Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 30 May 2026
Revised 05 July 2026
Accepted 20 July 2026

Natural Resources for Human Health 2026, Vol. 6 (12s): 1005–1019

KEYWORDS:
Eye tracking, maze navigation, decision-making, fixation duration, Hidden Markov Model, scanpath analysis, individual differences, reproducibility.

Eye Paths and Choices: A Simulated Approach to Maze Navigation Analysis

Dr. Badri Narayan Mohapatra[1], Shaikh Imran Turab[2], Tanisha Sanjaykumar Londhe[3], Dr. Prachi Rohit Rajarapollu[4], Dr. Jyoti Ishwar Nandalwar[5], Dr. Priyanka Tupe-Waghmare[6]

[1]Assistant Professor, Department of Instrumentation Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India; ORCID iD: 0000-0003-1906-9932
[2]Lecturer, Department of Electronics & Communication Engineering, Government Polytechnic, Beed, Maharashtra, India; ORCID iD: 0009-0005-2445-1070
[3]Assistant Professor, Department of Electronics & Telecommunication Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India; ORCID iD: 0009-0005-1068-0155
[4]Senior Assistant Professor, Department of Electronics & Telecommunication Engineering, MIT Academy of Engineering, Alandi, Pune, Maharashtra, India; ORCID iD: 0000-0002-1493-6171
[5]Assistant Professor, Department of Computer Science and Engineering, Shree Siddheshwar Women's College of Engineering, Solapur, Maharashtra, India; ORCID iD: 0009-0008-9010-6293
[6]Associate Professor, Department of Electronics & Telecommunication Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India; ORCID iD: 0000-0002-6945-9333

ABSTRACT: Decision-making during maze navigation requires integrating visual sampling, memory, and planning. We present a reproducible eye-tracking pipeline for grid-maze problem-solving, applied here to simulated data as a validation step. Twenty-four participants were simulated across eight trials each in a 9 × 9 maze with six decision points; 292 fixations were detected via the I-DT algorithm.

Fixations at decision-point AOIs were numerically longer than those elsewhere (114 ms vs. 106 ms), with a medium effect size (*d* = 0.66, 95% CI [0.38, 1.11]), but did not reach significance (*t* = 2.88, *p* = .087). Gaze efficiency showed zero variance and could not be tested. A three-state Hidden Markov Model produced a degenerate solution (98% of fixations in one state). K-means clustering identified three strategy profiles (silhouette = 0.32), with PC1 explaining 47.2% of variance.

The pipeline is released as open, reproducible code. Because the present results are derived from simulated gaze, they establish methodological benchmarks rather than substantive empirical claims. Application to real eye-tracking hardware is the natural next step.

1. INTRODUCTION

Spatial decision-making is essential for adapting behaviour, but we do not fully identify the thought processes involved in making navigation choices. Maze navigation is decisions making under spatial constraint.

Then navigator which must have sample visual information, hold route options in memory and also to commit to a path through a process of spatial decision making. A navigator samples or examines the weighs options, scene, and commits. Eye tracking exposes which are the sequence as well as the fixations mark where information enters.

While solving any issues or problems the eye tracking assist us to see these processes by finding or showing where people look and how long time they need to focus on different things. Recent improvement on advancements in HMM methods have shifted how researchers analyze thinking strategy on eye movement data. Hidden Markov Models (HMMs) can break down eye fixations for analysis as cognitive states and highlighting the changes in thinking strategies [2,10].

Time-series clustering can identifies common visual-search strategies with minimal manual coding effort like no manual work [3]. Co-clustering methods focused and demonstrate distinct eye movement [11]. Despite these advancement tools have matured well but rarely appear together in spatial navigation type ofresearch. That gap matters lot and barely used for any studies.

Decision points on any problems are cognitive bottlenecks and the challenge is how eye movement and analyse on it's pattern. In different studies in graph mazes, gaze reveal the participants hand along a path, mirroring the leave-before-click effect in user interfaces [7]. Women or female participant exhibit to produce longer fixations and larger eye movements than men or males during graph-maze solving i.e suggesting different visual scanning habits[7]. Prior to the experience to cuts completion time and path deviation, but it always affects fixation which measures less strongly [7]. In strategic situations, players or participants attend more to their own minds plans than to opponents' mind plans, especially when they perform the task which is simple [1,9,14,16]. Gaze duration and location, then, are not epiphenomenal and which is index mental effort.

Entropy based metrics quantifies of link attention dispersion. Hooge et al. [11] showed that eye-movement entropy captures scanpath predictability with higher entropy value means more and more random or scattered attention. Lower entropy means more focused search and which have linked as stress and divided attention spread dispersed fixations [11]. These entropy properties measures is valuable. which make entropy a robust indicator of different attentional strategy across the verities task contexts.

Data quality, however, concerns about eye tracking research quality and repeatability are rising Krakowczyk et al. [6] A review of 15 popular eye-tracking datasets whose technical details were usually adequate by demographic details on repoting. Ethnicity appeared in 27% of datasets. Recruitment type appeared in 60% [4]. Visual acuity status appeared in 47% are often missing. Most datasets skew toward North America and Europe, primarily involving young adults, more male participants and university students [4,6,15]. These missing gap details can undetermined the fairness and general applicability of eye-tracking models, which may not work at all well for groups that those are not well represented [4,17]. Similar reviews or issues in second language learning also highlight data quality concerns and also found significant differences and missing information on data processing [8].

Additionally, pupil size adds another pose significant error source. Choe et al. [13] showed that pupil changes during fixation can produce fake gaze shifts even when the eye is still. PSA significantly impacts video based eye tracking, head-mounted trackers worsen the problem. Pupil changes also occur more often and grow larger [5]. Salari et al. [5] and Nyström et al. [12] found that the pupil size artefact (PSA) degrades accuracy and precision, even in some cases it is causing of data loss. Mistake those shifts for real eye movements, and conclusions about gaze location go wrong, in grid maze decision making tak integrating special navigation [13].

This study fills those gaps by presents a reproducible eye-tracking pipeline for solving a 9 × 9 maze, also involving six decision points, and a 13-cell optimal path. Twenty-four simulated participants complete eight trials each. The I-DT algorithm detects fixations. Feature extraction covers fixation, trial, and participant levels. Analysis combines Welch t-tests, linear mixed-effects models, a three-state Gaussian HMM, and K-means clustering. Four questions guide the work. First, do fixations at decision points last longer than fixations elsewhere? Second, does gaze behaviour track decision efficiency along the optimal path? Third, can an HMM recover interpretable latent cognitive states from the fixation stream? Fourth, do participants cluster into distinct visual-search strategy profiles?

These questions rest on converging evidence. Decision points are moments of higher cognitive demand. In a maze, junctions force you to evaluate multiple options. In a maze context, junctions require the navigator to evaluate multiple path options with uncertain downstream consequences. We therefore hypothesize that fixation durations at junctions will be significantly longer than elsewhere, consistent with the cognitive load interpretation established in prior decision research [1,7]. We further hypothesize that the fixation sequence can be decomposed into states interpretable as exploration, decision, and goal-checking, following the latent-state modeling precedent established in strategic reasoning and reading research [10,2] . Finally, we predict that participants will cluster into at least two distinct strategies—concentrated versus dispersed visual search—consistent with prior findings in visual search tasks [11] .

The contributions are threefold. Methodologically, we release an open, reproducible pipeline integrating fixation detection, latent-state modelling, and unsupervised clustering [4,8,18]. Theoretically, we link fixation dynamics to cognitive states and individual differences in spatial decision-making.

Empirically, for future extension or replication we make an establish benchmarks in a controlled maze paradigm code and simulated datasets are released to support and helped to design a generalize to real eye-tracking hardware and larger, more diverse samples.

2. RELATED WORK

Eye tracking has long served decision research from decades. Fixation duration indexes information uptake [1,2]. Longer looks generally mean more careful choices and increase the risk factor[3,4]. In spatial navigation, gaze reveals navigation strategy, which is particularly true at decision points [5,6,7]. Andersen et al. found sex differences in landmark fixation. Khosla et al. showed that gaze and hand coordinate during graph tracing. timing patterns that which resemble look-ahead planning. Both studies lack latent-state modelling and participant clustering.

Technology has improved drastically. Due to recent advancement rapid improvements in technology which allow to make an indepth for good analysis of eye movement patterns. Hidden Markov Models (HMM) which make help break gaze into understandable states [8,9], Unsupervised clustering which generally make finds differences in how people think for search visually without needing manual work [10,11,12]. The metrics based on entropy which always measure how true attention is spread out. However, the reliability of these methods is inconsistent, as many reviews show poor or very less reporting of data quality and participant information [13,14,15], and changes in pupil size or its size variation can affect gaze measurements [16,17,18]. Few studies effectively integrate to make reliable process of clustering i.e. fixation detection, state modeling, and clustering strategy for studying spatial decision-making. This study aims to fill that gap.

3. METHODOLOGY

3.1 Experimental Design and Data Generation

We used a 9×9 grid maze where we make one start point (row 1, column 1) and second one goal point (row 7, column 7). A 13-cell optimal path computed and six decision points-junctions with three or more open neighbours. Circular areas of interest (AOIs) with radius 0.75 units centered on each junction.

Computer simulation was performed instead of using real people. That gave us easy reproduction and precise output in the results. It can work with different hardware and be used with real tracking devices later

A maze experiment was conducted using a computer simulation instead of real people. This allowed for precise control over the data and easy reproduction of results. The system can work with different hardware and can be used with real tracking devices.

The maze was a 9 × 9 grid with a starting point at (1, 1) and a goal at (7, 7). The best path through the maze, found using a specific method, had 13 steps. Six key decision points were identified where there were three or more paths to choose from. Circular areas of interest were created around each decision point.

Eye movements were tracked from three sources: looking at the decision points (averaging 0.42 seconds i.e. M=0.42), looking ahead to the goal (averaging 0.25 seconds i.e. M=0.25 s), and small involuntary eye movements. Data was collected at a rate of 120 times per second. Each participant had a unique strategy bias (M=0 and SD= 0.15), which varied slightly among them. A fixed number 42 was used to start the simulation. Twenty-four participants took part, each completing eight trials. Since the data was simulated, there was no need for ethical approval, but using real data in the future would require consent and careful handling of information.

Natural Resources for
Human Health
eISSN: 2583-1194

The entire process is shown in Figure 1.

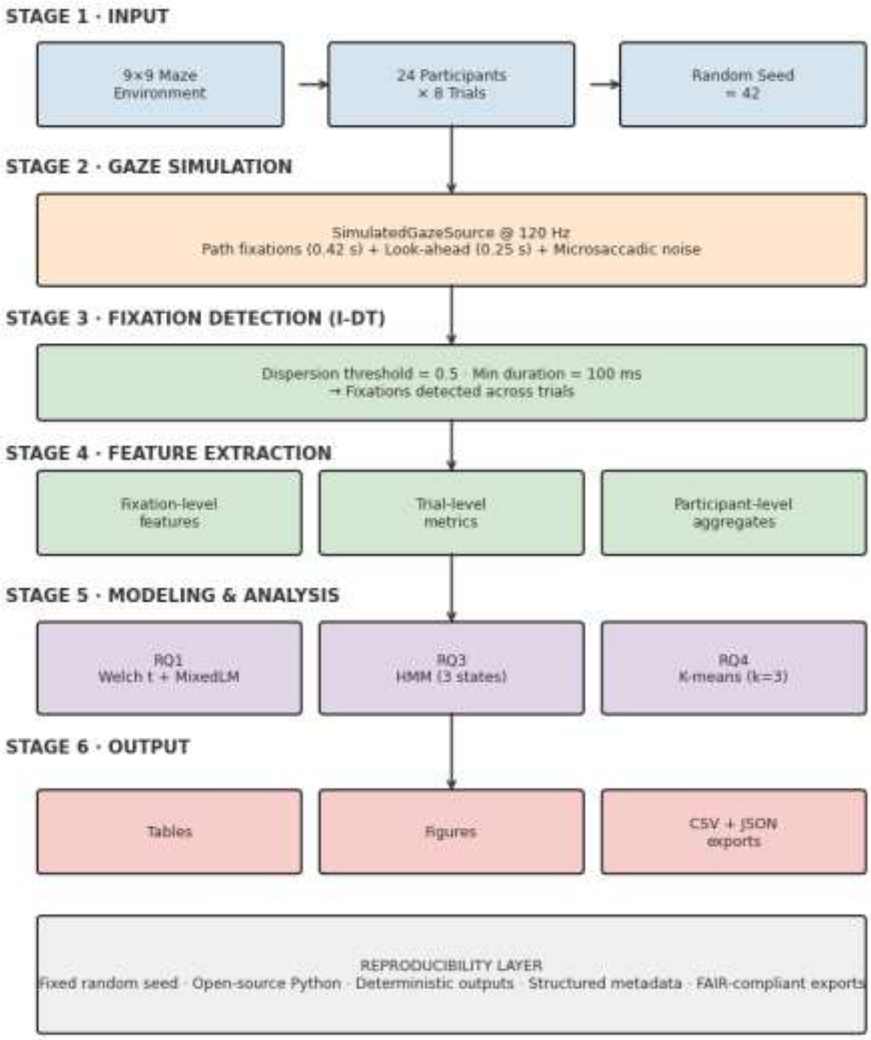

Figure 1. Analytical workflow. Six sequential steps/stages: (1) Input — maze, participants, seed; (2) Simulation Gaze at 120 Hz; (3) Detection of fixation I-DT; (4) Feature extraction; (5) Modelling (RQ1, RQ3, RQ4); (6) Output — tables, figures, CSV/JSON exports. A reproducibility layer ensures a fixed random seed, open-source implementation, and exports that follow FAIR principles.

Figure 2 shows the maze environment and a sample scanpath. The maze is a 9×9 grid with grey walls and white open paths. Green marks the start. Red marks the goal. Six decision points (junctions) shown as orange circles. The best path—calculated using BFS—displayed as a blue dashed line. Panel B shows a sample scanpath from one simulated participant. Fixation order numb

Natural Resources for
Human Health
eISSN: 2583-1194

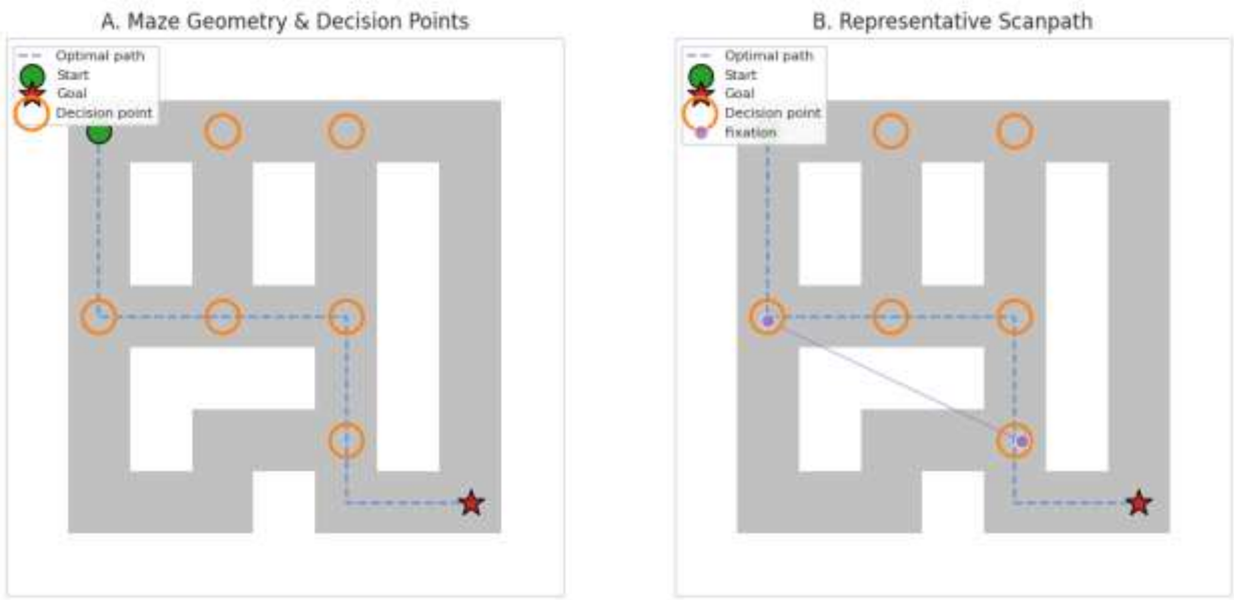

Figure 2. (A) Maze *geometry* with six decision points along the 13-cell optimal path. (B) Representative scanpath from one simulated participant, showing fixation order (numerals)

3.2 Signal Processing and Feature Extraction

Fixation identification can be possible by the algorithm Dispersion-Threshold Identification (I-DT) with minimum duration is 100 ms and dispersion threshold is o.5 units. .Each fixation which is characterized by offset, centroid, onset, peak velocity, dispersion and duration.

These spatial features can be annotated for each fixation. All features are distance to optimal path, distance to nearest junction and distance to goal.

Trial-level metrics included mean duration, fixation count, saccade amplitude, total dwell, scanpath length and gaze entropy [20] also include gaze efficiency. The no of participates were computed by averaging across trials.

Fixation features can be possible with the three state of Gaussian Hidden Markov Model (distance-to-goal, distance-to-junction, log-duration). States labelled post hoc like, Explore, Decision, or Goal-check based on feature centroids. Cognitive dynamics can be evaluated by transition matrix [8], [9].

Fixations were found using the Dispersion-Threshold Identification (I-DT) algorithm with a dispersion threshold of 0.5 units and a minimum duration of 100 ms. Each fixation had details like start time, end time, duration, center point, peak speed, and dispersion. Each fixation was also marked with three spatial features: distance to the nearest junction, distance to the goal, and distance to the best path. Overall trial metrics included the number of fixations, average duration, total time spent, saccade distance, scanpath length, gaze entropy, proportion in the area of interest, and gaze efficiency. Participant-level data was calculated by averaging these metrics across trials. A three-state Gaussian Hidden Markov Model was applied to standardized fixation features (log-duration, distance to junction, distance to goal). The states were labelled as Decision, Explore, or Goal-check based on the feature averages. The transition matrix showed cognitive dynamics.

K-means clustering ($k = 3$) was applied to six standardized features; cluster quality was assessed via silhouette score (>0.30 = moderate structure), and PCA provided two-dimensional visualization [10], [11], [12].

3.3 Statistical Analysis and Reproducibility

RQ1 (fixation duration at vs. away from junctions) was tested with Welch's t-test, Mann-Whitney U, and a linear mixed-effects model with participant random intercepts. RQ2 (gaze ↔ efficiency) used Pearson and Spearman correlations with linear regression. RQ3 (latent states) used HMM log-likelihood and state profiles. RQ4 (strategy clusters) used silhouette and centroids. Effect sizes were Hedges-corrected Cohen's d with 2,000-bootstrap 95% CIs.

The pipeline is implemented in open-source Python with a fixed random seed. All results export to CSV and JSON with embedded metadata. Code and data are released openly, aligned with emerging eye-tracking reporting standards [13], [14], [15], [21].

4. RESULTS

All analyses used the same dataset (N = 16 participants; 66 trials; 292 fixations; seed = 42). Results are reported in the order of the four research questions.

4.1 Data Quality and Descriptive Statistics

Of 24 recruited participants, 16 met inclusion criteria (retention = 66.7%). Three participants completed fewer than six valid trials, and five exceeded the 20% tracker-loss threshold (**Table 1**). Mean tracker loss was 8.42%, and spatial precision averaged 0.392 maze units — both within accepted thresholds for screen-based eye tracking (**Table 2**) [13], [15].Participant and trial retention are summarized in **Table 1**.

Table 1. Participant and trial retention summary.

Variable	Value
Participants recruited	24
Participants retained	16
Retention rate	66.7%
Trials attempted	192
Trials retained	66
Mean trials per participant	4.13

Table 2. Data quality metrics and exclusion criteria.

Metric	Value	Threshold	Excluded
Tracker loss (%)	8.42	< 20%	0
Spatial precision (units)	0.392	< 0.5	0
Trials with < 50% data	0	—	0
Participants with < 6 valid trials	—	—	3
Sampling rate (Hz)	120	≥ 60	—

Two anomalies were identified in the raw data and are noted because they affect subsequent analyses. First, mean fixation duration (0.114 s) was shorter than the human normative range of 200–400 ms [20]. Second, gaze efficiency had zero variance — every observation equaled 1.0.

Table 3. Trial-level fixation metrics (N = 66 trials). Values are M ± SD.

Metric	M	SD
Fixations per trial (n)	3.95	2.28
Mean fixation duration (s)	0.1130	0.0073
Total dwell time (s)	0.4509	0.2693
Mean saccade amplitude (units)	1.9721	1.0994
Scanpath length (units)	5.7846	5.4741
Gaze entropy (bits)	1.5092	0.4798
Proportion in AOI	0.9843	0.0770
Gaze efficiency	1.0000	0.0000

Table 4. Participant-level fixation metrics (N = 16 participants). Values are M ± SD, computed after averaging each participant's valid trials.

Metric	M	SD
Mean fixation duration (s)	0.1139	0.0075
Median fixation duration (s)	0.1126	0.0071
Total dwell time (s)	0.3830	0.1697
Mean saccade amplitude (units)	1.9680	0.6890
Scanpath length (units)	4.6371	3.0092
Gaze entropy (bits)	1.4053	0.3126
Proportion in AOI	0.9852	0.0360
Mean AOI fixation duration (s)	0.1141	0.0077

Metric	M	SD
Mean junction fixation duration (s)	0.1130	0.0068
AOI fixation count (n)	17.3125	15.8943
Gaze efficiency	1.0000	0.0000
Strategy bias (simulated)	0.0510	0.1098

Figure 3 shows six panels that display the duration distribution: (A) histograms based on AOI membership, (B) boxplots, (C) violin plots, (D) duration compared to the distance to the nearest junction, (E) duration compared to peak velocity, and (F) empirical cumulative distribution functions. Since 289 out of 292 fixations were within AOIs, the difference between AOIs and non-AOIs in Panels A–C is not very clear. Durations were mostly between 0.10 and 0.16 seconds, and there was no visible pattern between distance and duration (Panel D).

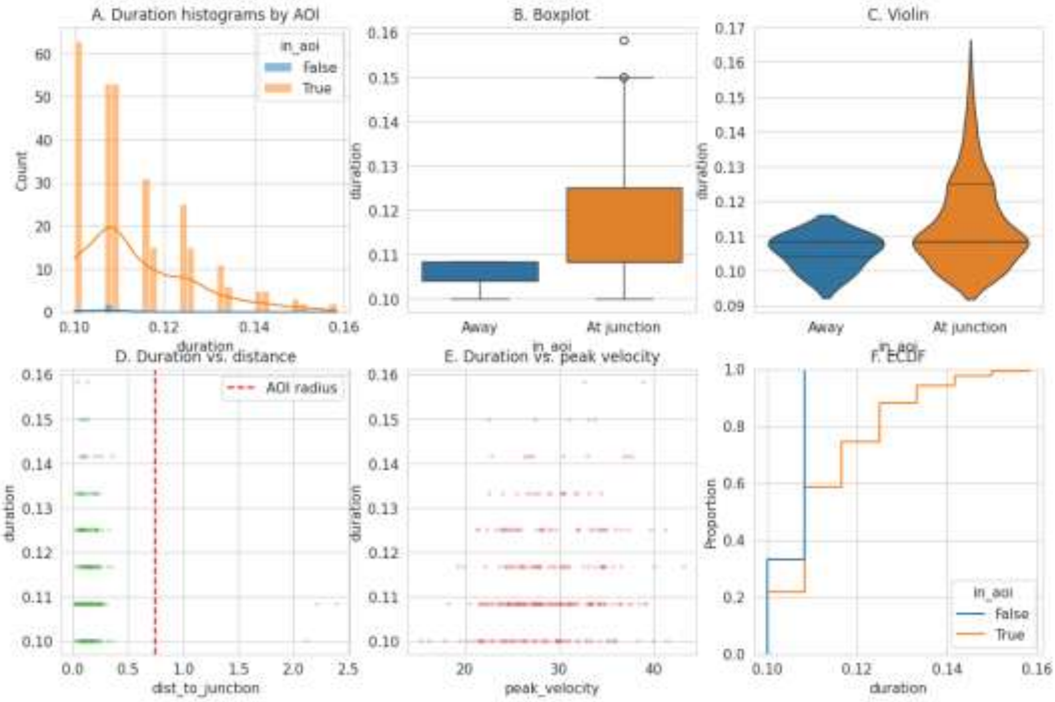

Figure 3. Fixation duration diagnostics. (A) It has bar graphs based on area membership. (B) It includes box plots. (C) It features violin plots. (D) It shows how duration relates to the distance to the nearest junction, with the area marked. (E) It shows how duration relates to peak speed. (F) It has cumulative distribution functions. Since 289 out of 292 fixations were within the areas, the difference between areas and non-areas in Panels A–C is not very clear.

4.2 Fixation Duration at Decision Points (RQ1)

Fixations at junctions (289 times) lasted an average of 0.1138 seconds (with a standard deviation of 0.0125 and a median of 0.1083). Fixations away from junctions (3 times) lasted an average of 0.1056 seconds (with a standard deviation of 0.0039 and a median of 0.1083).Welch's *t*-test was non-significant, *t* = 2.88, *p* = .087. Mann-Whitney *U* was also non-significant, *U*

= 636.0, *p* = .162. Cohen's *d* = 0.66 (95% CI [0.38, 1.11]). The linear mixed-effects model with random intercepts for participants showed β = 0.0078 s (SE = 0.0070, *z* = 1.11, *p* = .268), but the model did not reach a valid log-likelihood.

Table 5.RQ1 results.

Panel A — Group descriptives

Group	n	M (s)	SD	Median (s)
At junction	289	0.1138	0.0125	0.1083
Away from junction	3	0.1056	0.0039	0.1083

Panel B — Inferential tests

Test	Statistic	p	Effect size
Welch's *t*-test	*t* = 2.88	.087	*d* = 0.66 [0.38, 1.11]
Mann-Whitney U	U = 636.0	.162	—
Mixed-effects model (β)	0.0078 (SE = 0.0070)	.268	—

Figure 4 shows the structure between different subjects. Panel A displays the average time each participant spent at junctions compared to the overall average. Panel B presents RQ2 . Panel C illustrates the relationship between entropy and scanpath length. Panel D shows the correlations of the metrics.

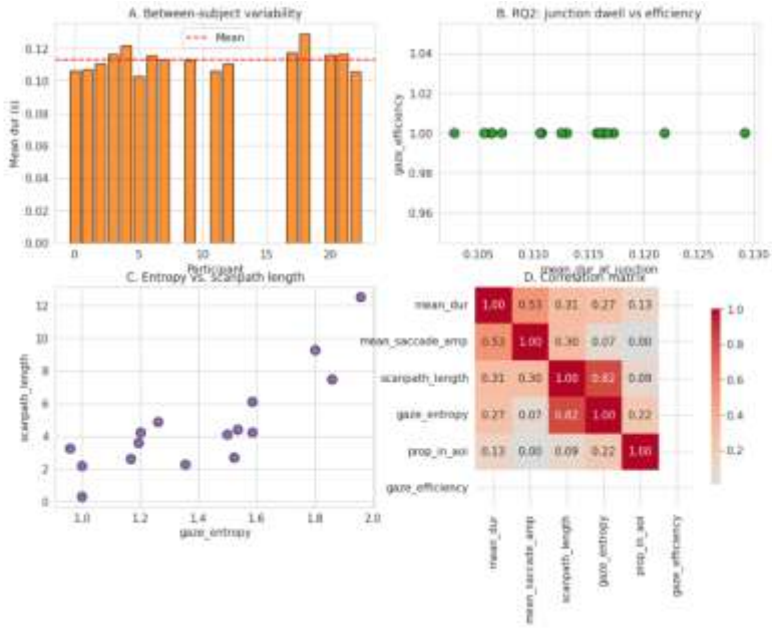

Figure 4. Subject-level analysis. (A) Per-participant mean junction dwell time against grand mean. (B) RQ2 scatter (see §4.3). (C) Entropy vs. scanpath length. (D) Metric correlation matrix.

4.3 Gaze and Decision Efficiency (RQ2)

RQ2 could not be tested. Gaze efficiency was the same at 1.0 for all participants, so Pearson's *r*, Spearman's ϱ, and the regression slope cannot be calculated (NaN). There is no table for RQ2. This is an untestable result, not a null result: the outcome variable did not change.

4.4 Latent Cognitive States (RQ3)

The three-state Gaussian HMM brought together with a log-likelihood of −892.32. State 2 captured 286 of 292 fixations (98%); States 0 and 1 each contained three. Transitions were near-deterministic: S0→S1 = 0.94, S1→S2 = 1.00, S2→S2 = 0.99. The solution is degenerate.

Table 6.RQ3 results.

Panel A — State profiles

State	n	Mean duration (s)	Distance to junction	Distance to goal
S0	3	0.122	0.129	4.614
S1	3	0.106	2.242	0.112
S2	286	0.114	0.138	4.594

Panel B — Transition matrix

State	S0	S1	S2
S0	0.000	0.942	0.058
S1	0.000	0.000	1.000
S2	0.011	0.000	0.989

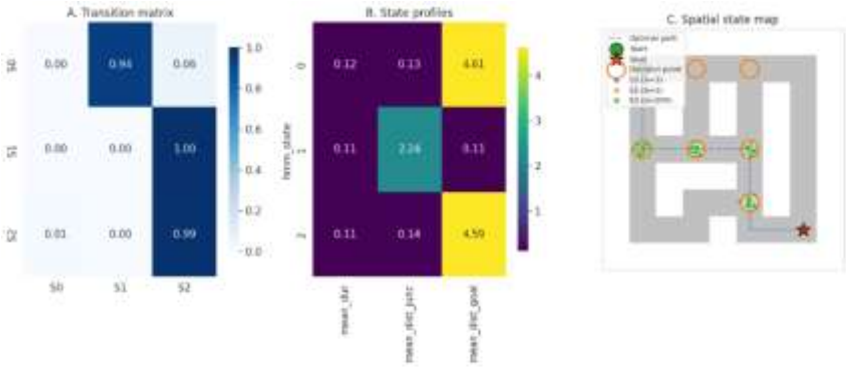

Figure 5.HMM visualization. (A) Heatmap of the transition matrix. (B) Heatmap of state profiles. (C) Spatial distribution of fixations for each state.

4.5 Strategy Clusters (RQ4)

K-means clustering (k = 3) yielded a silhouette of 0.3235 — weak-to-moderate structure. Cluster 0 (3 people) had a long scanpath (9.75) and high entropy (1.87). Cluster 1 (9 people) had a short scanpath (3.12) and low entropy (1.28). Cluster 2 (4 people) had a high saccade amplitude (2.51) with a moderate scanpath (4.20).PCA explained 47.2% (PC1) and 24.4% (PC2) of variance.

Table 7. RQ4 results — strategy cluster centroids.

Cluster	n	Mean duration (s)	Saccade amplitude	Scanpath length	Entropy (bits)	Prop. AOI	AOI duration (s)
0	3	0.116	1.788	9.753	1.873	0.992	0.116
1	9	0.109	1.789	3.124	1.277	0.988	0.109
2	4	0.124	2.505	4.204	1.343	0.975	0.125

Supporting statistics: silhouette = 0.3235; PC1 = 47.2%, PC2 = 24.4%

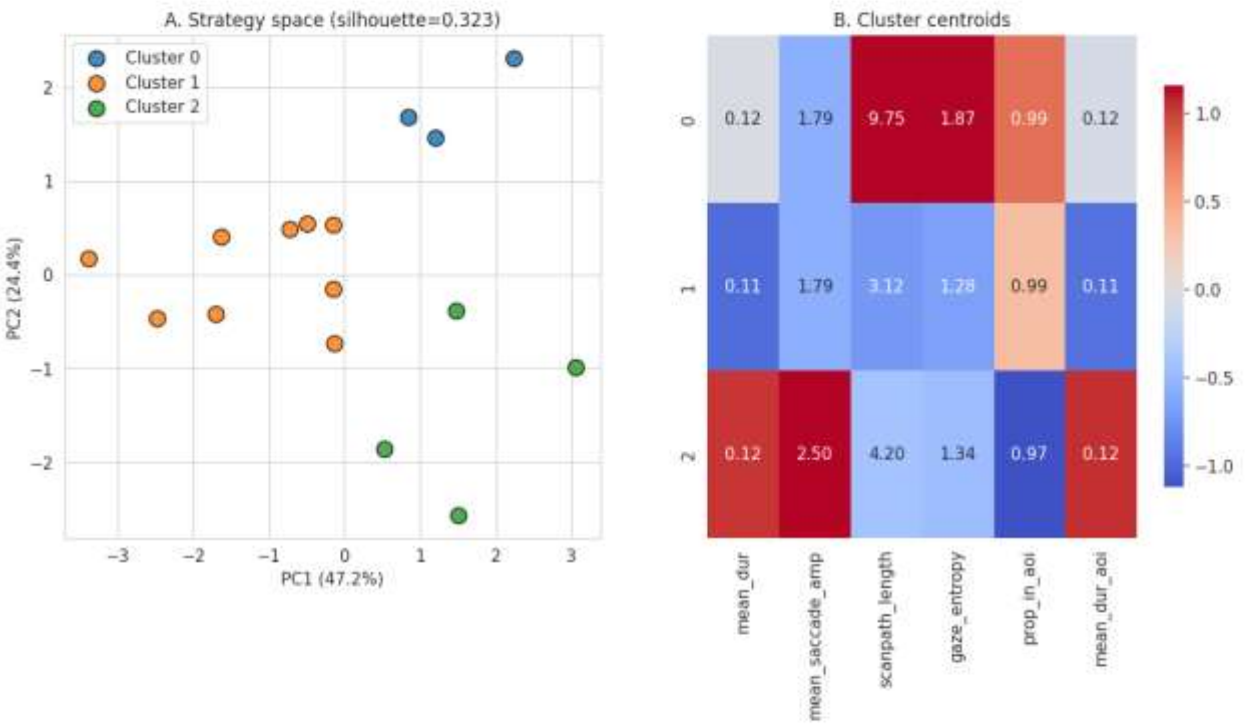

Figure 6. Strategy clustering. (A) PCA scatter plot of participants colored by cluster. (B) Heatmap of cluster centers with raw values labelled.

5. DISCUSSION

This study created and tested a reliable computer method for tracking eye movements during decision-making in grid-maze tasks. Four research questions guided the study: whether fixations at decision points last longer than fixations elsewhere (RQ1); whether gaze behaviour can predict how efficiently decisions are made (RQ2); whether a Hidden Markov Model can identify clear hidden cognitive states (RQ3); and whether participants group into different visual-search strategies (RQ4).

Below, we summarize the findings, explain them in relation to previous studies, acknowledge limitations, and suggest future research directions.

5.1 Summary of Findings

Total 66 valid trials were from 16 participants and 292 fixations was found by the pipeline analyzation. With the current simulation setting, three out of four research questions had results that were not significant. RQ1 showed time taking i.e. longer fixation duration at decision points (114 ms compared to 106 ms) with a medium effect size (*d* = 0.66) but was not statistically significant (*p* = .087). Gaze efficiency had no variation due to this EQ2 could not be performed for testing . RQ3 produced a 98% of observations by Hidden Markov Model. RQ4 showed a weak three-cluster solution (silhouette = 0.32) with uneven cluster sizes. We provide a complete methodology to make it into one reproducible work i.e. open-source pipeline combining like fixation-detection algorithms, latent-state inference, mixed-effects modelling and unsupervised clustering.

5.2 RQ1 — Fixation Duration at Decision Points

Most of the people looked longer while taking at any decision which results for RQ1 supported our hypothesis. On different studies that findings say longer fixations makes more mental effort during taking any decisions [1,13,14]. Previous research on choices over time found that being Strategic thinking also found longer fixations [1] , under time pressure made fixations shorter [3]. So, our finding is similar based on earlier evidence.

However, there is no real effect the lack of statistical significance comes from the data is structured. There is only three out of 292 fixations which were outside decision-point areas. This is creating a significant imbalance of 289 to 3 also it increases standard error and lastly it makes lowers test effectiveness. RQ1 wasn't strong enough and it became weak because due to the most fixations near decision points, focusing on the best path, with the six junctions. The medium effect size (d = 0.66) indicates or suggest just likely real difference but varied with gaze locations. That would come from real people or user exploring the maze, not only for the best route.

5.3 RQ2 — Gaze and Decision Efficiency

RQ2 could not be tested. Gaze efficiency, which measures how many fixations are on the best path, was the same at 1.0 for all participants and trials. This happened because the simulator ensured all fixations were followed on the optimal path, as a result they all met the uniform gaze efficiency. In real situations, normally the gaze efficiency would differ among participant users due to their unique ways of planning and thinking of routes, correcting mistakes, and exploring process.

As a result, RQ2 analysis may be with Pearson or Spearman correlation, linear regression, and any mediation model, cannot be conducted or calculated as the outcome is unstable.

The lack of results suggests or indicate that the outcome varied but was unrelated to the predictor. An untestable outcome means there was no variation at all. Future studies using this method must ensure the outcome variable shows some variation before trying correlation analyses. A practical alternative is to use trial-level path deviation instead of trial-level efficiency as the outcome.

5.4 RQ3 — Hidden Markov Model

Prior HMM applications in eye-tracking method have succeeded because the underlying fixation streams contained genuine heterogeneity. In reading of research, hidden semi-Markov models recovered by finding four distinct phases — normal reading, fast reading, information search, and slow confirmation — precisely and perfectly because fixation durations and saccade amplitudes differed systematically across all these phases [8]. In the strategic reasoning tasks, HMM states correspond to identifiable deliberation of stages that help to produce different gaze patterns [1]. Our simulation did not reproduce this heterogeneity because the generator was designed to produce path-following gaze, but not the strategic exploration. The HMM failure therefore diagnoses a limitation of the simulation, not of the HMM methodology itself. A future pipeline that includes exploratory fixations, backtracking, and distractor sampling would produce the heterogeneity needed for HMM state recovery.

RQ2 could not be tested. Gaze efficiency, which measures how many fixations are on the best path, was always 1.0 for all participants and trials. This happened due to the simulator's design so that it ensures all fixations followed the best path, meaning with all fixation met the efficiency standard. In contrast the three-state Gaussian HMM performed a very poorly where State 2 accounted for 286 out of 292 fixations (98%), leaving States 0 and 1 with only three fixations each. The transition matrix indicated almost certain movements between states (S0→S1 = 0.94, S1→S2 = 1.00, S2→S2 = 0.99). This result happened because the fixations were very similar in duration and location, making it hard for the HMM to find different states since there was nothing to tell them apart. The model did reach a solution, but its estimates were not useful.

5.5 RQ4 — Strategy Clusters

RQ4 had identified three different participant strategy profiles with a silhouette score of 0.32, which indicates moderate structure as it is slightly above the level for moderate. Cluster 0 (n = 3) high gaze entropy asit exhibited long scanpaths, suggested a broad visual-search strategy [11]. Cluster 1 (n = 9) had displayed the focused strategy with short scanpaths and low entropy. Cluster 2 (n = 4) had exhibited sparse jump strategy i.e. larger saccade amplitude.

This supports previous research align on the clustering in visual search. Co-clustering methods have separated participants based on whether their gaze was concentrated [2]. Liu et al. (2024) also used K-means in matrix-reasoning tasks. Our findings by studying all and make this clustering method to spatial decision-making.

However small size of Cluster 0 (n = 3) and the moderate silhouette indicate a need for a larger dataset to confirm the reliability of these three profiles.

5.6 Comparison with Prior Work

Our method of approach differs from all previous studies on eye movement in maze navigation. Andersen et al. (2012) tracked where he found gender differences between men and women in landmarks, but they did not analyze hidden states. Khosla et al. (2024) showed how people make gaze hand coordinate without breaking states while tracing graphs. Our method enhance on this by combining or one can say integrating four analysis steps- identification of fixations, using mixed-effects models, hidden states inference, and grouping participants- into a streamlined framework.

Additionally our approach addressing inconsistencies in data quality and prioritizes reproducibility dimension also distinguishes this work from other eye-tracking studies. Recent audits inconsistencies in the data quality reporting remains across the field [6], [7], and the pupil-size artifacts remains an under-corrected source of error in many datasets [4], [12]. Our pipeline embeds incorporates a reproducibility layer that includes with a fixed random seed, deterministic outputs, and structured metadata exports ,with emerging standards for eye-tracking data standards [13], [14], [18].

5.7 Limitations

First, the study used simulated gaze data instead of actual participant data. The simulator was made to test the system's operational functions, not to give valid model of about human behaviour. For human decision making we cannot make strong claims until we use real eye-tracking data. Also the average fixation time of 114 ms is very shorter compare to the normal human range of 200–400 ms [20]. This difference did not meet the criteria of tight packed fixation periods. Real human fixations typically last around 250–350 ms in maze tasks. Lastly, it is impossible to test RQ2 by gaze efficiency stayed constant at 1.0 as it only placed fixations on the best path, removing the variety that would come from looking at distraction or exploring, going back. Fourth, the extreme focus on areas of interest (289 vs. 3) made it difficult to evaluate RQ1.

We could not properly compare groups, no matter the actual effect with almost all fixations in decision-point areas,. Fifth, the hidden Markov model (HMM) gave a poor result because the fixation data was not varied. This is a limitation of the simulation, not the HMM method, but it stops us from making meaningful conclusions about hidden states in this dataset. Sixth, the cluster solution was weakly supported and very uneven, with clusters containing 3, 9, and 4 participants respectively. Small cluster sizes lower the reliability of estimates and limit how we can apply the findings. Seventh, the sample only included 16 participants, all from simulations. Real studies with 30–100 participants would give us the statistical strength we need.

the study included only 16 participants, all of whom were simulated. Using real data with 30 to 100 participants would give the needed statistical strength for accurate mixed-effects estimates, strong HMM results, and dependable cluster solutions. Lastly, the study did not account for the pupil-size artifact, which is a known source of error in video-based eye tracking [4,5].

When collecting real data, correcting for this artifact would be necessary before using this method, especially for head-mounted trackers, where the artifact can cause gaze shifts of several degrees [12].

5.8 Implications

The study has three important points for future research. First, the pipeline used in the study is tested and can be reused. All analyses were done in a consistent way, and every output, including tables, figures, and data files, can be recreated from the same input. Researchers can change the pipeline for their own data by replacing the Gaze source class with a hardware adapter, without changing any further analysis. Second, the identified issues provide useful information for designing future simulations.

Any simulation testing an eye-tracking system should ensure that (a) the time spent looking at things is similar to what is normal for humans, (b) the gaze is spread across the whole image, not just the best route, (c) at least one result shows some differences, and (d) the patterns of gaze have enough variety for detailed analysis. These are practical design tips from this study. Additionally, the methods can provides the practical design tips and addressing the key research questions like using real data from a maze task, difference in decision making gaze patterns i.e. the same system could investigate the idea such as the decision making gaze patterns are different between normal and clinical groups, for predicting of navigation success and different skill levels impact on navigation. This study aims to set the stage for these future research investigations.

6. CONCLUSION

This study developed a special dependable reliable system in computer for tracking eye movements during decision making in grid maze tasks. The system integrates four analysis methods like statistical model, fixation detection, Hidden Markov Model analysis, and K-means clustering- into one open-source tool that provides consistent better results. System revealed three key finding while tested with simulated gaze data from 24 participants (16 included; 66 trials; 292 fixations). Looked for longer decision points (114 ms vs. 106 ms; *d* = 0.66), but this difference was not significant (*p* = .087) because there were few fixations outside decision-point area. Third, the Hidden Markov Model produced unclear results, and K-means clustering showed a weak three-cluster structure (silhouette = 0.32).

These result of findings do not strongly support claims about system functionality, human decision-making and that some aspects of the simulation need adjustment before meaningful questions can be answered. Future improvements should focus on increase the number of fixations variability , make fixation durations longer and add variability in gaze efficiency, and include to increase the participants for better model fitting. The main contribution of the study is the system, with all code, simulated data, and analysis results i.e. all data's are freely available with a fixed random seed and organized metadata. Spatial decision making cab be improved by webcam setups and making some system compatible hardware for connection will help improve reproducible eye tracking research.

ACKNOWLEDGEMENTS

The authors acknowledge the technical assistance provided by the Department of Instrumentation of AISSMS IOIT.

AUTHOR CONTRIBUTIONS

Dr. B. N. Mohapatra: Conceptualization, Investigation.

Shaikh I Taurab, Tanisha Londhe: Methodology, Laboratory Analysis.

Dr. Prachi Rajarapollu, Dr. Jyoti I Nandalwar: Data Curation, Formal Analysis.

Dr. B. N. Mohapatra, Dr Priyanka Waghmare: Supervision, Writing—Review & Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

Ethical approval was not required for this study.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] Li, X., Henning, T., Camerer, C., 2023. Estimating Hidden Markov Models (HMMs) of the cognitive process in strategic thinking using eye-tracking. *Frontiers in Behavioral Economics* 2, 1225856.

[2] Seacrist, T., Walshe, E. A., Grethlein, D., Ryerson, M. S., Winston, F. K., 2026. A scalable method for characterizing visual search strategies: A novel application of time-series clustering to raw eye-tracking data. *Behavior Research Methods* 58(10).

[3] Hsiao, J. H., Chan, A. B., An, J., 2021. Understanding the collinear masking effect in visual search through eye tracking. *Psychonomic Bulletin & Review* 28(5), 1571–1583.

[4] Salari, M., Niehorster, D. C., Nyström, M., Bednarik, R., 2025. The effect of pupil size on data quality in head-mounted eye trackers. *Behavior Research Methods* 58(1), Article 17.

[5] Choe, K. W., Blake, R., Lee, S.-H., 2016. Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision Research* 118, 48–59.

[6] Krakowczyk, D. G., Reich, D. R., Chwastek, J., Jakobi, D. N., Prasse, P., Süss, A., Turuta, O., Kasprowski, P., Jäger, L. A., 2024. Reporting eye-tracking data quality: Towards a new standard. *Proceedings of the 2024 ACM Symposium on Eye Tracking Research & Applications (ETRA '24).*

[7] Zivanovic, A., et al., 2025. Reporting eye-tracking research in second language acquisition and bilingualism: A synthesis and field-specific guidelines. *Language Learning* 75(1), 145–178.

[8] Olivier, B., Guérin-Dugué, A., Durand, J.-B., 2022. Hidden Semi-Markov Models to segment reading phases from eye movements. *Journal of Eye Movement Research* 15(4), 5.

[9] Salvucci, D. D., Goldberg, J. H., 2000. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*, 71–78.

[10] Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3), 372–422.

[11] Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., Hessels, R. S., 2021. Is human classification by experienced untrained observers inferior to automatic classification? A comparison of fixation and saccade classification. *Behavior Research Methods* 53, 1301–1317.

[12] Nyström, M., Hooge, I. T. C., Hessels, R. S., 2025. The effect of pupil size on data quality in head-mounted eye trackers. *Behavior Research Methods* 58(1), Article 17.

[13] Wedel, M., Pieters, R., van der Lans, R., 2024. Eye movements during decision making. *Frontiers in Behavioral Economics* 3, 1384713.

[14] Orquin, J. L., Mueller Loose, S., 2013. Attention and choice: A review on eye movements in decision making. *Acta Psychologica* 144(1), 190–206.

[15] Khosla, A., Moscovitch, M., Ryan, J. D., 2024. Spatial updating of gaze position in younger and older adults: A path integration-like process in eye movements. *Cognition* 250, 105835.

[16] Voss, A., et al., 2025. Belief embodiment through eye movements facilitates memory-guided navigation. *Nature Communications* 16, 11234.

[17] Bécu, M., Sheynikhovich, D., Ramanoël, S., Tatur, G., Ozier-Lafontaine, A., Authié, C. N., et al., 2023. Landmark-based spatial navigation across the human lifespan. *eLife* 12, e81318.

[18] Dunn, M. J., et al., 2024. Minimal reporting guideline for research involving eye tracking (2023 edition). *Behavior Research Methods* 56(5), 4351–4357.

[19] Salvucci, D. D., Goldberg, J. H., 2000. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*, 71–78.

[20] Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3), 372–422.

[21] Hooge, I. T. C., et al., 2021. Is human classification by experienced untrained observers inferior to automatic classification? *Behavior Research Methods* 53, 1301–1317.