doi.org/10.1016/j.joep.2014.07.004

Murchie, P., Norwood, P. F., Pietrucin-Materek, M., Porteous, T., Hannaford, P. C., & Ryan, M. (2016). Determining cancer survivors' preferences to inform new models of follow-up care. *British Journal of Cancer*, *115*(12). https://doi.org/10.1038/bjc.2016.352

Murray, N. P., Hunfalvay, M., & Bolte, T. (2017). The Reliability, Validity, and Normative Data of Interpupillary Distance and Pupil Diameter Using Eye-Tracking Technology. *Translational Vision Science & Technology*, *6*(4). https://doi.org/10.1167/tvst.6.4.2

Nevalainen, S., & Sajaniemi, J. (2004). Comparison of Three Eye Tracking Devices in Psychology of Programming Research. *Proceedings of 16th Annual Workshop of the Psychology of Programming Interest Group (PPIG'04), April*.

Ngan, H. F. B., Bavik, A., Kuo, C. F., & Yu, C. E. (2022). Where you Look Depends on What you are Willing to Afford: Eye Tracking in Menus. *Journal of Hospitality and Tourism Research*, *46*(1). https://doi.org/10.1177/1096348020951226

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4). https://doi.org/10.3758/s13428-017-0998-z

Pearce, A. M., Mulhern, B. J., Watson, V., & Viney, R. C. (2020). How Are Debriefing Questions Used in Health Discrete Choice Experiments? An Online Survey. *Value in Health*, *23*(3). https://doi.org/10.1016/j.jval.2019.10.001

Pedersen, L. B., Kjær, T., Kragstrup, J., & Gyrd-Hansen, D. (2011). Does the inclusion of a cost attribute in forced and unforced choices matter? Results from a web survey applying the discrete choice experiment. *Journal of Choice Modelling*, *4*(3). https://doi.org/10.1016/S1755-5345(13)70044-7

Pernice, K., Nielsen, J., Usability, E., & Using, S. (2009). Eyetracking Methodology: How to Conduct and Evaluate Usability Studies Using Eyetracking. *Notes*, *August*.

Pesce, C., Guidetti, L., Baldari, C., Tessitore, A., & Capranica, L. (2005). Effects of aging on visual attentional focusing. *Gerontology*, *51*(4). https://doi.org/10.1159/000085123

Ratcliffe, J. (2000). The use of conjoint analysis to elicit willingness-to-pay values: Proceed with caution? *International Journal of Technology Assessment in Health Care*, *16*(1). https://doi.org/10.1017/S0266462300161227

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3). https://doi.org/10.1037/0033-2909.124.3.372

Rio, D., Woog, K., & Legras, R. (2016). Effect of age, decentration, aberrations and pupil size on subjective image quality with concentric bifocal optics. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *36*(4). https://doi.org/10.1111/opo.12300

Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: Implied social presence, eye tracking and attention. *Attention, Perception, and Psychophysics*, *73*(2). https://doi.org/10.3758/s13414-010-0042-1

Rowen, D., Stevens, K., Labeit, A., Elliott, J., Mulhern, B., Carlton, J., Basarir, H., Ratcliffe, J., & Brazier, J. (2018). Using a Discrete-Choice Experiment Involving Cost to Value a Classification System Measuring the Quality-of-Life Impact of Self-Management for Diabetes. *Value in Health*, *21*(1). https://doi.org/10.1016/j.jval.2017.06.016

Rubin, G., Bate, A., George, A., Shackley, P., & Hall, N. (2006). Preference for access to the GP: A discrete choice experiment. *British Journal of General Practice*, *56*(531).

Ryan, M., Krucien, N., & Hermens, F. (2018). The eyes have it: Using eye tracking to inform information processing strategies in multi-attributes choices. *Health Economics (United Kingdom)*, *27*(4). https://doi.org/10.1002/hec.3626

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, *103*(3). https://doi.org/10.1037/0033-295X.103.3.403

Scarpa, R., Gilbride, T. J., Campbell, D., & Hensher, D. A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*, *36*(2). https://doi.org/10.1093/erae/jbp012

Selivanova, A., & Krabbe, P. F. M. (2018). Eye tracking to explore attendance in health-state descriptions. *PLoS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0190111

Sever, I., Verbič, M., & Klarić Sever, E. (2019). Cost attribute in health care DCEs: Just adding another attribute or a trigger of change in the stated preferences? *Journal of Choice Modelling*, *32*. https://doi.org/10.1016/j.jocm.2018.03.005

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shic, F., Chawarska, K., Bradshaw, J., & Scassellati, B. (2008). Autism, eye-tracking, entropy. *2008 IEEE 7th International Conference on Development and Learning, ICDL*. https://doi.org/10.1109/DEVLRN.2008.4640808

Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, *7*(2). https://doi.org/10.1086/208799

Sillero-Rejon, C., Mahmoud, O., Tamayo, R. M., Clavijo-Alvarez, A. A., Adams, S., & Maynard, O. M. (2022). Standardised packs and larger health warnings: visual attention and perceptions among Colombian smokers and non-smokers. *Addiction*, *117*(6). https://doi.org/10.1111/add.15779

Smith, R. D. (2003). Construction of the contingent valuation market in health care: A critical assessment. In *Health Economics* (Vol. 12, Issue 8). https://doi.org/10.1002/hec.755

Smith, R. L. (1984). Sequential Treatment Allocation Using Biased Coin Designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, *46*(3). https://doi.org/10.1111/j.2517-6161.1984.tb01323.x

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1). https://doi.org/10.1037/1082-989X.11.1.54

Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete Choice Experiments in Health Economics: Past, Present and Future. In *PharmacoEconomics*. https://doi.org/10.1007/s40273-018-0734-2

Spinks, J., & Mortimer, D. (2016). Lost in the crowd? Using eye-tracking to investigate the effect of complexity on attribute non-attendance in discrete choice experiments Clinical decision-making, knowledge support systems, and theory. *BMC Medical Informatics and Decision Making*, *16*(1). https://doi.org/10.1186/s12911-016-0251-1

Spooner, J. W., Sakala, S. M., & Baloh, R. W. (1980). Effect of Aging on Eye Tracking. *Archives of Neurology*, *37*(9). https://doi.org/10.1001/archneur.1980.00500580071012

SPROTLES, G. B., & KENDALL, E. L. (1986). A Methodology for Profiling Consumers' Decision-Making Styles. *Journal of Consumer Affairs*, *20*(2). https://doi.org/10.1111/j.1745-6606.1986.tb00382.x

Szinay, D., Cameron, R., Naughton, F., Whitty, J. A., Brown, J., & Jones, A. (2021). Understanding uptake of digital health products: Methodology tutorial for a discrete choice experiment using the bayesian efficient design. *Journal of Medical Internet Research*, *23*(10). https://doi.org/10.2196/32365

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, *46*(12). https://doi.org/10.1016/j.visres.2005.12.005

Tinelli, M., Ryan, M., & Bond, C. (2016). What, who and when? Incorporating a discrete choice experiment into an economic evaluation. *Health Economics Review*, *6*(1). https://doi.org/10.1186/s13561-016-0108-4

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. In *Psychonomic Bulletin and Review* (Vol. 25, Issue 6). https://doi.org/10.3758/s13423-018-1432-y

Vass, C., Rigby, D., Tate, K., Stewart, A., & Payne, K. (2018). An Exploratory Application of Eye-Tracking Methods in a Discrete Choice Experiment. *Medical Decision Making*, *38*(6). https://doi.org/10.1177/0272989X18782197

Viney, R., Savage, E., & Louviere, J. (2005). Empirical investigation of experimental design properties of discrete choice experiments in health care. *Health Economics*, *14*(4). https://doi.org/10.1002/hec.981

Whitaker, K. L., Ghanouni, A., Zhou, Y., Lyratzopoulos, G., & Morris, S. (2017). Patients' preferences for GP consultation for perceived cancer risk in primary care: A discrete choice experiment. *British Journal of General Practice*, *67*(659). https://doi.org/10.3399/bjgp17X690905

Winn, B., Whitaker, D., Elliott, D. B., & Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology and Visual Science*, *35*(3).

Youn, H., & Kim, J. H. (2017). Effects of ingredients, names and stories about food origins on perceived authenticity and purchase intentions. *International Journal of Hospitality Management*, *63*. https://doi.org/10.1016/j.ijhm.2017.01.002

List of Tables

Table 1. Attributes and levels used to describe a GP appointment

Attributes	Definition	Levels			
		1	2	3	4
Flexibility	I can choose the time that suits me	No	Yes	-	-
Waiting time	Number of days I have to wait before the appointment	4 days	2 days	1 day	Same day
Continuity	I can choose the doctor I want to see	No	Yes	-	-
Length	Duration of the consultation	10 minutes	15 minutes	20 minutes	-
Cost*	The amount I have to pay at the end of the consultation	£30	£20	£10	£0

* In defining the cost attribute respondents were told: *"We are interested in how you would value a GP appointment. One way of doing this is to ask about the amount of money you would be willing to pay for a GP appointment. In the choice questions that follow, each GP appointment has a cost. Please assume that cost of a GP appointment is not covered by the NHS so you would have to pay this amount."*

Table 2. Respondents' characteristics and responses

	COST (N=55)[1]	NOCOST (N=60)	p-value
AGE			0.426[2]
Mean (SD)	35.06 (11.25)	37 (14.36)	
GENDER			0.748[3]
Male	19	18	
Female	36	42	
EXPERIENCE (*do you have experience paying for GP appointments?*)			0.934[3]
Yes	7	9	
No	48	51	
Cost ACCEPTABILITY (*do you find information about the cost of GP appointments acceptable?*)			0.544[3]
Completely acceptable	8	16	
Acceptable	33	25	
Not acceptable	11	11	
Not acceptable at all	3	8	
Task DIFFICULTY (*how did you find making choices between appointment options?*)[4]			0.140[3]
Very easy	13	14	
Easy	34	44	
Difficult	8	2	
Very difficult	0	0	

1. Due to technical issues, personal data for five respondents in the COST DCE was not recorded
2. T-test of mean equality; 3. Pearson's Chi-squared test with Yates' continuity correction; 4. Elicited after all 12 choices.

Table 3. Beta regression of the share of visual attention on the cost attribute

Effects (reference level)	Share of Visual Attention on cost attribute
1. Model parameters	
ACCEPTABILITY (not acceptable)	0.270 (0.080) ***
DIFFICULTY (Entropy of transitions)	2.522 (0.291) ***
BLOCK-1 (BLOCK-2)	-0.020 (0.078)
BLOCK-3 (BLOCK-2)	0.024 (0.079)
AGE	0.010 (0.003) ***
GENDER (female)	-0.235 (0.071) ***
Constant	-3.527 (0.225) ***
Precision	10.375 (0.632) ***
2. Model diagnostics	
Log-Likelihood	-579.782
Number of observations	611
Number of respondents	51

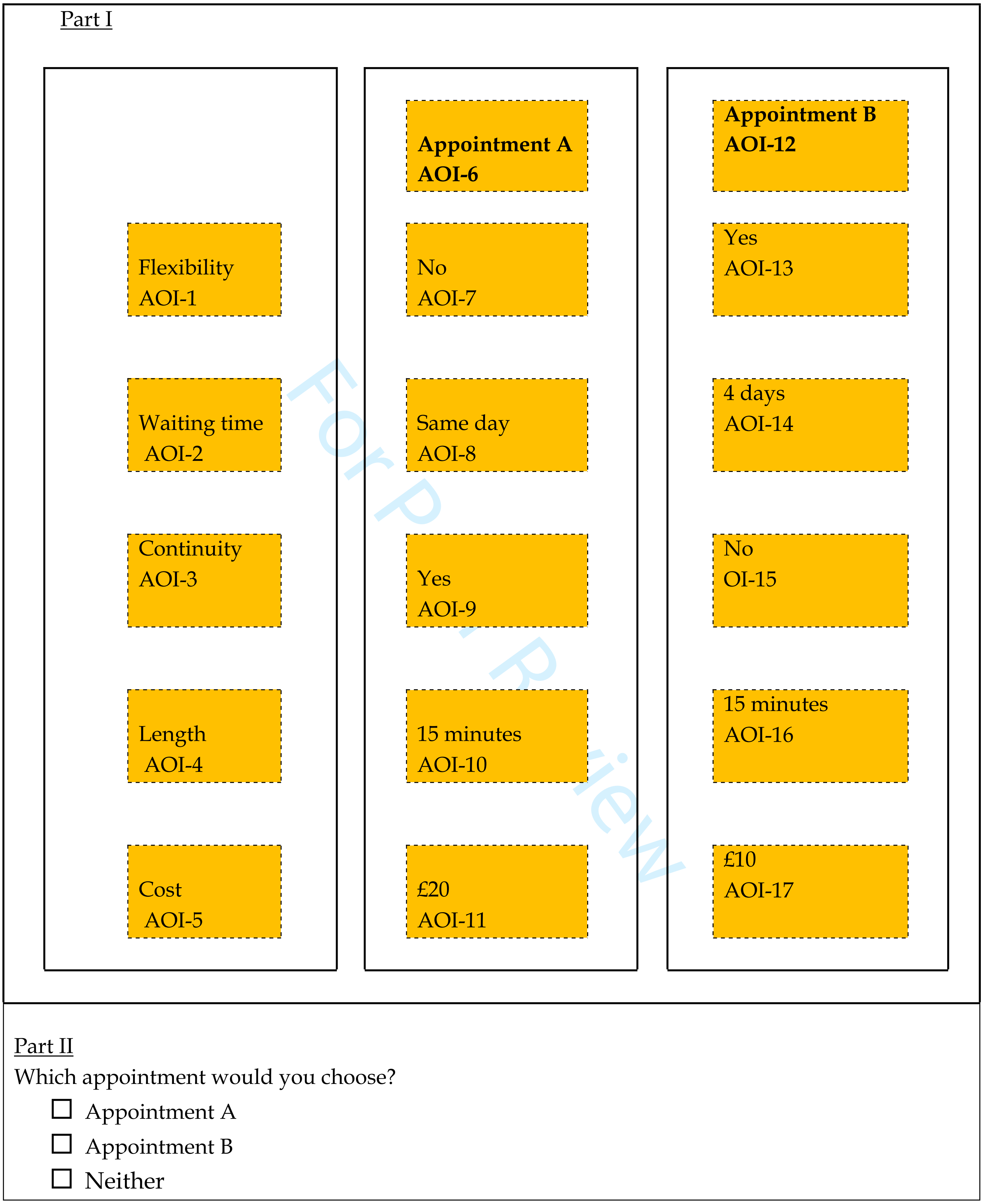
**** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis.*

Table 4. Linear mixed-effect regression results of eye-tracking metrics

Effects (reference level)	Fixation time	Dispersion of transitions (scan path length)	Change in pupil size (mental effort)
1. Model parameters			
EXPERIMENT (NOCOST)	0.311 (0.087) ***	-0.477 (0.035) ***	0.016 (0.03)
ACCEPTABILITY (not acceptable)	0.058 (0.097)	-0.045 (0.038)	-0.041 (0.033)
DIFFICULTY – Entropy of transitions	1.002 (0.15) ***	-0.005 (0.097)	-0.015 (0.023)
BLOCK-1 (Block-2)	0.037 (0.034)	-0.052 (0.023) **	0.029 (0.005) ***
BLOCK-3 (Block-2)	-0.071 (0.034) **	0.007 (0.023)	-0.012 (0.005) ***
AGE	-0.004 (0.004)	0.003 (0.001) *	0.0004 (0.001)
GENDER (female)	0.134 (0.092)	-0.031 (0.036)	0.023 (0.032)
Constant	7.571 (0.182) ***	1.925 (0.083) ***	-0.075 (0.057)
Individual errors	0.402	0.138	0.144
Observation errors	0.482	0.324	0.064
2. Model diagnostics			
Log-Likelihood	-926.260	-402.136	-1364.891
Number of observations	1246	1243	1246
Number of respondents	104	104	104

**** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis.*

Figure 1. Example of a choice task – COST DCE



Part I

	Appointment A **AOI-6**	**Appointment B** **AOI-12**
Flexibility AOI-1	No AOI-7	Yes AOI-13
Waiting time AOI-2	Same day AOI-8	4 days AOI-14
Continuity AOI-3	Yes AOI-9	No OI-15
Length AOI-4	15 minutes AOI-10	15 minutes AOI-16
Cost AOI-5	£20 AOI-11	£10 AOI-17

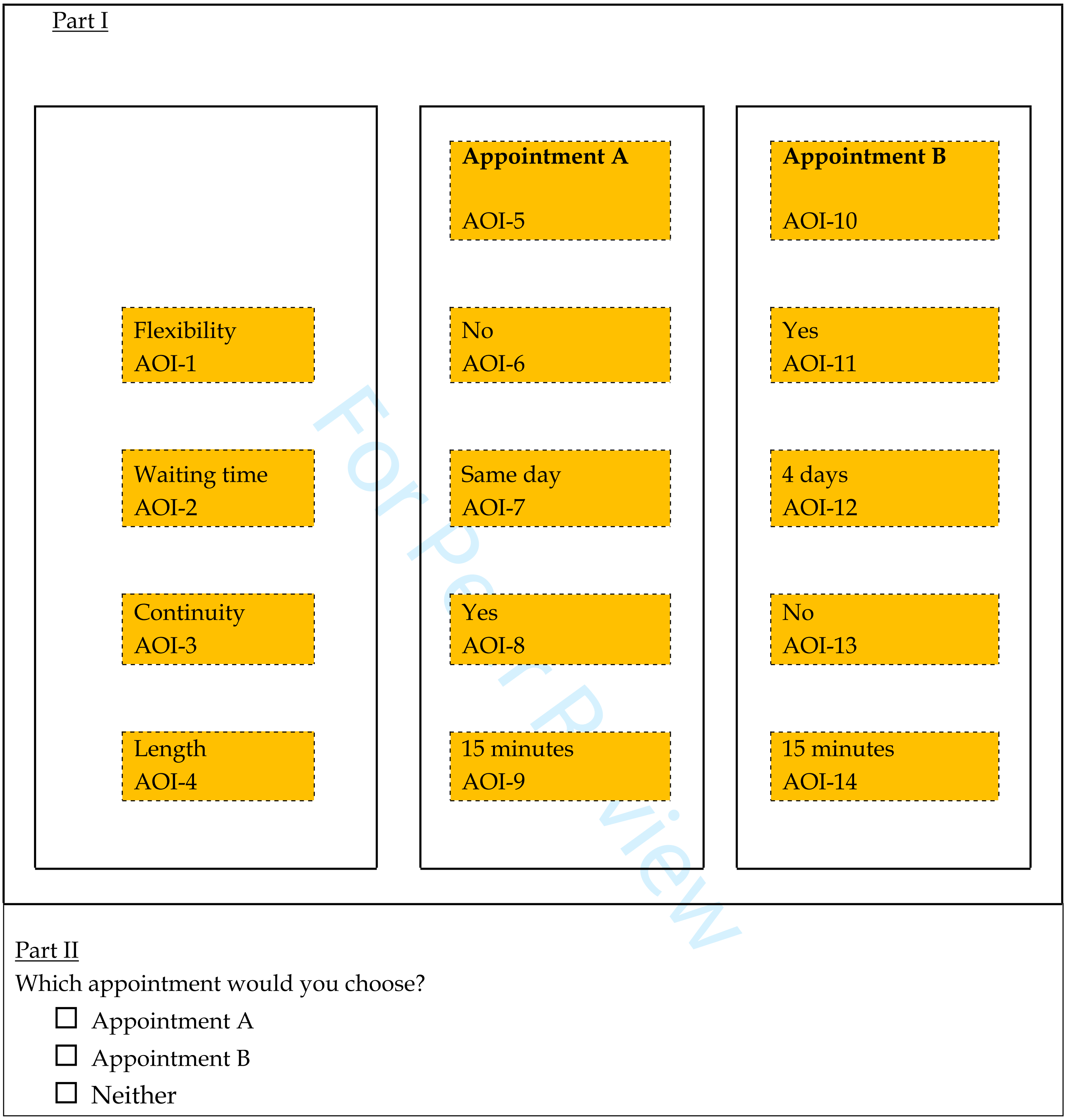
Part II

Which appointment would you choose?

- ☐ Appointment A
- ☐ Appointment B
- ☐ Neither

Highlighted squares indicated the Areas of Interest (AOI) for analysis and were not shown to respondents during the experiment.

Figure 2. Example of a choice task – NOCOST DCE



Part I

	Appointment A AOI-5	**Appointment B** AOI-10
Flexibility AOI-1	No AOI-6	Yes AOI-11
Waiting time AOI-2	Same day AOI-7	4 days AOI-12
Continuity AOI-3	Yes AOI-8	No AOI-13
Length AOI-4	15 minutes AOI-9	15 minutes AOI-14

Part II

Which appointment would you choose?

☐ Appointment A
☐ Appointment B
☐ Neither

Highlighted squares indicated the Areas of Interest (AOI) for analysis and were not shown to respondents during the experiment.

Figure 3. Average fixation times across choice tasks

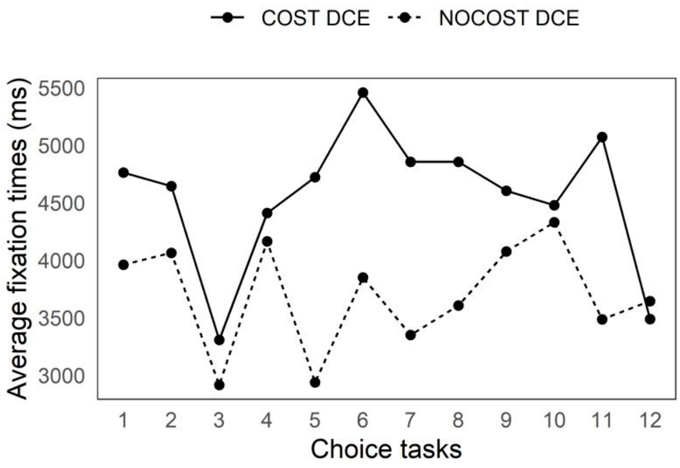


Figure 4. Comparison of average fixation times across attributes

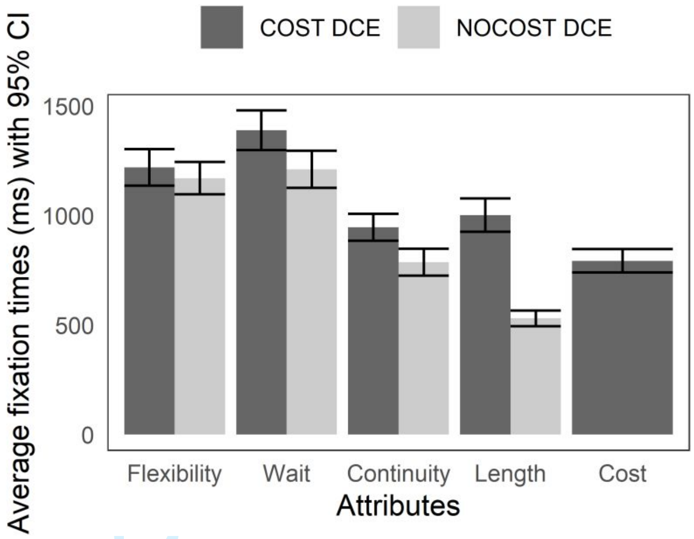


Figure 5. Average dispersion of transitions across choice tasks

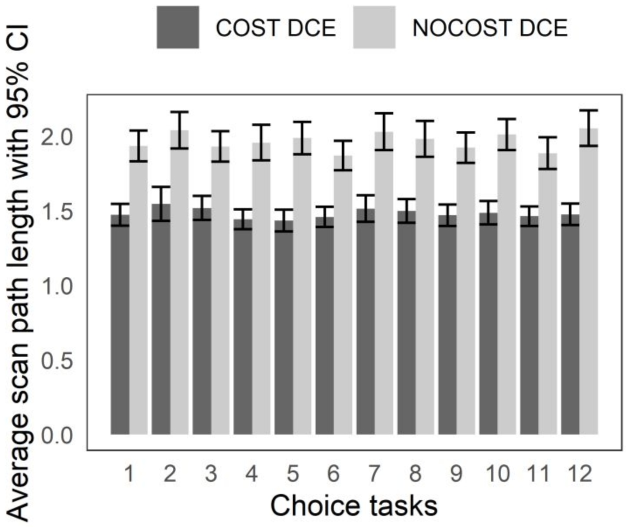
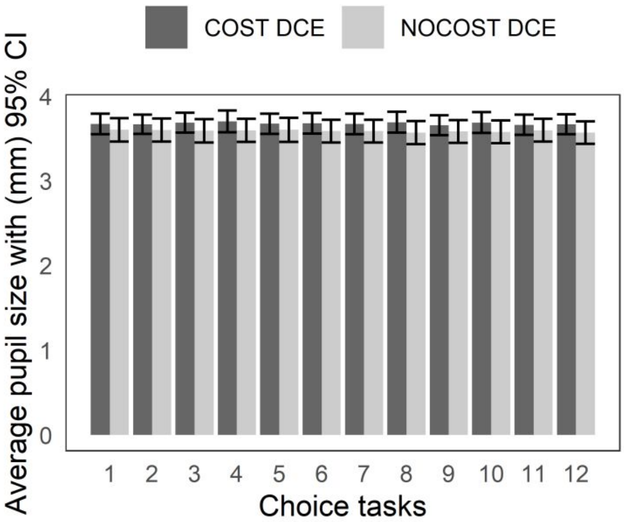


Figure 6. Changes in average pupil size across choice tasks



Online Supplementary Material A

Table A1. Studies using eye-tracking technology alongside discrete choice experiments in health

Author (Year)	Context	Sample/Country	Research question	Visual Attention Eye-tracking metrics
Spinks and Mortimer (2016)	Medicine for health conditions (insomnia or joint pain)	32 University Staff/Australia	Do consumers process additional information in an already complex decision problem/attribute non-attendance (ANA)?	Fixations (area, duration and order)
Krucien et al. (2017)	Health & lifestyle programmes	58 students/UK	Does visual attention (VA) influence processing multi-attribute information, and how does this relate to individuals' choices?	Fixations (time)
Ryan et al. (2018)	Health & lifestyle programmes	58 students/UK	How do respondents interact with choice tasks?	Fixations (time and transition)
Selivanova (2018)	Health state descriptions	10 public, PhD, Master students/Netherlands	How do respondents attend to various information cues presented in DCEs?	Fixations (time, paths) Heat maps
Vass et al. (2018)	Breast cancer screening	40 women from the public/UK	Do different ways of presenting risk influence VA and processing strategies?	Pupil size Direction of motion Dwell time
(Sillero-Rejon et al., 2022)	Cigarette packaging and health warning size	175 public/Colombia	How does information provided affect VA and preferences?	Fixations (number)

- **Saccade** is a quick, concurrent movement of eyes between two phases of fixation in the same direction (no encoding takes place during saccades).
- **Fixation** refers to the duration of time during which the eyes are at a halt for encoding data: **Number of fixations** refer to how many times a participant has fixated a particular area of interest (AOI) and **Duration of fixations** refer to for how long a participant has fixated a particular AOI.
- **Heatmaps** visualise the intensity levels where the participant is looking at the stimulus, described by a colour scale comprised of three colours: green (areas with lower fixation intensities), yellow (areas with higher fixation intensities), and red (areas with very high fixation intensities).
- **Dwell time** refers to the total amount of time spent looking within an AOI. This includes all fixations and saccades within the AOI, including revisits.

Online Supplementary Material B

Choice Model Results

Table B1. Results of MNL models

	COST		NOCOST
1. Preference parameters		**WTP (£)**	
Flexibility (Ref: No)	0.312 (0.079) ***	15.60	0.445 (0.081) ***
Waiting time	-0.075 (0.027) ***	3.75	-0.011 (0.028)
Continuity (Ref: No)	0.316 (0.079) ***	15.80	0.530 (0.081) ***
Length of consultation	0.059 (0.007) ***	2.95	0.062 (0.007) ***
Cost	-0.020 (0.004) ***	-	-
3. Model diagnostics			
LL at convergence	-623.651		-600.357
Number of observations	720		720
Number of parameters	5		4

**** indicate significance at 1%. Standard errors in parentheses.*

Online Supplementary Material C

Examples of information search pattern

Figure C1 Structured information search pattern

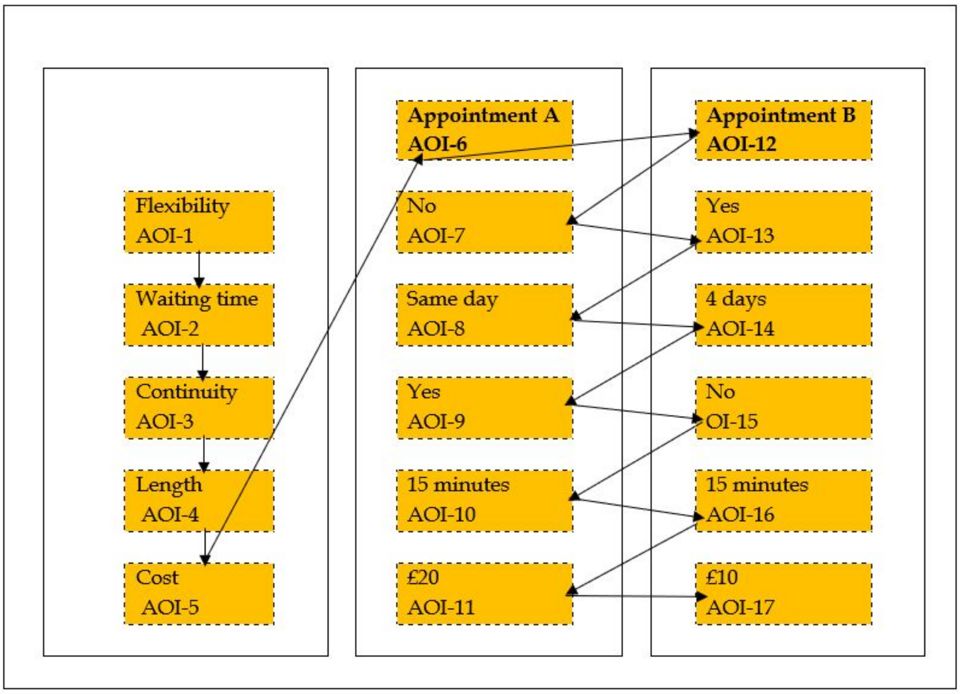
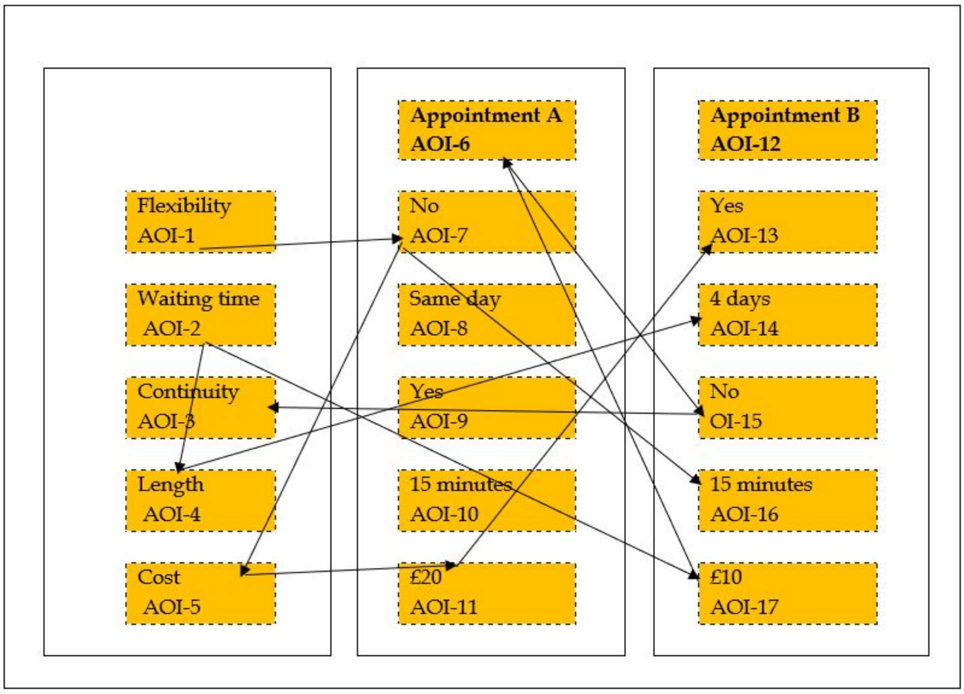


Figure C2 Less structured information search pattern



Online Supplementary Material D
Using different measures of choice task difficulty

Table D1. Beta regression of the share of visual attention on the cost attribute

Effects (reference level)	**Share of Visual Attention on cost attribute**	**Share of Visual Attention on cost attribute**
1. Model parameters		
ACCEPTABILITY (not acceptable)	0.260 (0.087) ***	0.319 (0.084) ***
Perceived DIFFICULTY (not difficult)	**0.287 (0.114)** **	**-**
DIFFICULTY - DSD	-	**0.23 (0.027)** ***
BLOCK-1 (BLOCK-2)	-0.032 (0.083)	-0.021 (0.084)
BLOCK-3 (BLOCK-2)	-0.030 (0.083)	-0.031 (0.083)
AGE	0.001 (0.004)	0.006 (0.004)
GENDER (female)	-0.256 (0.076) ***	-0.251 (0.075) ***
Constant	-1.623 (0.243) ***	-2.016 (0.161) ***
Precision	8.809 (0.536) ***	8.740 (0.532) ***
2. Model diagnostics		
Log-Likelihood	-536.568	-534.813
Number of observations	611	611
Number of respondents	51	51

**** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. DSD = dispersion of standard deviation*

Table D2. Linear mixed-effect regression results of eye-tracking metrics - difficulty measured as perceived difficulty

Effects (reference level)	Fixation time	Dispersion of transitions (scan path length)	Change in pupil size (mental effort)
1. Model parameters			
EXPERIMENT (NOCOST)	0.317 (0.1) ***	-0.479 (0.035) ***	0.012 (0.03)
ACCEPTABILITY (not acceptable)	0.033 (0.109)	-0.046 (0.038)	-0.042 (0.033)
Perceived DIFFICULTY (not difficult)	0.646 (0.199) ***	0.018 (0.069)	0.022 (0.06)
BLOCK-1 (Block-2)	0.034 (0.034)	-0.052 (0.023) **	0.029 (0.005) ***
BLOCK-3 (Block-2)	-0.075 (0.034) **	0.007 (0.023)	-0.012 (0.005) ***
AGE	-0.008 (0.004) **	0.003 (0.001)	0.0003 (0.001)
GENDER (female)	0.164 (0.104)	-0.03 (0.036)	0.023 (0.032)
Constant	8.88 (0.3) ***	1.944 (0.105) ***	-0.057 (0.091)
Individual errors	0.457	0.137	0.144
Observation errors	0.483	0.324	0.064
2. Model diagnostics			
Log-Likelihood	-942.96	-402.10	-1369.01
Number of observations	1247	1243	1243
Number of respondents	104	104	104

**** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis.*

Table D3. Linear mixed-effect regression results of eye-tracking metrics – difficulty measured as dispersion of standard deviation (DSD)

Effects (reference level)	Fixation time	Dispersion of transitions (scan path length)	Change in pupil size (mental effort)
1. Model parameters			
EXPERIMENT (NOCOST)	0.311 (0.087) ***	-0.477 (0.035) ***	0.016 (0.03)
ACCEPTABILITY (not acceptable)	0.058 (0.097)	-0.045 (0.038)	-0.041 (0.033)
DIFFICULTY - DSD	**-0.07 (0.011) *****	**0.001 (0.007)**	**0.001 (0.001)**
BLOCK-1 (Block-2)	0.037 (0.034)	-0.052 (0.023) **	0.029 (0.005) ***
BLOCK-3 (Block-2)	-0.071 (0.034) **	0.007 (0.023)	-0.012 (0.005) ***
AGE	-0.004 (0.004)	0.003 (0.001) *	0.0004 (0.001)
GENDER (female)	0.134 (0.092)	-0.031 (0.036)	0.023 (0.032)
Constant	7.571 (0.182) ***	1.925 (0.083) ***	-0.075 (0.057)
Individual errors	0.402	0.138	0.144
Observation errors	0.482	0.324	0.064
2. Model diagnostics			
Log-Likelihood	-947.352	-402.129	-1368.997
Number of observations	1246	1243	1246
Number of respondents	104	104	104

**** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. DSD = dispersion of standard deviation*

Online supplementary material E
Stated ANA versus Visual ANA

Stated attribute non-attendance (SANA)

We asked respondents the following question at the end of the choice tasks:

Please could you tell which features of GP appointments you have never considered in your choices?

Attributes were listed in the same order as in the DCE. We coded responses into a dummy variable: stated ANA=1 when respondents never considered a given attribute in their 12 choices and stated ANA=0 otherwise.

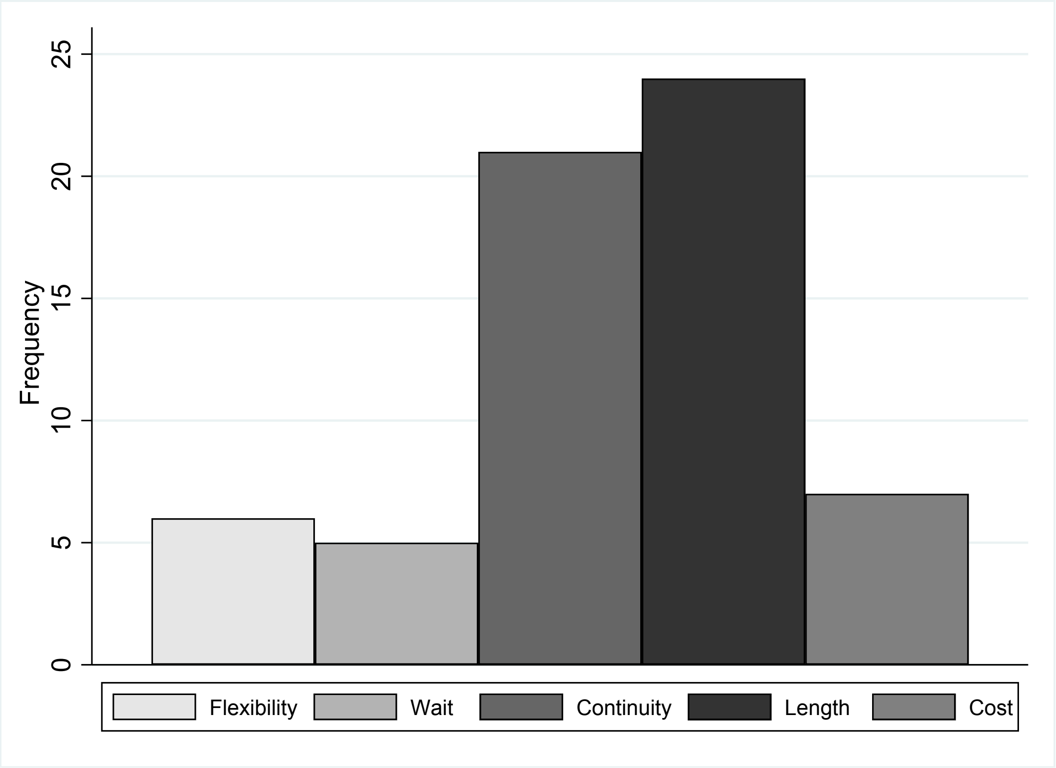
The SANA responses are summarised in Table E1. Of the 57 respondents who provided information, 12 (21.00%) stated that they attended all the attributes. When reaching their choice decisions, 16 (28.06%) participants indicated that they ignored two attributes and one individual indicated that they ignored three attributes.

Table E1. Stated attributes non-attendance

Attributes and combinations of attributes never considered	Number of respondents	%
None	12	21.05
Ignored one attribute		
Flexibility	4	7.02
Waiting time	3	5.26
Continuity	9	15.79
Length	10	17.54
Cost	2	3.51
Ignored two attributes		
Flexibility, Continuity	1	1.75
Waiting time, Continuity	1	1.75
Waiting time, Cost	1	1.75
Continuity, Length	9	15.79
Length, Cost	4	7.02
Ignored three attributes		
Flexibility, Continuity, Length	1	1.75
Total	57	100.00

Figure E1 shows stated ANA for each attribute. Of the total cases of stated ANA across all attributes, the most ignored attributes are consultation length (24 cases) and continuity (21 cases). The least stated attribute was ‘waiting time’ at around 8.8% (5/57) non-attended. The cost attribute was stated to be ignored by 12.28% (7/57) of respondents, indicating that the majority, 87.71% (50/57), considered cost in their choice decisions.

Figure E1. Frequency of stated attribute non-attendance



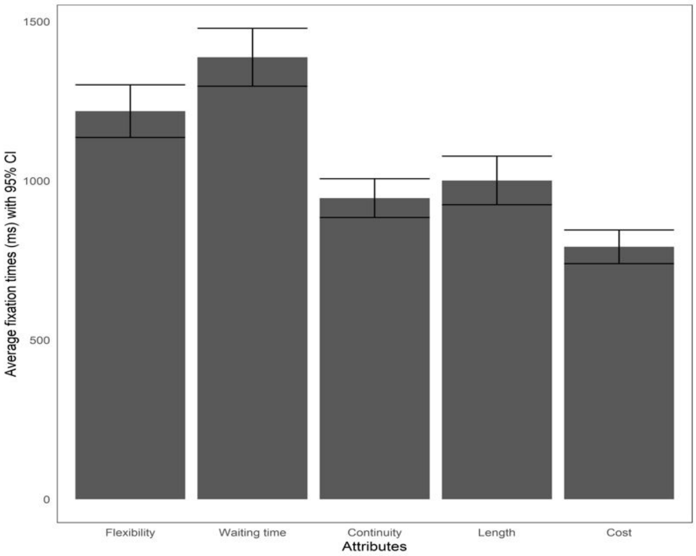
Visual attribute non-attendance (VANA)

We used fixation time to measure visual attribute non-attendance (VANA). An attribute is considered visually ignored if participants did not fixate on the levels of that attribute (i.e., duration of fixations is zero). Table E2 summarises VANA for each attribute. Removing individuals with poor eye-tracking data and participants without at least one fixation for every experimental task resulted in 51 respondents. Of the 3,055 observations (50 participants x 12 tasks x 5 attributes + 1 participant x 11 tasks x 5 attributes = 3,055), we observed 179 (5.9%) cases of VANA across the five attributes: flexibility, 42 (24%); waiting time, 21 (11.7%); continuity, 31 (17.3%); length of consultation, 35 (19.6%); cost, 49 (27.4%). Most respondents paid attention to most of the attributes (94.1%). V ANA was not uniformly distributed across the attributes (Chi-2 = 13.09; p<0.01). From Figure E2, we note that the first presented attributes (i.e., flexibility and waiting time) have the highest levels of visual attention. Second, the cost attribute is the lowest visually attended attribute – i.e., it has low attention relative to other attributes. Hence, the average fixation time and, therefore, our measure of visual attention varies across attributes (mean ranging from 792.796 ms for cost attribute to 1387.962 ms for waiting time) (Figure E2).

Table E2. Visual attribute non-attendance

Attributes	Cases of Visual ANA	% of visual ANA	Total observations (%)
Flexibility	42	24%	611 (7%)
Waiting time	21	12%	611 (3%)
Continuity	31	17%	611 (5%)
Length of consultation	35	20%	611 (6%)
Cost	49	27%	611 (8%)
Total (%)	179	100%	3055 (6%)

Figure E2. Average fixation times across attributes



Relationship between SANA and VANA

Table E3 shows raw data of VANA and SANA by each participant. Most respondents fixated on every attribute at least once whilst completing the 12 choices. For example, individual (ID1) visually attended to flexibility in 11 of the 12 choices; waiting time in all 12 choices; continuity 8 of the 12 choices; length 11 of the 12 choices; and cost 1 of 12 choices. There is evidence that respondents visually ignored attributes when they stated they attended them. For example, the VANA data suggests that ID1 visually ignored cost on eleven out of the twelve choice tasks, whilst the SANA data indicates individuals never ignored cost. Similarly, for ID2, the visual ANA data suggests cost visual ANA occurs on three occasions/decisions, whilst the stated ANA data indicates ID2 never ignored cost. Further, the opposite holds: some respondents stated they ignored attributes when they visually attended to them. For instance, for ID5, the visual ANA data indicates that the respondent visually attended the cost attribute whilst the stated ANA shows ID5 ignored the cost attribute. Interestingly, six respondents who stated that they attended to all the attributes also visually attended all the attributes in the experiment. However, for the majority of the respondents, we find divergence between stated ANA and visual ANA.

Table E3. Raw data comparing stated ANA and visual ANA

	Visual ANA					Stated ANA				
ID	**Flexibility**	**Waiting time**	**Continuity**	**Length**	**Cost**	**Flexibility**	**Waiting time**	**Continuity**	**Length**	**Cost**
1	1	0	4	1	11	0	0	0	0	0
2	7	1	1	3	3	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0
4	0	0	0	0	1	0	0	0	1	1
5	1	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	1	1	0
9	1	2	1	1	0	0	0	0	1	1
10	0	0	0	0	0	1	0	0	0	0
11	0	2	2	2	3	0	0	0	1	0
12	0	0	1	2	0	0	0	0	0	0
13	0	1	0	0	0	0	0	0	1	0
14	0	0	1	0	0	1	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	1	0	0
17	0	0	0	0	0	0	1	0	0	0
18	0	0	0	0	0	0	0	1	1	0
19	0	0	1	0	0	0	0	1	1	0
20	0	0	0	0	0	0	0	1	0	0
21	0	0	0	0	0	0	0	0	0	0
22	1	3	5	10	7	0	0	1	1	0
23	9	3	0	2	0	0	0	1	1	0
24	0	1	0	0	0	0	0	0	0	1
25	0	0	1	0	0	0	0	1	0	0
26	1	0	0	0	1	0	0	1	0	0
27	0	0	0	0	0	0	0	0	1	0
28	2	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	1	0
30	0	0	0	0	0	1	0	1	0	0
31	0	0	2	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	1	1	0
33	0	1	1	1	1	0	0	0	0	0
34	0	0	2	0	1	0	0	1	0	0
35	0	0	0	0	0	0	0	0	1	0
36	1	0	0	0	1	0	0	0	1	0
37	0	0	0	0	0	0	0	0	0	0
38	1	0	2	2	0	0	0	0	1	0
39	0	0	0	0	0	0	0	0	1	1

40	0	0	1	0	0	0	0	1	0	0
41	0	0	0	0	0	0	0	0	0	0
42	1	0	0	0	0	0	1	0	0	1
43	2	2	1	2	2	0	1	0	0	0
44	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	1	1
47	0	0	0	0	0	1	0	1	1	0
48	0	0	0	0	0	0	0	1	1	0
49	5	0	2	2	5	0	0	0	1	0
50	0	0	1	0	0	1	0	0	0	0
51	0	0	0	0	0	0	0	1	1	0
52	0	0	0	0	0	0	1	0	0	0
53	1	0	0	1	0	0	0	1	1	0
54	5	4	0	6	11	0	0	0	1	0
55	0	0	0	0	0	0	0	0	1	0
56	0	1	2	0	2	0	0	1	0	0
57	4	0	0	0	0	0	1	1	0	0