

Article

Visual Markers of Hesitation in Multi-Alternative Product Selection: An Eye-Tracking Study

Vicente Bayarri-Porcar ¹, Maria-Jesus Agost ², Joaquín L. Sancho-Bru ^{2,*} and Margarita Vergara ²¹ Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano (Human-Tech), Universitat Politècnica de València, 46022 Valencia, Spain; vbayarri@upv.es² Department of Mechanical Engineering and Construction, Universitat Jaume I, 12071 Castelló, Spain; magost@uji.es (M.-J.A.); vergara@uji.es (M.V.)

* Correspondence: sancho@uji.es

Abstract

Product selection often requires visual comparison between multiple alternatives, and hesitation during this process remains less understood than the prediction of final choice. This study investigated whether hesitation during product evaluation could be characterized using eye-tracking metrics in a task involving six jar-opening devices evaluated for robustness, comfort, and effort level required. Sixty-seven participants completed a visual selection task in which each trial was classified according to participants' self-reported hesitation after each question. Analyses examined first-fixation behavior, global trial-level eye-tracking metrics, attention allocation across grouped alternatives within hesitation trials, pupil diameter, and qualitative written reflections. Hesitation trials were associated with longer decision times, more saccades, and greater total saccadic amplitude. Furthermore, early orienting also differed, with first fixations occurring earlier in hesitation trials. Within hesitation trials, attention was not concentrated on the selected option; candidate alternatives were the most intensively inspected options when attention was considered per individual model. By contrast, baseline-corrected pupil diameter did not significantly differentiate hesitation from no-hesitation trials. These findings suggest that hesitation can be characterized through a structured pattern of visual behavior involving early attention shifts, more extensive exploration, and attention redistributed across competing alternatives.

Keywords: eye-tracking; decision-making; hesitation; gaze patterns; product selection

1. Introduction

Decision-making in product selection often involves the comparison of multiple competing alternatives, a process that may give rise to uncertainty and hesitation before a final choice is made. While visual attention has been shown to play a central role in guiding choice, much of the existing research has focused on predicting which option will ultimately be selected [1–5], rather than on understanding how the decision process unfolds when users hesitate between alternatives.

Eye-tracking (ET) has been widely used to study how users allocate attention when inspecting products and how this relates to preferences and final choices [6–8]. By recording gaze position and eye movements, ET provides objective measures such as fixations, saccades, scan paths, gaze plots, and heat maps, which allow researchers to quantify how visual information is explored and which elements attract greater attention [6,7]. In product selection contexts, multiple studies have shown that the option ultimately chosen



Academic Editor: João M. F. Rodrigues

Received: 5 June 2026

Revised: 6 July 2026

Accepted: 16 July 2026

Published: 19 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

tends to receive greater visual attention, reflected in longer fixation durations, a higher number of fixations, and more visits [2–4,9]. Furthermore, attention has been identified as a fundamental driver of choice, with evidence suggesting that the last fixation is typically directed towards the selected alternative and tends to be shorter than the average fixation duration [1,10]. Importantly, gaze bias towards the eventually chosen item may also emerge early in the decision process. A visual decision-making study including an eight-alternative forced-choice condition [11] showed that dwell duration was biased towards the chosen item from the early stages of the trial, suggesting that gaze bias is not limited to the final moments before response.

Beyond predicting final choices, ET has also been used to examine how users interpret product features through visual information. For example, Bayarri-Porcar et al. [12] applied ET to analyse usability-related product features in jar-opening (JO) assistive devices, showing that fixation- and visit-based metrics—such as total fixation duration, number of fixations, and total visit duration—were informative for understanding product evaluation. The study also demonstrated that ET can capture how users infer functional properties from rendered images, supporting its usefulness for early-stage design assessment.

ET has also been applied to characterise visual exploration strategies during decision-making. Studies analysing gaze plots and scan paths have shown that users tend to inspect the most relevant elements first and follow structured visual exploration strategies influenced by usual reading patterns [13]. A recent work has shown that visual exploration patterns vary according to task context and users' prior knowledge, with attention dynamically shifting depending on users' goals [14]. Recent research further suggests that attention dynamically modulates the weighting of attributes and alternatives during decision making. Eye movements not only reflect information acquisition but also the evolving evaluation process, with shifts in gaze being predictive of subsequent choices and preference formation [15,16]. Research combining ET with physiological measures has suggested that decision-making processes can be anticipated based on visual and physiological responses, with longer fixation times associated with greater interest and physiological signals such as blinking occurring prior to selection [17]. More recent studies have demonstrated that gaze behaviour in the final moments before a decision can be used to predict users' choices [4,5].

Despite this growing body of research, most studies have focused on identifying the selected option, describing gaze bias effects, or predicting the final decision. In contrast, considerably less attention has been paid to how hesitation manifests during the evaluation phase preceding selection, particularly in contexts involving multiple competing alternatives. This limitation is relevant because hesitation may reflect increased cognitive effort, difficulty in discriminating between options, or ambiguity in the interpretation of product features. Importantly, hesitation is not merely an abstract psychological state but may be expressed through observable behavioural patterns, such as repeated comparisons between alternatives or more complex visual exploration strategies.

In decision-making research, hesitation has often been associated with decision difficulty, choice conflict, and uncertainty during value-based comparison processes. When several alternatives compete and are difficult to discriminate, individuals tend to engage in iterative information search and repeated comparisons between promising options. Recent works propose that attention is dynamically allocated toward relevant information and likely choice candidates [16], with gaze patterns reflecting the ongoing accumulation of evidence and uncertainty reduction processes [18]. Therefore, hesitation may be understood as an observable behavioural manifestation of these underlying mechanisms during visual comparison and product evaluation.

Agost and Bayarri-Porcar [19] analysed gaze behaviour in a product-choice context. Their results confirmed that the selected option generally received longer total fixation duration, but also identified distinct gaze patterns suggestive of hesitation, including alternating attention between candidate options, the use of one option as a comparison anchor, and a rapid inspection followed by selection. The study also introduced the concept of “finalist alternatives” that attracted substantial attention without being selected, and highlighted the presence of hesitation-related behaviour prior to the final decision. However, this work remained largely exploratory, relying on qualitative inspection of gaze plots and heat maps, thus pointing to the need for more systematic quantitative approaches to characterise hesitation-related behaviour.

Product evaluation tasks provide a suitable framework to study these processes, as they typically involve visual comparison between multiple alternatives under conditions of limited information. In the case of assistive devices, such as JO aids, decision-making may require users to infer functional features—such as robustness, comfort, or required effort—from visual information alone, potentially increasing uncertainty and comparison behaviour.

In this context, the present study aims to examine whether hesitation during product selection can be systematically characterised using ET metrics. Specifically, it investigates whether shorter decision times are associated with less hesitant visual behaviour, whether hesitation is reflected in a greater number of gaze transitions between competing alternatives, and whether fixation- and visit-based ET metrics can more effectively capture hesitation-related behaviour than early-fixation indicators. In addition, the study analyses how visual attention is distributed across selected options, candidate alternatives, and remaining models during hesitation. By moving from exploratory descriptions to a structured, metric-based analysis, this work seeks to contribute to a better understanding of how hesitation emerges during visually guided decision-making and to the development of objective methods for detecting uncertainty in product evaluation contexts.

2. Materials and Methods

2.1. Participants and Equipment

ET data were collected from 67 volunteers (25 female, 42 male; mean age 23.73 ± 3.60 years, most of them were fourth-year undergraduate students) with normal or corrected-to-normal vision. The sample was therefore homogeneous and should be understood as a young-adult sample used for an initial, controlled investigation of hesitation during visual product evaluation. All participants provided written informed consent. Ethical approval was granted by the Ethics Committee of Universitat Jaume I (approval number CD/79/2021).

Gaze data were recorded using a Tobii Pro X2-60 system (Tobii Technology AB, Danderyd, Sweden), offering a gaze accuracy of 0.4° and a precision of 0.32° . The device was placed below a 24-inch display and operated at a sampling rate of 60 Hz. Participants were positioned approximately 60 cm from the screen. Prior to the task, a 9-point calibration procedure was conducted for each participant, following the manufacturer’s instructions. The ET session lasted approximately 5 min on average.

2.2. Experiment: Stimuli, Questions and Timeline

To create the experimental stimuli, six JOs were modelled in 3D using SolidWorks® 2020 and subsequently rendered with Blender® 2.83.18. Rendered static images were used deliberately to study an early-stage visual evaluation context, in which participants had to infer product attributes from visual information alone. This controlled setting allowed us to examine whether ET metrics could provide objective behavioural indicators of hesitation while keeping stimulus presentation standardized and avoiding variability associated with physical manipulation, haptic feedback, or differences in product handling. Each model

was presented in its intended functional orientation placed on the jar lid to reflect realistic usage conditions (Figure 1).



Figure 1. Models 1–6 of the JOs used in the stimuli.

The experiment began with the presentation of an image containing general instructions. Participants were then shown a slide displaying the six JO models to visually familiarise with the products prior to the evaluation. This step was intended to help participants focus on the evaluation task rather than on interpreting the product's design. Participants were allowed to take as much time as needed, with no time limit imposed.

Participants were asked to evaluate selected dimensions associated with product usability of the six JOs by answering three questions (Q1–Q3), each focusing on a specific criterion: robustness (Q1), comfort (Q2), and effort level required (Q3). These three criteria were selected based on previous findings [12] involving the same product family, which suggested different levels of visual interpretability and, consequently, different expected levels of decisional hesitation: robustness appeared to be more readily conveyed, effort level less readily conveyed, and comfort occupied an intermediate position. For each question, participants selected the JO they considered best suited to the specified evaluation criterion. To minimize potential order effects, the order of the three evaluation criteria (robustness, comfort, and effort level) was randomised across participants. Each criterion was presented with a predefined spatial arrangement of the six JO models; therefore, model positions varied across criteria but were kept constant across participants within each criterion. This procedure ensured that all participants evaluated the same visual configuration for a given criterion, while avoiding a fixed sequence of evaluation criteria.

Before each evaluation, a fixation slide was presented, displaying a blue dot at the centre of the screen. This procedure was based on the fixation protocol used in a previous study with the same product family [12], in which a central fixation point was used to ensure consistent starting points for gaze plots across stimuli. In the present study, the fixation duration was set to 2 s to provide a stable initial gaze position before the onset of each evaluation. Once the fixation was detected, the system automatically advanced to the corresponding evaluation slide, where the six JO models were displayed around the dot. After making their decision, participants were required to return their gaze to the blue dot. When a fixation of 2 s was detected, the blue dot changed to yellow, indicating that they should click on their selected model. This two-step fixation protocol dissociated the decision-making phase (blue dot) from the motor response phase (yellow dot), ensuring that gaze data reflected the evaluation process itself while minimising contamination from motor planning or cursor repositioning prior to clicking. Although this protocol introduced an artificial constraint compared with natural product choice behaviour, it was used to standardise the start and end of the decision interval and to isolate visual evaluation from the motor response. This method also provided consistent starting points for gaze plots and enhanced the reliability of inter-stimulus comparisons. A schematic representation of the experimental timeline is provided in Figure 2.

Prior to the main task, participants completed a trial designed to familiarise them with the experimental procedure. This task followed the same structure as the main experiment but used a different product category (drinking glasses) instead of JOs. Participants were asked to select the glass they found most aesthetically pleasing. The same sequence of actions and fixation-based interaction was employed, allowing participants to become

accustomed to the screen layout, task dynamics, and response protocol before engaging with the actual stimuli.

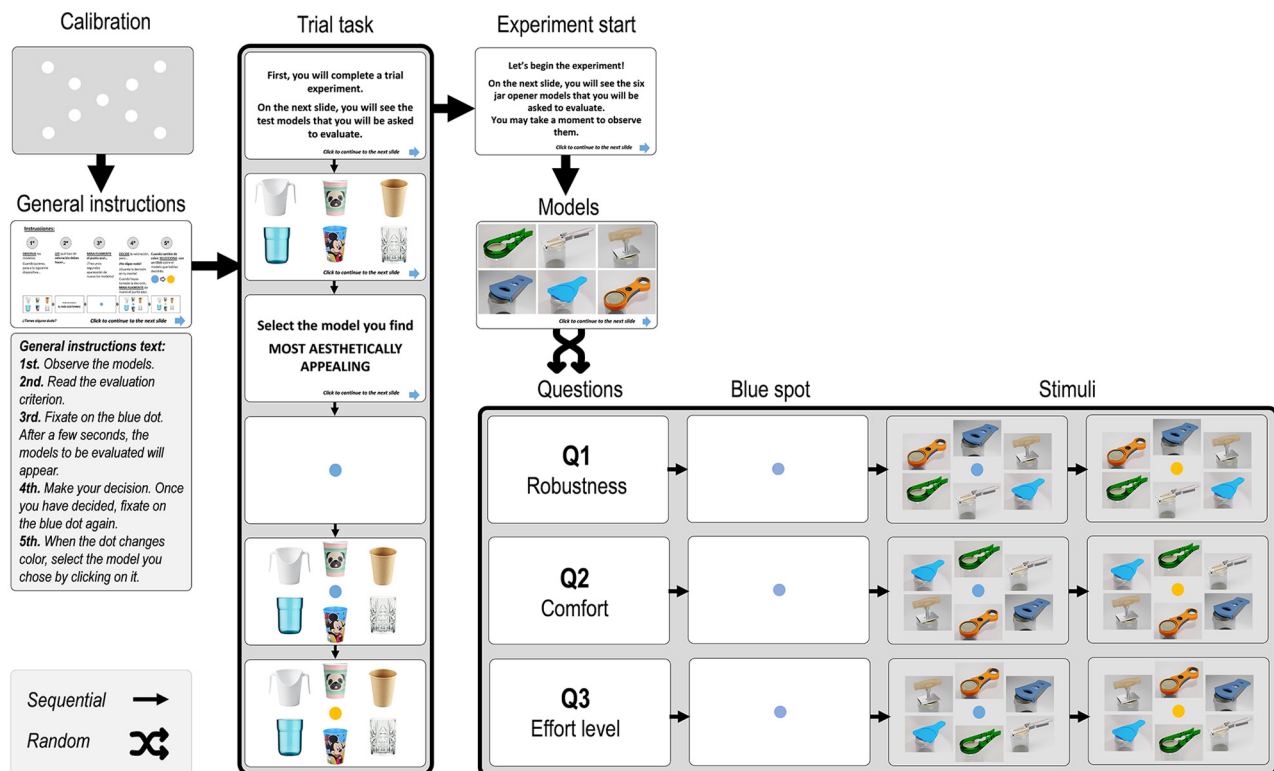


Figure 2. Timeline illustrating the sequence of questions and stimuli displayed. Blue arrows indicate progression to the next screen; blue and yellow dots represent the decision and model-selection phases, respectively.

2.3. Areas of Interest

To enable the analysis of ET metrics, Areas of Interest (AoIs) were defined for each stimulus image, delineating specific regions within the visual content. Table 1 provides illustrative examples of the AoIs. In accordance with the recommendations of Goldberg and Helfman [20], a margin of 4–5 mm was added around each AoI on the display. This adjustment accounted for the technical specifications of the ET system and the approximate viewing distance between participants and the device.

Table 1. List of AoIs and their definitions, accompanied by example image.

AoI Name	Definition	Example Image
AoI Model and AoI Position	Six rectangular AoIs, each enclosing one JO image. Slide with blue dot. Used for data analysis.	
AoI Model Selection	Rectangular area enclosing each individual model. Slide with yellow dot. Used only to record the selected model.	

Two AoI identifiers were used: AoI Model (1–6, the JO identity from Figure 1) and AoI Position (1–6, the spatial position on the screen, Figure 3). AoI Model was used to analyse

attention towards each product, whereas AoI Position was used to analyse spatial viewing patterns. The AoI Model Selection was used only to record which model was selected in each question.

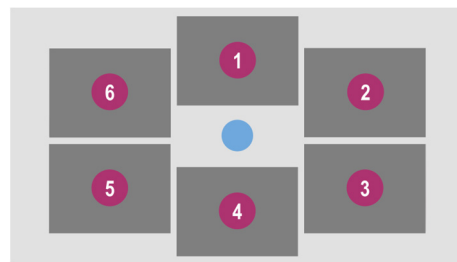


Figure 3. Stimulus positions used in the analysis.

After completing the ET experiment, participants were asked whether they had hesitated during the evaluation of each of the three questions (yes/no) and, if so, between which models. The yes/no format was used to obtain a binary trial-level classification of self-reported hesitation, which was required for the subsequent comparison between hesitation and no-hesitation trials. To support recall, this post-task query was conducted using a visual sheet showing the six models in the same order used during the corresponding evaluation trial. Self-reported hesitation was used as an operational indicator of subjective decision uncertainty, enabling the classification of trials based on participants' perceived hesitation. This information was then used to categorise trials according to the presence or absence of hesitation and to analyse visual behaviour accordingly.

In addition, those participants who were 4th-year students were provided with their own gaze-plot outputs and associated ET data (TDoF for each model in each question and the model selected, see Table 2). A detailed explanation of the plots and data was provided, and afterwards they were asked to reflect on their own visual exploration and decision process. Specifically, they were invited to comment on their personal decision in relation to the observed gaze pattern and fixation times, and to report whether they had hesitated when selecting the best option for robustness, comfort, and effort, and whether they considered this was reflected in their decision time. These written reflections were collected as qualitative material to complement the quantitative ET analyses.

2.4. Data Analysis

2.4.1. ET Metrics

Table 2 lists the ET metrics used in the analysis.

Table 2. ET metrics used in the analysis.

Metric	Definition
DoI	Duration of Interval (seconds). The total time spent observing the stimulus.
NoS	Number of Saccades. Saccades are defined as rapid eye movements.
AAoS	Average Amplitude of Saccades (degrees).
TAoS	Total Amplitude of Saccades (degrees).
TtFF	Time to First Fixation in AoI (seconds). A fixation is an eye movement with a velocity of less than 30°/s.
DoFF	Duration of First Fixation in AoI (seconds).
TDoF	Total Duration of Fixations in AoI (seconds).
NoF	Number of Fixations in AoI.
TDoV	Total Duration of Visits in AoI (seconds). A visit corresponds to all data (including saccadic movements, blinks, or invalid gaze data) from the start of the first fixation inside an AoI to the end of the last fixation in the same AoI.
NoV	Number of Visits in AoI.
Pupil diameter	Pupil diameter (mm). Exported separately for the left and right eye.

2.4.2. Data Extraction and Curation

Raw gaze data and AoI-level metrics were extracted in Tobii Pro Lab for each participant and question. Considering the fixation-based interaction protocol (blue-dot decision phase followed by yellow-dot motor phase), only gaze samples from the decision phase were retained for analysis.

The unit of analysis was the trial (participant $\times$ question; $67 \times 3 = 201$ trials), with each trial containing the six product alternatives and their corresponding spatial positions on the screen. Each trial was classified as hesitation or no-hesitation on the basis of participants' responses.

Analyses were conducted at two complementary levels: (a) trial-level metrics, capturing global decision behaviour, and (b) AoI-level metrics, capturing the distribution of visual attention across alternatives.

For hesitation trials, AoIs were further classified using a selection-based AoI grouping. This grouping was defined according to the role of each model in the participant's reported decision process: the selected model, candidate alternatives, and remaining models. The terminology used for these groups is summarised in Table 3. This selection-based group enabled the analyses of how visual attention was distributed across the decision space during hesitation trials.

Table 3. Definition of the selection-based AoI groups used in hesitation-trial analyses.

Term	Definition	Number of Models per Hesitation Trial
Selected model	The JO model ultimately selected by the participant.	Always 1
Candidate alternatives	JO models reported by the participant as alternatives considered during hesitation but not finally selected.	1 or more
Remaining models	JO models that were neither selected nor reported as candidate alternatives.	4 or less

For these selection-based AoI-group analysis, both absolute and per-model versions of the AoI-group metrics were considered (TDoF, NoF, TDoV and NoV). Absolute AoI-group metrics captured the total amount of attention allocated to each group within each hesitation trial. However, the three selection-based AoI groups were structurally unequal in size: the selected-model group always contained one model, whereas the candidate-alternative and remaining-model groups could contain different numbers of models depending on participants' responses. Therefore, to make the groups comparable at the level of individual models, per-model AoI-group metrics were computed by dividing each absolute group-level metric by the number of models included in that group. These per-model metrics represented the average value of each fixation- or visit-based measure per model within each selection-based AoI group.

Pupil diameter was exported separately for the left and right eye in millimetres. To reduce the influence of individual differences and trial-level variations in recording conditions, including possible differences in ambient lighting, pupil diameter was analysed using trial-level baseline correction. Given that luminance was not strictly controlled across all sessions, pupil-based results were treated as exploratory. For each trial, binocular pupil diameter was obtained by averaging the available left- and right-eye pupil values. The primary baseline-corrected pupil measure was computed as the median pupil diameter during the decision screen minus the median pupil diameter during the final 1000 ms of the preceding blue-dot fixation period. This median-based measure was selected to reduce the influence of transient artefacts or extreme samples. Additional corrected pupil variables

were computed for sensitivity analyses using a 500 ms baseline window while maintaining the median-based decision-period measure.

2.4.3. Statistical Analysis

All statistical analyses were conducted in jamovi (v2.6.44) [21–23] using two-tailed tests with $\alpha = 0.05$. An approximate sensitivity power analysis was conducted to contextualize the available sample size. Treating the 201 trials as observational units, the design would have 0.80 power to detect a minimum effect size of approximately Cohen's $d = 0.41$. This estimate was used as an approximate indication of detectable effect size for the trial-level analyses. Normality was assessed using Shapiro–Wilk tests and homogeneity of variances using Levene's test. When assumptions for parametric tests were not met, non-parametric tests were used as the primary inferential approach, with Welch's t -test additionally reported as a sensitivity analysis when the assumption of homogeneity of variances was violated.

The statistical analysis was organised around five aims:

- i. Initial orienting behaviour was examined using first-fixation information, including first-fixation position, whether the first fixation landed on the selected model, and first-fixation timing and duration (TtFF and DoFF).
- ii. Global decision behaviour was analysed by comparing trial-level ET metrics related to decision duration and oculomotor activity (DoI, NoS, TAoS and AAoS) between hesitation trials ($n = 126$) and no-hesitation trials ($n = 75$), and by examining associations among these metrics both overall and stratified by hesitation status. Exploratory logistic regression models were also used to assess whether trial-level ET metrics predicted self-reported hesitation. Given the strong correlation between NoS and TAoS, they were not entered simultaneously in the same model. The logistic regression models were interpreted as exploratory because they did not include mixed-effects terms to account for repeated measures within participants. As a sensitivity analysis addressing the repeated-trial structure of the data, within-participant comparisons were also conducted for the main trial-level metrics. This analysis included only participants who had at least one hesitation trial and at least one no-hesitation trial ($n = 54$). For each participant, mean values were computed separately across hesitation and no-hesitation trials. The analysis included the main global metrics (DoI, NoS, TAoS and AAoS) and complementary AoI-based inspection metrics aggregated across the six JO model AoIs within each trial (TDoF, NoF, TDoV and NoV). Because normality of the paired differences was not supported for most variables, paired comparisons were conducted using Wilcoxon signed-rank tests.
- iii. Within the 126 hesitation trials, the distribution of visual attention across the three selection-based AoI groups (the selected model, candidate alternatives and remaining models) was examined. Both absolute and per-model AoI-group metrics were analysed, and first-fixation timing was also compared across these groups.
- iv. Baseline-corrected pupil diameter was compared between hesitation and no-hesitation trials, and sensitivity analyses were conducted using alternative baseline windows and decision-period summaries.
- v. Participants' written reflections were analysed using an AI-assisted qualitative synthesis supported by Microsoft Copilot (accessed in May 2026) to contextualise the quantitative findings. The qualitative material was provided to Copilot, which was asked to synthesize the responses and identify recurring themes and common points across participants. The AI-assisted synthesis was reviewed by the authors against the original written responses and the structured post-task hesitation responses. Quantitative summaries derived from the qualitative analysis were checked against the

coded dataset. The qualitative synthesis was used to contextualize the quantitative ET findings rather than as a standalone qualitative validation procedure.

Table 4 summarises the analytical aims, analysis blocks, statistical procedures, dependent variables, grouping factors, and null hypotheses used in the study. Metric abbreviations are defined in Table 2.

Table 4. Summary of the analytical aims and statistical procedures used in the study.

Analysis Block	Objective of the Analysis	Statistical Test	Dependent Variable(s)	Independent Variable/ Grouping Factor	Null Hypothesis
(i) Initial attention: first-fixation position by hesitation status	To test whether hesitation affected the first-fixation position	χ^2 test of independence	First-fixation position (6 locations)	Hesitation status (hesitation vs. no-hesitation)	First-fixation position is independent of hesitation status
(i) Initial attention: overall start-position bias	To test whether first-fixation positions were uniformly distributed	χ^2 goodness-of-fit test	First-fixation position (6 locations)	Not applicable	First-fixation positions follow a uniform distribution across the 6 locations
(i) Initial attention: first fixation on selected option	To test whether hesitation affected first fixation on the selected model	χ^2 test of independence	First fixation on selected model (yes/no)	Hesitation status	The probability of first fixating the selected model is independent of hesitation status
(i) Initial attention: first-fixation timing	To test whether hesitation affected first-fixation timing	Mann–Whitney U test (primary); Welch’s <i>t</i> -test (sensitivity)	TtFF; DoFF	Hesitation status	TtFF and DoFF do not differ between hesitation and no-hesitation trials
(ii) Global ET metrics by hesitation status	To test whether hesitation was associated with differences in global ET behavior	Mann–Whitney U test (primary); Welch’s <i>t</i> -test (sensitivity)	DoI; NoS; TAoS; AAoS	Hesitation status	DoI, NoS, TAoS, and AAoS do not differ between hesitation and no-hesitation trials
(ii) Within-participant sensitivity analysis by hesitation status	To test whether the main hesitation-related pattern was preserved when comparing hesitation and no-hesitation trials within the same participants	Wilcoxon signed-rank test	Participant-level mean values of DoI, NoS, TAoS, AAoS, TDoF, NoF, TDoV and NoV	Hesitation status within participant; participants with both hesitation and no-hesitation trials	Within participants, hesitation and no-hesitation trials do not differ in the analyzed ET metrics
(ii) Associations among global ET metrics	To test associations among global ET metrics	Spearman correlation	DoI; NoS; TAoS; AAoS	Overall and stratified by hesitation status	For each pair, the correlation coefficient equals 0
(ii) Exploratory prediction of hesitation	To test whether trial-level ET metrics predicted hesitation	Binomial logistic regression; likelihood-ratio tests for nested models; AIC and McFadden’s pseudo- R^2 for model comparison; VIF for collinearity assessment	Hesitation status (binary)	Baseline: DoI; additional predictors: NoS or TAoS	Regression coefficients equal 0; added predictors do not improve model fit

Table 4. Cont.

Analysis Block	Objective of the Analysis	Statistical Test	Dependent Variable(s)	Independent Variable/ Grouping Factor	Null Hypothesis
(iii) Selection-based AoI-group analyses: absolute metrics	To test differences across AoI groups within hesitation trials	Friedman test; Durbin–Conover pairwise comparisons (post hoc)	TDoF; NoF; TDoV; NoV	Selection-based AoI group (selected model, candidate alternatives, remaining models)	The distributions of the absolute AoI-group metrics are equal across the 3 groups
(iii) Selection-based AoI-group analyses: per-model metrics	To test differences across AoI groups, calculated by dividing each group metric by the number of models included in that group	Friedman test; Durbin–Conover pairwise comparisons (post hoc)	TDoF; NoF; TDoV; NoV per model	Selection-based AoI group	The distributions of the per-model AoI-group metrics are equal across the 3 groups
(iii) First-fixation timing across selection-based AoI groups within hesitation trials	To test whether first-fixation timing differed across AoI groups within hesitation trials	Friedman test; Durbin–Conover pairwise comparisons (post hoc)	TtFF; DoFF	AoI group	The distributions of TtFF and DoFF are equal across the 3 AoI groups within hesitation trials
(iv) Baseline-corrected pupil diameter by hesitation status	To test whether baseline-corrected pupil diameter differed by hesitation status	Mann–Whitney U test; sensitivity analyses with alternative baseline window and decision-period summary	Baseline-corrected pupil diameter	Hesitation status	Baseline-corrected pupil diameter does not differ between hesitation and no-hesitation trials
(v) Qualitative synthesis	To identify recurring themes related to hesitation and decision experience	AI-assisted qualitative synthesis	Written reflections (49 reports)	Structured hesitation responses used as contextual information	Not applicable

3. Results

3.1. Early Orienting and First-Fixation Behavior

To examine whether hesitation affected initial gaze allocation, first-fixation behaviour was analysed in terms of spatial position, first fixation on the selected model, and timing (TtFF and DoFF). The distribution of first-fixation positions did not differ between hesitation and no-hesitation trials, $\chi^2(5, N = 201) = 3.07, p = 0.690$, indicating that both conditions shared a similar pattern of initial spatial allocation across the six screen locations. However, when compared against a uniform distribution, a strong start-position bias was observed ($\chi^2(5, N = 201) = 101, p < 0.001$), with first fixations concentrated at positions 1 (38.81%) and 6 (25.37%).

The probability that the first fixation landed on the model that was ultimately selected did not differ between hesitation and no-hesitation trials ($\chi^2(1, N = 201) = 2.06, p = 0.152, \phi = 0.101$).

First-fixation timing differed between conditions being TtFF shorter in hesitation trials than in no-hesitation trials (Table 5), with consistent results in the Welch sensitivity analysis. In contrast, DoFF did not differ between conditions.

Table 5. Descriptive and inferential statistics for first-fixation ET metrics (TtFF and DoFF) by hesitation condition.

Metrics	Hesitation Condition	Mean	SD	Median	Mann–Whitney U	<i>p</i> -Value	Welch <i>p</i>
TtFF *	No-hesitation	0.629	0.460	0.468	3800	0.020	0.015
	Hesitation	0.494	0.277	0.418			
DoFF	No-hesitation	0.198	0.133	0.167	4248	0.232	0.183
	Hesitation	0.176	0.096	0.150			

Note. No-hesitation trials, $n = 75$; hesitation trials, $n = 126$. * Significant difference between hesitation and no-hesitation trials ($p < 0.05$).

Overall, hesitation was not associated with differences in the spatial location of the first fixation or in the probability of first fixating the selected model. However, first-fixation timing differed between conditions, with shorter TtFF in hesitation trials. Descriptive statistics for the first-fixation timing ET metrics by hesitation condition are presented in Table 5.

3.2. Global Decision Behavior

3.2.1. Differences Between Hesitation and No-Hesitation Trials

Table 6 summarises the descriptive statistics for the trial-level global ET metrics (DoI, NoS, TAoS and AAoS) by hesitation condition.

Table 6. Descriptive and inferential statistics for trial-level global ET metrics (DoI, NoS, TAoS and AAoS) by hesitation condition.

Metrics	Hesitation Condition	Mean	SD	Median	IQR	Mann–Whitney U	<i>p</i> -Value
DoI *	No-hesitation	10.257	3.354	9.850	7.508–12.242	3449	0.001
	Hesitation	13.725	8.938	11.625	8.737–16.179		
NoS *	No-hesitation	19.907	10.204	18.000	13–25.5	3293	<0.001
	Hesitation	27.921	18.264	24.500	17–34		
TAoS *	No-hesitation	121.595	73.660	113.220	70.585–169.530	3542	0.003
	Hesitation	155.329	86.798	153.730	98.455–197.985		
AAoS	No-hesitation	5.875	1.634	5.950	4.845–6.970	4463	0.511
	Hesitation	5.677	1.598	5.820	4.830–6.665		

Note. No-hesitation trials, $n = 75$; hesitation trials, $n = 126$. * Significant difference between hesitation and no-hesitation trials ($p < 0.05$).

At the trial level, hesitation was associated with differences in several global ET metrics (Table 6). DoI was longer in hesitation trials than in no-hesitation trials. NoS was likewise higher in hesitation trials than in no-hesitation trials. Similarly, TAoS was higher in hesitation trials. By contrast, AAoS did not differ between conditions.

Overall, hesitation was associated with longer decision times, a higher number of saccades, and a greater total saccadic amplitude, whereas average saccade amplitude did not differ between conditions.

The results of the within-participant sensitivity analysis (Table 7) were consistent with the trial-level analyses: hesitation trials showed longer decision times, a higher number of saccades, greater total saccadic amplitude, and higher fixation- and visit-based inspection metrics aggregated across the six JO model AoIs. Average saccade amplitude did not differ between conditions. These results indicate that the hesitation-related pattern was preserved when comparing hesitation and no-hesitation trials within the same participants.

Table 7. Descriptive and inferential statistics for the within-participant sensitivity analysis by hesitation condition.

Metrics	Hesitation Condition	Mean	SD	Median	Wilcoxon W	p-Value
DoI *	No-hesitation	10.218	3.294	9.825	352	<0.001
	Hesitation	13.399	7.355	11.758		
NoS *	No-hesitation	20.565	9.782	19.500	353	0.001
	Hesitation	27.778	16.149	26.000		
TAoS *	No-hesitation	125.102	71.101	117.200	433	0.008
	Hesitation	157.320	77.103	149.555		
AAoS	No-hesitation	5.874	1.574	6.020	757	0.717
	Hesitation	5.799	1.497	5.865		
TDoF *	No-hesitation	4.836	2.412	4.508	295	<0.001
	Hesitation	7.206	5.489	6.537		
NoF *	No-hesitation	22.037	9.905	18.250	319	<0.001
	Hesitation	30.389	19.031	27.000		
TDoV *	No-hesitation	5.328	2.579	5.017	287	<0.001
	Hesitation	8.008	6.161	7.087		
NoV *	No-hesitation	10.889	4.382	9.250	372	0.006
	Hesitation	13.037	4.884	12.500		

Note. $n = 54$ participants with both hesitation and no-hesitation trials. * Significant difference between hesitation and no-hesitation trials ($p < 0.05$).

3.2.2. Associations Among Global ET Metrics

Spearman correlations revealed consistent associations among the global ET metrics. DoI correlated positively with NoS ($\rho = 0.642, p < 0.001$) and TAoS ($\rho = 0.509, p < 0.001$), and negatively with AAoS ($\rho = -0.275, p < 0.001$). As expected, NoS and TAoS were strongly correlated ($\rho = 0.919, p < 0.001$).

When analysed separately by hesitation condition, the same general pattern was retained, although the strength of the associations between DoI and the saccade-based metrics was greater in hesitation trials. In no-hesitation trials, DoI correlated with NoS ($\rho = 0.463, p < 0.001$) and TAoS ($\rho = 0.363, p = 0.001$), whereas in hesitation trials these correlations were stronger (NoS: $\rho = 0.700, p < 0.001$; TAoS: $\rho = 0.562, p < 0.001$). In addition, AAoS was negatively associated with DoI only in hesitation trials, whereas no such association was observed in no-hesitation trials.

3.2.3. Exploratory Predictive Modelling

Exploratory binomial logistic regression models were estimated to examine whether trial-level ET metrics predicted the probability of reporting hesitation. In the baseline model, DoI alone significantly predicted hesitation (OR = 1.129, 95% CI [1.051, 1.214], $p < 0.001$), although model fit was modest (AIC = 254; McFadden's $R^2 = 0.0582$). When NoS was added, model fit improved only marginally (AIC = 253), and both predictors became marginally significant, reflecting shared variance between them. A complementary model including TAoS instead of NoS showed that DoI remained significant, whereas TAoS did not provide reliable incremental predictive value. Overall, within this exploratory trial-level analysis, decision time emerged as the most consistent predictor of reported hesitation, while the incremental contribution of additional trial-level metrics was limited.

3.3. Attention Allocation Across Selection-Based AoI Groups in Hesitation Trials

To characterise how attention was distributed within hesitation trials, ET metrics (TDoF, NoF, TDoV and NoV) were analysed across the three selection-based AoI groups: the selected model, the candidate alternatives, and the remaining models.

3.3.1. Absolute AoI-Group Metrics

All absolute selection-based AoI-group metrics differed significantly across the three groups (Table 8). For TDoF and TDoV, the same pattern was observed, with candidate alternatives receiving the highest values, followed by remaining models, and the selected model receiving the lowest values. For NoF, candidate alternatives and remaining models both received more fixations than the selected model, while they did not differ reliably from each other. For NoV, remaining models received the highest number of visits, followed by candidate alternatives, with the selected model again receiving the lowest values.

Table 8. Descriptive and inferential statistics for absolute AoI-group metrics (TDoF, NoF, TDoV and NoV) in hesitation trials.

Metrics	AoI Groups	Mean	SD	Median	Friedman $\chi^2(2)$	<i>p</i> -Value
TDoF *	Selected model	1.167	1.251	0.734	104	<0.001
	Candidate alternatives	3.495	4.221	2.692		
	Remaining models	2.701	2.455	2.117		
NoF *	Selected model	5.016	4.189	4.000	96.9	<0.001
	Candidate alternatives	13.310	13.603	11.000		
	Remaining models	12.349	9.872	10.500		
TDoV *	Selected model	1.276	1.352	0.800	100	<0.001
	Candidate alternatives	3.892	4.809	2.925		
	Remaining models	2.999	2.757	2.350		
NoV *	Selected model	2.270	1.602	2.000	131	<0.001
	Candidate alternatives	4.817	2.806	4.000		
	Remaining models	5.770	3.406	5.000		

Note. $n = 126$. * Significant difference across selection-based AoI groups within hesitation trials ($p < 0.05$).

Across all absolute AoI-group metrics, the selected model consistently received the least visual attention, whereas candidate alternatives and remaining models attracted substantially more fixations and visits. Descriptive statistics for the absolute AoI-group metrics in hesitation trials are presented in Table 8.

3.3.2. Per-Model AoI-Group Metrics

To account for the unequal number of models included in each selection-based AoI group, the absolute AoI-group metrics were also expressed as per-model values by dividing each group metric by the number of models in that group. All per-model AoI-group metrics differed significantly across the three groups (Table 9). For TDoF, NoF, TDoV and NoV, candidate alternatives showed the highest values per model, the selected model showed intermediate values, and remaining models showed the lowest values. Pairwise Durbin–Conover comparisons indicated significant differences between all three groups for each metric. Most pairwise comparisons were significant at $p < 0.001$, except for the comparison between the selected model and remaining models for TDoF ($p = 0.004$) and TDoV ($p = 0.002$).

Table 9. Descriptive and inferential statistics for per-model AoI-group metrics (TDoF, NoF, TDoV and NoV) in hesitation trials.

Metrics	AoI Groups	Mean	SD	Median	Friedman $\chi^2(2)$	<i>p</i> -Value
TDoF per model *	Selected model	1.167	1.251	0.734	91.4	<0.001
	Candidate alternatives	1.952	2.110	1.513		
	Remaining models	0.835	0.694	0.671		
NoF per model *	Selected model	5.016	4.189	4.000	74.8	<0.001
	Candidate alternatives	7.363	6.666	6.000		
	Remaining models	3.834	2.796	3.000		
TDoV per model *	Selected model	1.276	1.352	0.800	90.3	<0.001
	Candidate alternatives	2.162	2.394	1.659		
	Remaining models	0.928	0.785	0.729		
NoV per model *	Selected model	2.270	1.602	2.000	68.4	<0.001
	Candidate alternatives	2.688	1.365	2.500		
	Remaining models	1.796	0.910	1.667		

Note. $n = 126$ hesitation trials. * Significant difference across selection-based AoI groups within hesitation trials ($p < 0.05$).

3.3.3. Early Orienting Within Hesitation Trials

Within hesitation trials, TtFF differed significantly across selection-based AoI groups ($\chi^2(2) = 8.87$, $p = 0.012$). The selected model was fixated later than both candidate alternatives ($p = 0.027$) and remaining models ($p = 0.004$), whereas candidate alternatives and remaining models did not differ reliably. By contrast, DoFF did not differ across groups ($\chi^2(2) = 0.650$, $p = 0.723$).

3.4. Pupil Diameter

Baseline-corrected pupil diameter did not differ significantly between hesitation and no-hesitation trials (Table 10). For the primary measure, computed as the median pupil diameter during the decision screen minus the median pupil diameter during the final 1000 ms of the preceding blue-dot fixation period, median values were slightly higher in hesitation trials than in no-hesitation trials, but the difference was not statistically significant. The same conclusion was obtained using a 500 ms baseline window ($p = 0.886$).

Table 10. Descriptive and inferential statistics for baseline-corrected pupil diameter by hesitation condition.

Analysis	Metric	Group	Median	IQR	Mann–Whitney U	<i>p</i> -Value
By hesitation condition	Baseline-corrected pupil diameter	No-hesitation	0.242	0.168–0.351	4649	0.850
		Hesitation	0.257	0.160–0.372		

Note. No-hesitation trials, $n = 75$; hesitation trials, $n = 126$.

3.5. Qualitative Reflections on Gaze Plots and Decision Experience

Before examining the written reflections, the distribution of self-reported hesitation was summarised for the full sample. Across the full sample, hesitation was reported in 126 of the 201 trials (62.7%). By evaluation criterion, hesitation was most frequent for effort (49/67, 73.1%), followed by comfort (41/67, 61.2%) and robustness (36/67, 53.7%). At the participant level, 66 of the 67 participants (98.5%) reported hesitation in at least one of the three evaluation categories.

Written reflections were available for 49 of the 67 participants who completed the ET experiment (73.1%), and the following qualitative synthesis is therefore based on this subset of reports. In the current sample, almost all participants (48/49, 98%) reported hesitation

in at least one of the three evaluation categories, indicating that hesitation was a common feature of the task rather than an isolated occurrence. At the category level, hesitation was reported most frequently for effort (38/49, 77.6%), followed by comfort (29/49, 59.2%) and robustness (25/49, 51%). Most participants reported hesitation in two categories (28/49, 57.1%), whereas 12/49 (24.5%) reported hesitation in only one category, 8/49 (16.3%) in all three categories, and only one participant (2%) reported no hesitation at all.

A recurring theme in the comments was that hesitation was experienced as a process of repeated comparison between a limited number of alternatives rather than as a diffuse search across all options. Participants frequently described moving back and forth between two or three models, revisiting the same candidates several times before making a final choice. This pattern is consistent with the structured hesitation reports in the dataset, where hesitation most often involved two alternatives rather than a larger set of competing options.

The comments also showed that participants commonly associated hesitation with longer decision times. When they recalled uncertainty, they often described spending more time comparing products, re-checking features, and delaying the final decision. Several responses explicitly linked hesitation to a denser or less linear gaze path, characterised by repeated returns to the same models. This subjective account is broadly consistent with the quantitative findings reported above, in which hesitation was associated with longer decision times and with visual behaviour indicative of more extensive comparison.

Another relevant theme was that greater visual attention did not always correspond directly to the finally selected model. Many participants stated that the chosen option was also the one they had looked at the most, and interpreted this as confirmation that their gaze behaviour reflected their final decision. However, other participants reported the opposite pattern: some of the longest fixations were directed to alternatives that were ultimately discarded, because these required closer inspection or generated uncertainty before rejection. The qualitative material therefore suggests that prolonged visual attention may reflect both preference and deliberation, rather than final choice alone.

The category most frequently described as difficult was effort. Participants often explained that this judgement required them to imagine how the product would work in use, how much force would be needed, or how the mechanism would behave when opening the jar. As a result, the evaluation of effort was often described as slower, more comparative, and more uncertain than the other two criteria. This qualitative pattern is consistent with the quantitative analyses, in which effort was the category with the highest frequency of reported hesitation.

Participants occasionally referred to specific visual cues when explaining why hesitation emerged, such as grip areas, apparent leverage, or perceived structural solidity, but these comments were generally framed in terms of how they supported comparison between alternatives rather than as design evaluations in themselves. Similarly, although most participants viewed ET positively as a useful method for understanding user behaviour, this point appeared mainly as a secondary reflection rather than as the central focus of their reports. Some participants also noted that judging the products from images alone could make certain decisions more difficult, particularly when trying to infer effort or functional performance without physical interaction.

Overall, the qualitative responses consistently described hesitation as a process involving repeated comparison between alternatives, extended decision time, and iterative inspection of competing options.

4. Discussion

The present study sought to characterise hesitation during product selection through ET metrics, moving the focus beyond predicting which option is ultimately chosen towards understanding how the decision process unfolds when users hesitate between alternatives. Overall, the findings indicate that hesitation was associated with a distinct and structured pattern of visual behaviour. At the global level, hesitation trials showed longer decision times and more extensive visual exploration, reflected in a higher number of saccades and greater total saccadic amplitude. At the level of early orienting, hesitation was not associated with differences in the spatial location of the first fixation or in the probability that the first fixation landed on the selected model, but it was associated with first-fixation timing. At the level of attention allocation, the selected model received lower total group-level fixation- and visit-based values than candidate alternatives and remaining models. However, the per-model analysis showed a more nuanced pattern: candidate alternatives received the highest values per individual model, the selected model showed intermediate values, and remaining models showed the lowest values. By contrast, baseline-corrected pupil diameter did not significantly differentiate hesitation from no-hesitation trials. Taken together, these findings suggest that hesitation was not merely reflected in longer decision times, but in a structured pattern involving differences in early orienting, increased exploratory activity, and a redistribution of attention across the decision space.

A first relevant contribution of the study concerns the earliest stages of visual exploration. Previous ET research has often emphasised gaze bias towards the selected option and the role of the final fixation prior to choice [1,10], although early gaze bias has also been documented in multi-alternative visual decision tasks [11]. The present results extend this perspective by showing that differences between hesitant and non-hesitant trials were also visible at the onset of the decision process. Although hesitation was not associated with changes in the spatial location of the first fixation or with the probability that the first fixation landed on the selected model, it was associated with first-fixation timing. Specifically, TtFF was shorter in hesitation trials than in no-hesitation trials, whereas DoFF did not differ between conditions. Moreover, within hesitation trials, the selected option was first fixated later than both candidate alternatives and remaining options. Taken together, these findings suggest that hesitation is reflected in early orienting dynamics, not through a different initial spatial bias, but through differences in the timing with which visual information was first accessed and in the later arrival of gaze to the option ultimately selected.

At the global level, hesitation was associated with longer decision times, a higher number of saccades, and greater total saccadic amplitude, whereas average saccadic amplitude did not differ between conditions. The observed pattern is consistent with recent theories proposing that attention is allocated adaptively during multi-alternative decisions, with individuals repeatedly sampling information from promising alternatives until sufficient evidence has accumulated to support a choice. Increased visual exploration during hesitation may therefore reflect pro-longed evidence accumulation and uncertainty reduction processes rather than inefficient search behaviour [16]. The increase in NoS suggests more frequent shifts in attention during hesitant trials, while the increase in TAOs indicates greater cumulative oculomotor activity across the decision interval. The absence of differences in AAOs further suggests that hesitation affected the frequency and cumulative extent of eye movements rather than their average scale. In this respect, the findings are consistent with previous work linking visual attention to information accumulation and comparison during decision-making [2,3]. At the same time, the present study shows more specifically that these exploratory dynamics were stronger when participants reported hesitation. The stronger associations observed between DoI and the saccade-based metrics

in hesitation trials are compatible with a tighter coupling between decision duration and visual exploration under conditions of hesitation.

The selection-based AoI-group analyses provide one of the clearest indications that hesitation is associated with a qualitatively different distribution of attention. At the group level, candidate alternatives and remaining models accumulated more fixation- and visit-based attention than the selected model. This indicates that, during hesitation trials, visual attention was not concentrated primarily on the option ultimately selected, but was distributed across alternatives involved in comparison and evaluation. At first sight, this result may appear to contrast with previous studies showing that the selected option often receives greater visual attention than non-selected alternatives [2,5,9]. However, the increased inspection of candidate alternatives observed during hesitation is consistent with recent models proposing that attention is dynamically allocated toward promising options and relevant attributes during evidence accumulation [15]. Repeated comparisons between competing alternatives may therefore reflect an extended evaluation process aimed at reducing uncertainty before commitment to a final choice. The present study focuses specifically on hesitation trials, in which prolonged attention is likely to reflect comparison, uncertainty, and elimination processes rather than preference alone. In this context, attention should not be interpreted only as a marker of emerging commitment to the final choice, but also as a marker of active comparison between alternatives.

The per-model analysis refined this interpretation by accounting for the unequal number of models included in each selection-based AoI group. Candidate alternatives received the highest values per individual model across TDoF, NoF, TDoV and NoV, indicating that their higher attention was not simply a consequence of the candidate group containing more than one model. In contrast, remaining models, which accumulated substantial attention at the group level, showed the lowest values after per-model adjustment. The selected model occupied an intermediate position. This pattern suggests that candidate alternatives were the most intensively inspected options on a per-model basis, whereas remaining models attracted attention more diffusely as a group. The qualitative reflections reinforce this interpretation, as several participants explicitly reported that some of the alternatives receiving the longest inspection were not the ones eventually selected, but rather those that required closer examination before being discarded.

These findings extend previous exploratory work by Agost and Bayarri-Porcar [19], which identified gaze patterns suggestive of hesitation, such as alternating attention between alternatives, reference-based comparison, and the presence of finalist alternatives. The present study advances that line of work by translating those qualitative patterns into a more systematic quantitative framework based on trial-level metrics, first-fixation indicators, AoI-group analyses, and per-model adjustment. In addition, the results complement previous work on JO assistive devices showing that ET metrics can capture how users interpret product-related functional cues from rendered images [12]. Whereas that earlier study focused primarily on usability-related evaluation, the present study shows that ET can also be used to characterise hesitation during visually guided product selection. Taken together, the two studies support the value of ET not only for examining which product is preferred or how product features are interpreted, but also for revealing how uncertainty emerges during comparison and selection.

The findings also have a broader conceptual implication for the study of decision-making in product evaluation. Rather than treating hesitation as a vague or purely subjective internal state, the present results suggest that it can be identified through a structured pattern of visual behaviour. Studying hesitation is therefore relevant because it provides process-level information that is not captured by the final choice alone, and may reflect increased cognitive effort, difficulty in discriminating between alternatives, or ambiguity

in interpreting product features. In the present task, hesitation was reflected in three complementary components: differences in early orienting, increased exploratory activity over the course of the trial, and a redistribution of attention across the decision space. Importantly, the per-model analysis showed that candidate alternatives were the most intensively inspected options after accounting for group size, supporting the interpretation of hesitation as a focused comparison process involving competing alternatives. This is especially relevant in product contexts where users must infer functional properties from visual information alone, as was the case here for robustness, comfort, and effort. Under such conditions, hesitation may arise when alternatives are difficult to discriminate visually or when product cues do not support an immediate interpretation of likely product performance. From an applied perspective, these findings suggest a possible way of using ET in early-stage product evaluation. Rather than replacing simple indicators such as decision time, ET metrics can help explain why a decision takes longer by showing whether the additional time is associated with broader visual exploration, repeated shifts between alternatives, or attention concentrated on candidate options that are ultimately not selected. In practice, designers could use this information to identify product attributes or alternatives that generate uncertainty, to detect visually ambiguous design cues, and to compare whether alternative design versions support more confident selection. In this sense, the proposed metric-based approach provides a diagnostic framework for interpreting hesitation-related visual behaviour and for identifying design alternatives that may require further refinement.

Trial-level baseline correction was applied using the blue-dot fixation period preceding each decision screen to reduce the effect of the data being collected under different lighting conditions. Baseline-corrected pupil diameter did not differ significantly between hesitation and no-hesitation trials. Therefore, the absence of significant pupil differences should not be interpreted as definitive evidence that pupil responses are unrelated to hesitation, but rather as an indication that pupil diameter did not provide reliable discrimination under the present experimental conditions. Recent evidence suggests that pupil dilation may primarily capture large differences in uncertainty and may be relatively insensitive to more subtle variations in uncertainty or decision conflict [24]. Future studies should control luminance and testing environment more strictly to assess the potential contribution of pupil-based measures to hesitation detection.

The qualitative reflections provide complementary support for the interpretation of the quantitative results. Participants frequently described hesitation as involving repeated comparison between a limited number of alternatives, spending more time on the decision, and iteratively inspecting competing options. They also indicated that greater visual attention did not always correspond directly to the chosen option, as some highly inspected alternatives were ultimately discarded, which is consistent with the AoI-group and per-model findings. Particularly noteworthy was the prominence of effort in the reflections and in the distribution of reported hesitation across criteria. In the present observational design, hesitation was not experimentally induced but analysed as it naturally emerged during the evaluation task. Effort was the criterion with the highest frequency of reported hesitation, and participants often described this judgement as requiring greater inference about future use, required force, and mechanism performance. This pattern is consistent with the intended differences in visual interpretability across the three criteria: required effort was less directly inferable from static visual information than robustness and, to some extent, comfort. Therefore, hesitation may have increased when the evaluated attribute required participants to mentally simulate product use rather than relying on directly visible design cues. Although these qualitative materials are complementary rather than confirmatory, they reinforce the interpretation of hesitation

as an organised process of comparison and re-evaluation rather than as a diffuse feeling of indecision.

Several limitations should be considered when interpreting these findings, and they point to specific methodological improvements for future research. Hesitation was operationalised from participants' binary self-reports after each trial. This approach provided direct information about perceived hesitation and allowed trials to be classified as hesitation or no-hesitation, supported by a visual sheet showing the six models in the same order as in the corresponding trial. Future studies could complement this procedure with more fine-grained measures, such as continuous confidence or hesitation ratings, response dynamics, or behavioural indicators of repeated comparison. In addition, although the main trial-level analyses treated trials as observational units, the within-participant sensitivity analysis showed that the main hesitation-related pattern was preserved when participants were compared with themselves. Future work could further extend this approach using mixed-effects or hierarchical models to account explicitly for trial nesting within participants.

The scope of the study was also defined by the use of a homogeneous young-adult sample, static rendered stimuli, and a single product category. This design was appropriate for an initial controlled investigation of hesitation during early-stage visual product evaluation, although previous reviews of eye movements during decision making have noted that controlled search and choice tasks may not fully capture the complexity of real-world decision contexts [18]. Future research should therefore examine whether the observed hesitation-related ET patterns are preserved in more diverse samples, during physical product interaction or dynamic VR/AR environments, and across other product categories. Similarly, possible learning, carryover, fatigue, or position-related effects cannot be completely ruled out, as all participants evaluated the same six products and each criterion used a predefined spatial arrangement. These effects could be assessed more directly in future studies using fully counterbalanced layouts or randomised product positions. Finally, pupil-diameter findings should be interpreted cautiously because luminance and testing conditions were not strictly controlled across all sessions. Future studies using pupil-based indicators should apply stricter control of ambient illumination, display luminance, and viewing conditions, and could also incorporate scan path complexity, transition matrices, or other sequence-based measures to better capture comparison dynamics during hesitation.

In conclusion, the present study shows that hesitation during product selection can be characterised through a coherent set of ET-based indicators. Hesitation was not only associated with longer decision times, but also with differences in early orienting, more extensive exploratory eye-movement behaviour, and a redistribution of attention across the decision space. The per-model analysis further showed that candidate alternatives were the most intensively inspected options after accounting for the number of models within each group. These findings support the usefulness of ET for moving beyond simple attention-choice associations and towards a more process-oriented understanding of decision-making under hesitation in multi-alternative product evaluation.

5. Conclusions

This study shows that hesitation during product selection can be characterised through a coherent set of eye-tracking indicators. Rather than being reflected only in longer decision times, hesitation is associated with a broader and more structured pattern of visual behaviour involving differences in early orienting, more extensive exploratory eye-movement activity, and a redistribution of attention across the decision space. In particular, hesitation

trials are characterised by longer decision times, a higher number of saccades, greater total saccadic amplitude, and earlier first fixations than no-hesitation trials.

Within hesitation trials, visual attention was not concentrated exclusively on the option ultimately selected. At the group level, candidate alternatives and remaining models accumulated more fixation- and visit-based attention than the selected model. Candidate alternatives received the highest values per individual model, whereas the selected model showed intermediate values and remaining models showed the lowest values. These findings indicate that, in hesitation contexts, prolonged visual attention is more closely related to comparison, evaluation, and rejection processes across alternatives than to commitment to the final choice alone.

By contrast, baseline-corrected pupil diameter did not significantly distinguish hesitation from no-hesitation trials. In the present task, fixation-, visit-, and saccade-based metrics therefore appeared to be more informative indicators of hesitation than pupil diameter. The qualitative reflections converged with the quantitative results, describing hesitation as a process of repeated comparison between a limited number of alternatives, extended decision time, and iterative inspection of competing options.

Taken together, the findings support the usefulness of eye-tracking for moving beyond simple attention–choice associations and towards a more process-oriented understanding of hesitation in multi-alternative product evaluation. The original contribution of this study is threefold. First, it operationalises hesitation as an intermediate decision state that can be examined through trial-level and AoI-level ET metrics, rather than focusing only on the final selected option. Second, it shows that hesitation-related behaviour is reflected in a structured visual signature that goes beyond longer decision times, involving temporal, oculomotor and attention-allocation indicators such as saccadic activity, early orienting, fixation- and visit-based inspection, and per-model attention to candidate alternatives. This pattern was also supported by the within-participant sensitivity analysis, which showed that the main differences between hesitation and no-hesitation trials were preserved when participants were compared with themselves. Third, the study proposes a structured metric-based approach that may serve as a diagnostic framework for interpreting hesitation-related visual behaviour during early-stage product evaluation. This approach can help identify product alternatives or attributes that generate uncertainty and may require further design refinement.

Author Contributions: Conceptualization, V.B.-P., J.L.S.-B. and M.V.; methodology, V.B.-P. and M.V.; software, M.-J.A.; validation, V.B.-P., M.-J.A., J.L.S.-B. and M.V.; formal analysis, V.B.-P.; investigation, V.B.-P., M.-J.A., J.L.S.-B. and M.V.; resources, V.B.-P., M.-J.A., J.L.S.-B. and M.V.; data curation, V.B.-P.; writing—original draft preparation, V.B.-P.; writing—review and editing, V.B.-P., M.-J.A., J.L.S.-B. and M.V.; visualization, V.B.-P., M.-J.A., J.L.S.-B. and M.V.; supervision, V.B.-P., M.-J.A., J.L.S.-B. and M.V.; project administration, J.L.S.-B. and M.V.; funding acquisition, J.L.S.-B. and M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universitat Jaume I, grant number GACUJIMB/2024/08.

Institutional Review Board Statement: This study was approved by the Deontological Committee of the Universitat Jaume I (approval number CD/79/2021 and date of approval 15 July 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data will be made available by the authors on request.

Acknowledgments: During the preparation of this manuscript, the authors used Microsoft Copilot (accessed in May 2026) for language editing and text refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ET	Eye-tracking
JO	Jar-opening
AoI	Area of Interest

References

- Orquin, J.L.; Mueller Loose, S. Attention and Choice: A Review on Eye Movements in Decision Making. *Acta Psychol.* **2013**, *144*, 190–206. [CrossRef] [PubMed]
- Orquin, J.L.; Lahm, E.S.; Stojic, H. The Visual Environment and Attention in Decision Making. *Psychol. Bull.* **2021**, *147*, 597–617. [CrossRef] [PubMed]
- Bialkova, S.; Grunert, K.G.; van Trijp, H. From Desktop to Supermarket Shelf: Eye-Tracking Exploration on Consumer Attention and Choice. *Food Qual. Prefer.* **2020**, *81*, 103839. [CrossRef]
- Rojas, J.C.; Marín-Morales, J.; Ausín Azofra, J.M.; Contero, M. Recognizing Decision-Making Using Eye Movement: A Case Study With Children. *Front. Psychol.* **2020**, *11*, 2542. [CrossRef] [PubMed]
- Palacios-Ibáñez, A.; Marín-Morales, J.; Contero, M.; Alcañiz, M. Predicting Decision-Making in Virtual Environments: An Eye Movement Analysis with Household Products. *Appl. Sci.* **2023**, *13*, 7124. [CrossRef]
- Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*; Springer: London, UK, 2003.
- Köhler, M.; Falk, B.; Schmitt, R. Applying Eye-Tracking in Kansei Engineering Method for Design Evaluations in Product Development. *Int. J. Affect. Eng.* **2015**, *14*, 241–251. [CrossRef]
- Kuo, J.Y.; Chen, C.H.; Koyama, S.; Chang, D. Investigating the Relationship between Users' Eye Movements and Perceived Product Attributes in Design Concept Evaluation. *Appl. Ergon.* **2021**, *94*, 103393. [CrossRef] [PubMed]
- Fernández-Serrano, P.; Tarancón, P.; Bonet, L.; Besada, C. Consumers' Visual Attention and Choice of 'Sustainable Irrigation'-Labeled Wine: Logo vs. Text. *Agronomy* **2022**, *12*, 685. [CrossRef]
- Callaway, F.; Griffiths, T.L. Attention in Value-Based Choice as Optimal Sequential Sampling. *PsyArXiv* **2019**, 1–12. [CrossRef]
- Glaholt, M.G.; Reingold, E.M. The Time Course of Gaze Bias in Visual Decision Tasks. *Vis. Cogn.* **2009**, *17*, 1228–1243. [CrossRef]
- Bayarri-Porcar, V.; Roda-Sales, A.; Sancho-Bru, J.L.; Vergara, M. Exploring the Use of Eye Tracking to Evaluate Usability Affordances: A Case Study on Assistive Device Design. *Appl. Sci.* **2025**, *15*, 8376. [CrossRef]
- Rebollar, R.; Lidón, I.; Martín, J.; Puebla, M. The Identification of Viewing Patterns of Chocolate Snack Packages Using Eye-Tracking Techniques. *Food Qual. Prefer.* **2015**, *39*, 251–258. [CrossRef]
- Dosso, C.; Chevalier, A.; Tamine, L.; Paubel, P.V.; Salmerón, L. People Also Ask: How Does This Tool Affect Exploration-Exploitation Strategies with Regard to Prior Domain Knowledge and Search Context? An Eye-Tracking Study. *Appl. Ergon.* **2024**, *121*, 104367. [CrossRef] [PubMed]
- Yang, X.; Krajbich, I. A Dynamic Computational Model of Gaze and Choice in Multi-Attribute Decisions. *Psychol. Rev.* **2023**, *130*, 52–70. [CrossRef] [PubMed]
- Gluth, S.; Deakin, J.; Rieskamp, J. A Theory of Multiattribute Search and Choice. *Psychol. Rev.* **2026**, Online ahead of print. [CrossRef]
- Wang, Y.; Yu, S.; Ma, N.; Wang, J.; Hu, Z.; Liu, Z.; He, J. Prediction of Product Design Decision Making: An Investigation of Eye Movements and EEG Features. *Adv. Eng. Inform.* **2020**, *45*, 101095. [CrossRef]
- Wedel, M.; Pieters, R.; van der Lans, R. Modeling Eye Movements During Decision Making: A Review. *Psychometrika* **2022**, *88*, 697–729. [CrossRef] [PubMed]
- Agost, M.J.; Bayarri-Porcar, V. The Use of Eye-Tracking to Explore the Relationship Between Consumers' Gaze Behaviour and Their Choice Process. *Big Data Cogn. Comput.* **2024**, *8*, 184. [CrossRef]
- Goldberg, J.H.; Helfman, J.I. Comparing Information Graphics: A Critical Look at Eye Tracking. In Proceedings of the Conference on Human Factors in Computing Systems—Proceedings, Atlanta, GA, USA, 10–15 April 2010; pp. 71–78.
- Jamovi. *The Jamovi Project*, Version 2.6; [Computer Software]; Jamovi: Sydney, Australia, 2024. Available online: <https://www.jamovi.org/> (accessed on 20 March 2026).
- R Core Team. *R: A Language and Environment for Statistical Computing*, Version 4.4; [Computer Software]. Available online: <https://cran.r-project.org> (accessed on 7 August 2024).

23. Fox, J.; Weisberg, S.; Price, B.; Adler, D.; Bates, D.; Baud-Bovy, G.; Bolker, B.; Ellison, S.; Firth, D.; Friendly, M.; et al. *Companion to Applied Regression*, R Package Car Version 3.1-5; CRAN: Contributed Packages; Sage: Thousand Oaks, CA, USA, 2026. [[CrossRef](#)]
24. Becker, J.; Viertler, M.; Korn, C.W.; Blank, H. The Pupil Dilation Response as an Indicator of Visual Cue Uncertainty and Auditory Outcome Surprise. *Eur. J. Neurosci.* **2024**, *59*, 2686–2701. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.