Supplementary Information

Archaeological expertise shapes the visual exploration and recognition of Palaeolithic art

Olivia Rivero[1*], M. Soledad Beato[2], Mar Suarez[3], Ana M. Mateo-Pellitero[1], Alicia Alvarez-Martinez[2], Javier Eseverri[2], Xabier Eguilleor-Carmona[1] & Miguel García-Bustos[1]

[1]Faculty of Geography and History, University of Salamanca, Salamanca, Spain. [2]Faculty of Psychology, University of Salamanca, Salamanca, Spain. [3]Faculty of Education Sciences, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain. [*]E-mail: oliviariver@usal.es

Contents

Supplementary Methods 1. Exploratory gaze–recognition models

Supplementary Methods 2. Eye-tracking models and AOI operations

Supplementary Methods 3. Taxonomic recognition and response coding

Supplementary Tables 1–11

Supplementary Figs. 1–3

Throughout, the *20-s window* denotes the complete presentation (primary analysis) and the *17-s window* the dependent sensitivity window that excludes the first 3 s of viewing. Abbreviations: AOI, area of interest; BH, Benjamini–Hochberg; CI, confidence interval; CrI, credible interval; E, experts; GEE, generalized estimating equations; HR, hazard ratio; N, non-experts; OR, odds ratio; pp, percentage points; RR, rate ratio; s.d., standard deviation.

Supplementary Methods

Supplementary Methods 1. Exploratory gaze–recognition models

These exploratory analyses related gaze allocation to strict taxonomic identification (binary outcome) in 172 responses from 28 participants (90 from experts and 82 from non-experts) to images 01, 02, 03, 06, 07, 08, 09 and 10. Unscored, partial-credit and broadly compatible responses were excluded from the likelihood rather than recoded as failures.

Two Bernoulli–logit models were fitted, one for each gaze predictor: the proportion of AOIs fixated (X1) and the AOI fixation assignment rate per second of exposure (X2). Each model included centred group membership, within- and between-participant components of the gaze predictor, a non-centred participant intercept and eight image effects constrained to sum to zero through seven orthonormal contrasts. Participant means were computed separately for each window before outcomes were selected, and predictors in both windows were scaled by the standard deviations of the 20-s window. Assignments to overlapping AOIs were retained for X2.

Models were fitted in PyMC 6.3.2 (ArviZ 1.3.0 for diagnostics) with the No-U-Turn sampler, using four chains, a target acceptance rate of 0.95 and a maximum tree depth of 12. Reference fits in the 20-s window used 2,000 warm-up and 2,000 retained draws per chain; all other fits used 1,500 of each. Fits were accepted when all parameters had R-hat < 1.01, bulk and tail effective sample sizes were ≥ 400, there were no divergent transitions or maximum-depth saturations and the minimum chain BFMI was ≥ 0.3. All 108 fits met these criteria.

Sensitivity analyses examined the 17-s window, recording validity $\geq 80\%$, prolonged invalid intervals, continuous recording quality, image exclusions, regularization strength, participant and image omissions, and a power likelihood that equalized participant contributions. Two bounding scenarios assigned all-incorrect or all-correct outcomes to the 37 unscored responses eligible for completion (N = 209); eleven further unscored responses were not eligible because a hypothetical correct outcome would contradict the existing coding: the ten unscored responses to image 07, for which the coding established that no participant had identified the depicted animal (wolf or bear), and one expert response to image 10 that had been coded as describing pigment only, without a figure. This restriction was introduced after an initial unrestricted version of the scenarios (N = 220) had been fitted; it was based on the coding constraints, not on the sign or size of the estimates, and the unrestricted scenarios are not reported. These scenarios did not alter the 172 observed scores. Supplementary Table 9 reports six conditions for each predictor, chosen by type of sensitivity analysis rather than by result; all other results are available with the analysis code.

A complementary conditional logistic model conditioned on the number of correct responses per image. Image 07 contributed no information because all its outcomes were incorrect, leaving seven informative images with 154 responses (82 from experts and 72 from non-experts) and all 28 participants. Confidence intervals were obtained from 1,999 participant bootstrap resamples stratified by group, treating repeatedly sampled participants as distinct individuals; resamples in which the model failed were discarded, with a maximum acceptable failure rate of 5%. Holm correction was applied to the two within-participant terms in each window (Supplementary Table 10).

Hierarchical odds ratios are expressed per standard deviation of the predictor in the 20-s window, with central 95% credible intervals, which are not adjusted for multiplicity; conditional models report bootstrap confidence intervals. Because the associations depended on regularization and data assumptions, neither approach established a consistent association between gaze and recognition.

Supplementary Methods 2. Eye-tracking models and AOI operations

Rates were computed using actual recorded exposure. Coverage was the number of AOIs fixated divided by the number of AOIs in the image, not the proportion of image area inspected. Fixation counts and durations retained multiple assignments in overlapping polygons, whereas recording-quality summaries counted time within any AOI only once, using the geometric union of the AOIs.

Coverage was analysed with a participant-clustered quasi-binomial GEE with image effects, an independence working correlation and the bias-reduced (Mancl–DeRouen) sandwich covariance,

weighting images equally. Image 01, which all participants fixated, was retained in descriptive and marginal summaries but could not contribute to finite logistic estimation. Restricted times to first AOI fixation assigned non-events a common horizon no longer than the shortest exposure, and participant means (with images weighted equally) were compared with Welch's *t*-test. AOI fixation assignments were analysed with a negative-binomial GEE with image effects and a log-exposure offset. Holm correction was applied to the three primary comparisons within each window (Supplementary Table 2 for the 17-s window).

The exploratory BH family comprised four tests: visit rate, glance rate, assigned fixation duration per second of exposure and mean assigned fixation duration, analysed with Poisson, Poisson, Tweedie (log link) and gamma (log link) GEEs, respectively. Visit and glance counts were identical in these data and are therefore shown as a single row in Supplementary Table 3; both tests were retained in the four-test family. Mean fixation duration is conditional on at least one assigned fixation.

AOI-level models included image-centred log polygon area and weights inversely proportional to the number of AOIs in each image. Because the Wald covariance matrices for fixation probability, assigned fixation count and assigned fixation duration were not positive definite, these outcomes were analysed with crossed participant/AOI bootstrap inference (1,999 resamples, with participants resampled within group and AOIs within image; seed 20261013). Fixation probability models failed in 12 (20-s window) and 5 (17-s window) resamples; count and duration models had no failures. Intervals are percentile intervals, and bootstrap *P* values are centred with a plus-one correction (minimum attainable $P = 5 \times 10^{-4}$). First AOI entry was analysed with an image-stratified Cox model with crossed participant/AOI covariance. Group and area effects formed separate four-test BH families within each window (Supplementary Table 4).

Recording-quality restrictions and adjustments for valid recording time, prolonged invalid intervals, age and prior familiarity were sensitivity analyses and did not change the primary sample. Supplementary Table 5 and Supplementary Fig. 2 report these variants for the 20-s window. Supplementary Table 11 reports further sensitivity analyses for self-reported visual problems and for reported familiarity with the stimuli. Supplementary Table 6 summarizes leave-one-participant-out estimates as ranges, which should not be read as confidence limits.

Supplementary Methods 3. Taxonomic recognition and response coding

Strict taxonomic identification was assessed for images 01, 02, 03, 06, 07, 08, 09 and 10. Images 04 and 05 were included in gaze and descriptive analyses but excluded from this outcome: although image 05 is archaeologically attributed to a right-facing goat, the fragment presented was not diagnostic enough to require that identification. Taxon, anatomy, figuration and technical or material observations were coded separately; describing a figure or a colour did not by itself establish a taxon or the presence of pigment.

Only responses scored 0 or 1 entered the strict outcome: 90 from experts and 82 from non-experts, of which 46 and 10, respectively, were correct. Mixed alternatives receiving partial credit, broadly compatible responses without a partial score and responses that could not be scored for strict taxonomic identification were kept as separate categories (Supplementary Table 7). The absence of a taxon was not automatically treated as an error or as a failure to perceive a figure.

Each participant contributed their proportion of correct responses among binary-evaluable responses, with participants weighted equally. Groups were compared with a two-sided studentized participant-label permutation test (99,999 permutations), and pointwise percentile intervals were obtained from 19,999 group-stratified participant bootstrap resamples, conditional on the fixed image set and on binary evaluability. The 0/0.5/1 partial-credit index and image-exclusion analyses were complementary (Supplementary Table 8). Image-specific recognition counts, including those for image 01, are descriptive; no test comparing painted and engraved stimuli, and no equivalence test, was performed.

Supplementary Tables

Supplementary Table 1 | Archaeological descriptions of the 37 areas of interest

AOI	Archaeological description
01_1	Outline of a headless, left-facing bison painted in red, comprising the cervico-dorsal line, haunch line, beginning of the hind leg and ventral line. Natural rock relief partly forms the haunch and the beginning of the hind limb.
02_1	Cervico-dorsal curve of an engraved, headless bison facing left.
02_2	Tail of an engraved bison depicted vertically and facing downwards.
02_3	Hind leg of the same vertical, downward-facing engraved bison as AOI 02_2.
02_4	Foreleg of the same vertical, downward-facing engraved bison.
02_5	Hump line, horn, fronto-nasal line and eye of the vertical, downward-facing engraved bison.
02_6	Head and mane of a right-facing horse: fronto-nasal line, muzzle and jaw, with eye, nose, mouth, two ears and a double linear mane.
02_7	Left-facing goat represented by two horns, head, chest line, ventral line and cervico-dorsal line.
03_1	Hump line of the bison.
03_2	Two bison horns superimposed on remains of red pigment.
03_3	Hind leg of the bison.
04_1	Engraved head of an aurochs or goat facing left, including the fronto-nasal line, eye, jaw line and part of the horns.
04_2	Leg of an animal, possibly a goat, whose taxon cannot be determined because distinctive anatomical details are absent; it appears to face right. The region also includes part of the haunch and cervico-dorsal lines.
04_3	Cervico-dorsal line of an indeterminate animal lacking diagnostic anatomical details; orientation indeterminate.
04_4	Aurochs horn oriented to the right.
04_5	Two hind legs shown in anatomically correct perspective, belonging to an indeterminate right-facing animal; the tail and haunch line are also preserved.
05_1	Goat ear or horn (the ambiguity is retained).
05_2	Forelegs of the goat shown in perspective.
05_3	Cervico-dorsal line, tail, haunch line and two hind legs of the goat.
06_1	Horse mane formed by short parallel strokes.
06_2	Cervico-dorsal line of the horse.
06_3	Ventral line of the horse.
06_4	Chest and jaw lines of the horse.
07_1	Animal head, including muzzle, fronto-nasal line, ears, chest line and beginning of the back line.
07_2	Forelegs of the animal.
07_3	Ventral line formed by a series of short parallel strokes.
07_4	End of the hind leg.
08_1	Head of the goat.
08_2	Forelegs and parts of the chest and ventral lines of the goat.
08_3	Hind legs of the goat extended backwards.
08_4	Cervico-dorsal line of the goat.
09_1	Muzzle and jaw line of the animal.
09_2	Eye of the animal.
09_3	Ear or horn of the animal.
09_4	Ear or horn of the animal.
10_1	Bison horns, with remains of red pigment within the AOI.
10_2	Cervico-dorsal line, haunch line and hind leg of the bison.

Descriptions retain the anatomical and taxonomic uncertainties of the archaeological analysis. The interpretation of an image as a whole is not automatically transferred to each of its regions, and the content listed for an AOI does not by itself imply that the AOI is taxonomically diagnostic.

Supplementary Table 2 | Primary eye-tracking outcomes in the 17-s sensitivity window

Outcome	Experts	Non-experts	Effect	95% CI	*P*	P_{Holm}
AOIs fixated (%)	82.3 (6.2)	72.7 (8.7)	OR 1.84	1.25 to 2.70	0.003	0.003
Restricted time to first AOI fixation (s)	0.96 (0.46)	2.45 (1.09)	Δ −1.50	−2.17 to −0.83	1.8×10^{-4}	3.6×10^{-4}
AOI fixation assignment rate (s^{-1})	1.16 (0.15)	0.72 (0.23)	RR 1.65	1.35 to 2.02	2.5×10^{-5}	7.5×10^{-5}

Group values are means (s.d.) of participant means over ten equally weighted images (n = 14 per group). OR, quasi-binomial GEE; Δ, difference in restricted time, experts minus non-experts (Welch's t-test); RR, negative-binomial GEE with log-exposure offset. Holm correction was applied across the three contrasts. Results for the 20-s window are given in Table 3 of the main text.

Supplementary Table 3 | Complementary eye-tracking outcomes in both windows

Window	Outcome (unit)	Experts	Non-experts	Ratio	95% CI	*P*	q_{BH}	*n*
20 s	Visit/glance rate (s^{-1})	0.582 (0.076)	0.435 (0.082)	1.34	1.17 to 1.53	1.0×10^{-4}	1.4×10^{-4}	280
20 s	Assigned fixation duration per exposure (s s^{-1})	0.340 (0.031)	0.210 (0.069)	1.65	1.35 to 2.01	1.9×10^{-5}	7.5×10^{-5}	280
20 s	Mean assigned fixation duration (s)	0.303 (0.029)	0.293 (0.042)	1.03	0.94 to 1.14	0.53	0.53	280
17 s	Visit/glance rate (s^{-1})	0.576 (0.077)	0.430 (0.081)	1.34	1.17 to 1.53	1.1×10^{-4}	1.5×10^{-4}	280
17 s	Assigned fixation duration per exposure (s s^{-1})	0.351 (0.030)	0.214 (0.072)	1.66	1.36 to 2.03	1.7×10^{-5}	6.9×10^{-5}	280
17 s	Mean assigned fixation duration (s)	0.306 (0.030)	0.296 (0.042)	1.04	0.94 to 1.14	0.48	0.48	279

Group values are means (s.d.) of participant means; *n*, presentations (all models include 28 participants). Visit and glance counts were identical and are shown in one row; both tests remain in the four-test BH family. Assigned duration may count the same fixation in overlapping AOIs and is therefore not unique dwell time. Mean duration is conditional on at least one assigned fixation.

Supplementary Table 4 | AOI-level models adjusted for AOI area

Window	Outcome	Term	*n*	Estimate	95% CI	*P*	q_{BH}
20 s	Fixation probability (OR)	Group (E vs N)	1,008	2.50	1.30 to 5.79	0.025	0.025
20 s	Fixation probability (OR)	Log area within image	1,008	9.41	4.94 to 51.60	0.012	0.012
20 s	Assigned fixation count (ratio)	Group (E vs N)	1,036	1.67	1.31 to 2.08	5.0×10^{-4}	0.0010
20 s	Assigned fixation count (ratio)	Log area within image	1,036	2.88	2.06 to 4.22	5.0×10^{-4}	6.7×10^{-4}
20 s	Assigned fixation duration (ratio)	Group (E vs N)	1,036	1.63	1.31 to 2.03	0.0010	0.0013
20 s	Assigned fixation duration (ratio)	Log area within image	1,036	2.87	2.05 to 4.35	5.0×10^{-4}	6.7×10^{-4}
20 s	First AOI entry (HR)	Group (E vs N)	1,036	1.59	1.35 to 1.87	3.5×10^{-6}	1.4×10^{-5}
20 s	First AOI entry (HR)	Log area within image	1,036	2.97	2.17 to 4.05	1.4×10^{-7}	5.5×10^{-7}
17 s	Fixation probability (OR)	Group (E vs N)	1,008	2.27	1.22 to 4.89	0.027	0.027
17 s	Fixation probability (OR)	Log area within image	1,008	6.82	4.17 to 20.70	0.0035	0.0035
17 s	Assigned fixation count (ratio)	Group (E vs N)	1,036	1.67	1.27 to 2.10	5.0×10^{-4}	6.7×10^{-4}

17 s	Assigned fixation count (ratio)	Log area within image	1,036	2.85	2.03 to 4.22	5.0×10^{-4}	6.7×10^{-4}
17 s	Assigned fixation duration (ratio)	Group (E vs N)	1,036	1.63	1.29 to 2.06	5.0×10^{-4}	6.7×10^{-4}
17 s	Assigned fixation duration (ratio)	Log area within image	1,036	2.86	2.00 to 4.20	5.0×10^{-4}	6.7×10^{-4}
17 s	First AOI entry (HR)	Group (E vs N)	1,036	1.62	1.31 to 2.02	1.0×10^{-4}	4.1×10^{-4}
17 s	First AOI entry (HR)	Log area within image	1,036	2.54	1.92 to 3.36	3.2×10^{-7}	1.3×10^{-6}

Records are AOI × participant observations from all 28 participants. Image 01, fixated by all participants, was excluded from finite estimation of fixation probability (1,008 records); other models used 1,036. Fixation probability, count and duration use crossed participant/AOI bootstrap inference (1,999 resamples; 5.0×10^{-4} is the smallest attainable *P*); first AOI entry uses an image-stratified Cox model. Group and area effects form separate four-test BH families within each window.

Supplementary Table 5 | Sensitivity of the primary outcomes to recording quality and covariates (20-s window)

Outcome	Variant	*n*	Estimate	95% CI	*P*	P_{Holm}
AOIs fixated (OR)	Reference	252 presentations	1.88	1.29 to 2.75	0.0021	0.0021
AOIs fixated (OR)	Valid-time adjustment	252 presentations	1.80	1.23 to 2.63	0.0040	–
AOIs fixated (OR)	Invalid-run adjustment	252 presentations	1.86	1.27 to 2.73	0.0027	–
AOIs fixated (OR)	≥80% valid recording time	249 presentations	1.87	1.27 to 2.74	0.0026	–
AOIs fixated (OR)	Age and familiarity	252 presentations	1.74	1.15 to 2.64	0.011	–
Time to first AOI fixation (Δ, s)	Reference	28 participants	−1.05	−1.60 to −0.50	7.9×10^{-4}	0.0016
Time to first AOI fixation (Δ, s)	Valid-time adjustment	28 participants	−1.23	−1.77 to −0.70	7.1×10^{-5}	–
Time to first AOI fixation (Δ, s)	Invalid-run adjustment	28 participants	−1.11	−1.71 to −0.51	8.2×10^{-4}	–
Time to first AOI fixation (Δ, s)	≥80% valid recording time	277 presentations	−1.05	−1.55 to −0.56	2.0×10^{-4}	–
Time to first AOI fixation (Δ, s)	Age and familiarity	28 participants	−1.03	−1.64 to −0.42	0.0020	–
AOI fixation rate (RR)	Reference	280 presentations	1.64	1.34 to 2.00	3.0×10^{-5}	9.0×10^{-5}
AOI fixation rate (RR)	Valid-time adjustment	280 presentations	1.61	1.31 to 1.97	6.4×10^{-5}	–
AOI fixation rate (RR)	Invalid-run adjustment	280 presentations	1.63	1.33 to 1.99	3.7×10^{-5}	–
AOI fixation rate (RR)	≥80% valid recording time	277 presentations	1.62	1.33 to 1.99	3.9×10^{-5}	–
AOI fixation rate (RR)	Age and familiarity	280 presentations	1.65	1.34 to 2.04	4.6×10^{-5}	–

The primary sample comprises 280 presentations from 28 participants; image 01 is omitted from finite estimation of coverage (252 presentations). The ≥80% restriction changes the presentations analysed, not the participants included. Covariate models for restricted time use participant-level ordinary least squares with HC3 standard errors; the ≥80% restriction for restricted time uses image-standardized means with a participant-stratified bootstrap (4,999 resamples; 2.0×10^{-4} is the smallest attainable *P*). Holm-adjusted values are given for the reference analyses only. Plotted in Supplementary Fig. 2.

Supplementary Table 6 | Leave-one-participant-out ranges of the primary effects

Window	Outcome	Scale	Minimum	Maximum	Direction (28 fits)
20 s	AOIs fixated	OR	1.76	2.03	All > 1

20 s	Time to first AOI fixation	Δ (s)	−1.16	−0.94	All < 0
20 s	AOI fixation rate	RR	1.59	1.70	All > 1
17 s	AOIs fixated	OR	1.70	1.99	All > 1
17 s	Time to first AOI fixation	Δ (s)	−1.59	−1.28	All < 0
17 s	AOI fixation rate	RR	1.60	1.72	All > 1

Each range summarizes 28 point estimates, each obtained after omitting one participant (27 participants and ten images per fit). Ranges are minima and maxima of point estimates, not confidence intervals. All fits converged.

Supplementary Table 7 | Taxonomic response categories by image

Group	Image	Presentations	Correct (1)	Incorrect (0)	Partial credit	Broadly compatible	Unscored (strict taxon)	Binary-evaluable
E	01	14	11	2	0	1	0	13
E	02	14	9	3	0	0	2	12
E	03	14	1	8	0	0	5	9
E	06	14	10	4	0	0	0	14
E	07	14	0	8	0	0	6	8
E	08	14	8	5	1	0	0	13
E	09	14	5	8	0	0	1	13
E	10	14	2	6	0	0	6	8
E	**Total**	**112**	**46**	**44**	**1**	**1**	**20**	**90**
N	01	14	6	6	0	0	2	12
N	02	14	2	6	1	0	5	8
N	03	14	0	10	0	0	4	10
N	06	14	0	8	1	0	5	8
N	07	14	0	10	0	0	4	10
N	08	14	0	9	0	0	5	9
N	09	14	2	11	0	0	1	13
N	10	14	0	12	0	0	2	12
N	**Total**	**112**	**10**	**72**	**2**	**0**	**28**	**82**

Only correct (1) and incorrect (0) responses enter the strict binary outcome; the binary-evaluable column is a subtotal, not an additional category. Images 04 and 05 were not scored for taxonomic identification. In image 01, correct identifications were 11/14 and 6/14 presentations (78.6% and 42.9%), or 11/13 and 6/12 binary-evaluable responses (84.6% and 50.0%); six of the ten correct non-expert identifications across the eight images occurred in image 01. These counts are descriptive.

Supplementary Table 8 | Recognition outcomes and image-exclusion sensitivity analyses

Outcome	Images excluded	Responses (E/N)	Mean, E/N (%)	Difference (pp)	95% CI (pp)	P	P_{Holm}
Strict identification	04, 05	90/82	50.9/13.7	37.2	22.4 to 51.7	9.0×10^{-5}	9.0×10^{-5}
Partial-credit index (0/0.5/1)	04, 05	91/84	51.0/14.8	36.2	21.7 to 50.7	1.0×10^{-4}	1.0×10^{-4}
Strict, excluding image 02	02, 04, 05	78/74	47.4/12.2	35.1	20.3 to 49.9	1.7×10^{-4}	5.6×10^{-4}
Strict, excluding image 08	04, 05, 08	77/73	49.3/16.1	33.2	16.8 to 48.9	8.6×10^{-4}	0.0017
Strict, excluding images 02 and 08	02, 04, 05, 08	65/65	44.2/14.3	29.9	13.3 to 45.6	0.0022	0.0022
Strict, excluding images 07 and 09	04, 05, 07, 09	69/59	61.3/16.0	45.4	27.1 to 62.7	1.4×10^{-4}	5.6×10^{-4}
Proportion of unscored responses	04, 05	112/112	17.9/25.0	−7.1	−20.5 to 6.3	0.39	0.39

Means weight each of the 14 participants per group equally and are conditional on the evaluable responses and fixed image set of each row. Intervals are pointwise percentile intervals from 19,999 group-stratified participant

bootstrap resamples; *P* values are from two-sided studentized label permutation tests (99,999 permutations). Holm correction was applied within families: the four image-exclusion analyses formed one family, and the strict outcome, the partial-credit index and the proportion of unscored responses were each tested on their own. The partial-credit index and the proportion of unscored responses are distinct outcomes, not replacements for the strict comparison.

Supplementary Table 9 | Hierarchical models of gaze and recognition (selected conditions)

Predictor	Condition	*N* responses	OR (median)	95% CrI
X1, AOIs fixated	Reference (20-s window)	172	1.305	0.649 to 2.708
X1, AOIs fixated	17-s window	172	1.006	0.525 to 1.857
X1, AOIs fixated	Stronger regularization (20-s window)	172	1.216	0.703 to 2.108
X1, AOIs fixated	Images 02 and 08 excluded	130	1.224	0.523 to 2.891
X1, AOIs fixated	Bounding scenario: unscored responses incorrect	209	1.224	0.635 to 2.428
X1, AOIs fixated	Bounding scenario: unscored responses correct	209	0.982	0.633 to 1.548
X2, AOI fixation rate	Reference (20-s window)	172	1.929	1.004 to 3.815
X2, AOI fixation rate	17-s window	172	2.005	1.107 to 3.713
X2, AOI fixation rate	Stronger regularization (20-s window)	172	1.634	0.969 to 2.844
X2, AOI fixation rate	Images 02 and 08 excluded	130	1.601	0.736 to 3.597
X2, AOI fixation rate	Bounding scenario: unscored responses incorrect	209	1.658	0.925 to 3.121
X2, AOI fixation rate	Bounding scenario: unscored responses correct	209	1.336	0.840 to 2.149

Odds ratios are within-participant effects per s.d. of the predictor in the 20-s window, with central 95% credible intervals (not adjusted for multiplicity). The same six conditions were reported for both predictors, selected by type of sensitivity analysis rather than by result. $N = 172$ observed responses; bounding scenarios ($N = 209$) hypothetically complete 37 unscored responses.

Supplementary Table 10 | Within-participant conditional logistic models of gaze and recognition

Window	Predictor	OR	95% CI	*P*	P_{Holm}	Failed resamples	Responses (E/N)
20 s	X1, AOIs fixated	1.09	0.56 to 3.14	0.80	0.80	13/1,999	82/72
20 s	X2, AOI fixation rate	1.47	0.78 to 4.37	0.33	0.67	10/1,999	82/72
17 s	X1, AOIs fixated	0.86	0.44 to 2.00	0.69	0.69	8/1,999	82/72
17 s	X2, AOI fixation rate	1.56	0.81 to 4.42	0.26	0.52	8/1,999	82/72

Seven informative image strata contained 154 binary responses from all 28 participants; image 07 (all responses incorrect) contributes no conditional information. Odds ratios are per s.d. of the predictor in the 20-s window. Intervals are participant bootstrap percentile intervals (1,999 resamples); *P* values are approximate centred bootstrap values, with Holm correction over the two within-participant terms in each window.

Supplementary Table 11 | Sensitivity of the primary outcomes to visual problems, reported familiarity and eye-tracker validation accuracy (20-s window)

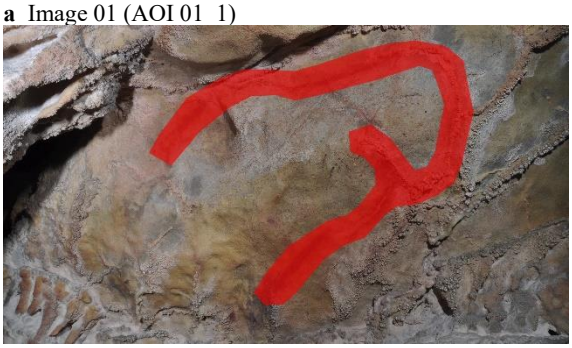
Variant	Outcome	*n*	Estimate	95% CI	*t*	*P*
Reference	AOIs fixated (OR)	28	1.88	1.29 to 2.75	3.42	0.0021
Reference	AOI fixation rate (RR)	28	1.64	1.34 to 2.00	5.04	3.0×10^{-5}
Reference	Time to first AOI fixation (Δ, s)	28	−1.05	−1.60 to −0.50	−3.96	7.9×10^{-4}
Reference	Identification (Δ, pp)	28	37.2	22.4 to 51.7	4.83	9.0×10^{-5}
Adjusted for visual problems	AOIs fixated (OR)	28	1.69	1.18 to 2.41	3.00	0.0059
Adjusted for visual problems	AOI fixation rate (RR)	28	1.55	1.27 to 1.89	4.49	1.3×10^{-4}

Adjusted for visual problems	Time to first AOI fixation (Δ, s)	28	−1.03	−1.57 to −0.49	−3.90	6.4×10^{-4}
Adjusted for visual problems	Identification (Δ, pp)	28	34.2	15.4 to 53.0	3.75	9.4×10^{-4}
Participants reporting visual problems only	AOIs fixated (OR)	22	1.68	1.12 to 2.51	2.66	0.015
Participants reporting visual problems only	AOI fixation rate (RR)	22	1.62	1.31 to 2.01	4.76	1.2×10^{-4}
Participants reporting visual problems only	Time to first AOI fixation (Δ, s)	22	−1.00	−1.60 to −0.39	−3.45	0.0029
Participants reporting visual problems only	Identification (Δ, pp)	22	32.9	11.4 to 54.5	3.34	0.0062
Excluding the two experts reporting familiarity	AOIs fixated (OR)	26	1.71	1.18 to 2.47	2.97	0.0067
Excluding the two experts reporting familiarity	AOI fixation rate (RR)	26	1.63	1.32 to 2.01	4.84	6.2×10^{-5}
Excluding the two experts reporting familiarity	Time to first AOI fixation (Δ, s)	26	−1.04	−1.61 to −0.47	−3.80	0.0011
Excluding the two experts reporting familiarity	Identification (Δ, pp)	26	36.3	18.0 to 54.6	4.19	6.7×10^{-4}
Excluding presentations of images reported as known	AOIs fixated (OR)	28	1.82	1.25 to 2.65	3.25	0.0031
Excluding presentations of images reported as known	AOI fixation rate (RR)	28	1.64	1.34 to 2.01	4.97	3.6×10^{-5}
Excluding presentations of images reported as known	Time to first AOI fixation (Δ, s)	28	−0.91	−1.54 to −0.28	−2.96	0.0066
Adjusted for validation accuracy	AOIs fixated (OR)	28	1.89	1.29 to 2.77	3.41	0.0021
Adjusted for validation accuracy	AOI fixation rate (RR)	28	1.65	1.36 to 1.99	5.38	1.3×10^{-5}
Adjusted for validation accuracy	Time to first AOI fixation (Δ, s)	28	−1.04	−1.61 to −0.46	−3.72	0.0010
Adjusted for validation accuracy	Identification (Δ, pp)	28	36.5	20.5 to 52.5	4.71	7.9×10^{-5}
Excluding presentations with validation accuracy > 1.5°	AOIs fixated (OR)	28	1.88	1.27 to 2.79	3.33	0.0026
Excluding presentations with validation accuracy > 1.5°	AOI fixation rate (RR)	28	1.62	1.32 to 1.98	4.93	4.1×10^{-5}
Excluding presentations with validation accuracy > 1.5°	Time to first AOI fixation (Δ, s)	28	−1.17	−1.75 to −0.59	−4.22	4.3×10^{-4}

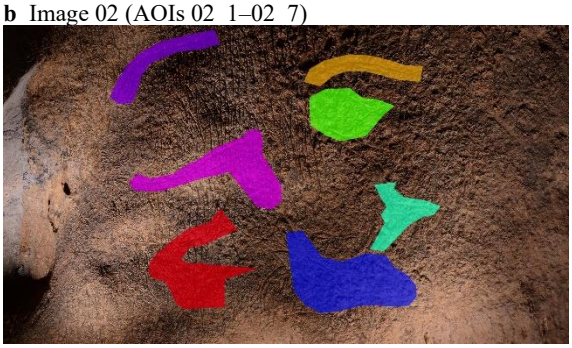
n, participants. Visual problems were self-reported by 9 of 14 experts and 13 of 14 non-experts. Coverage and fixation rate use the primary GEE specifications (participant clusters, independence working correlation, bias-reduced sandwich covariance, *t* reference with $n - 2$ degrees of freedom), with a participant-level visual-problem indicator added where stated. Time to first AOI fixation uses Welch's test on participant means, or ordinary least squares with HC3 standard errors when adjusted. Identification uses Welch's test on participant rates, or ordinary least squares with HC3 standard errors when adjusted. The photographs were unpublished; two experts nevertheless stated that they recognised three and five images, respectively (plus two doubtful), which was attributed to confusion with other representations. Both were retained in the primary analyses; two variants exclude them, or only their presentations of those images. Validation accuracy is the mean angular offset of the four-point validation preceding each image (Tobii Pro Lab); it enters the GEEs as a presentation-level covariate and the participant-level models as the participant mean. The 1.5° threshold excluded 31 of 280 presentations (13 from non-experts and 18 from experts; 30 of 252 in the coverage model), with time to first AOI fixation averaged over the remaining images. Results for the 17-s window were concordant and are provided with the analysis code. Values are unadjusted for multiplicity.

Supplementary Figures

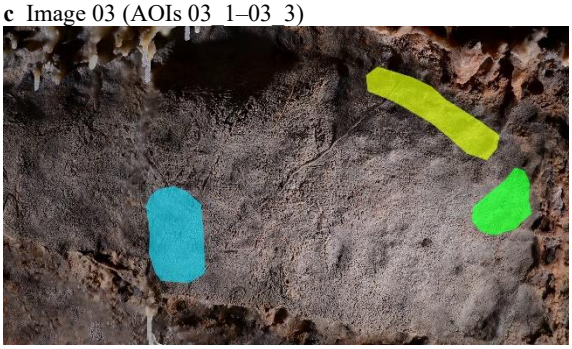
a Image 01 (AOI 01_1)



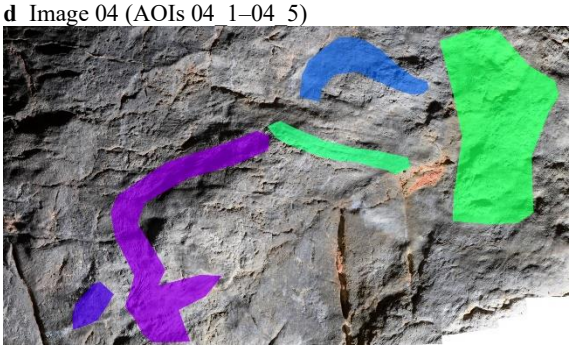
b Image 02 (AOIs 02_1–02_7)



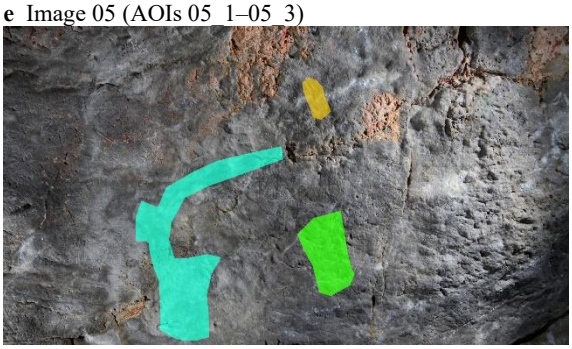
c Image 03 (AOIs 03_1–03_3)



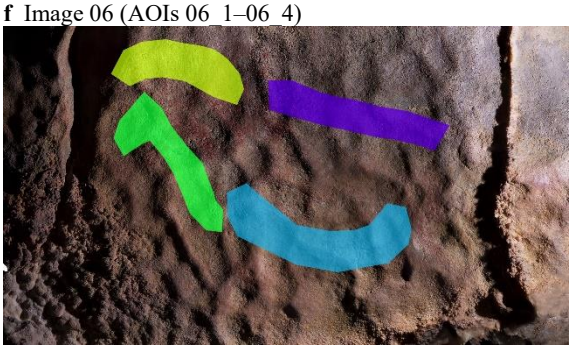
d Image 04 (AOIs 04_1–04_5)



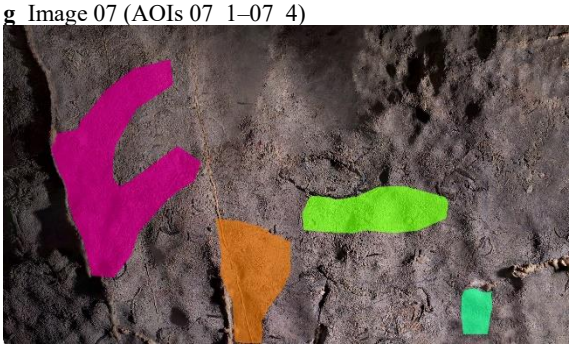
e Image 05 (AOIs 05_1–05_3)



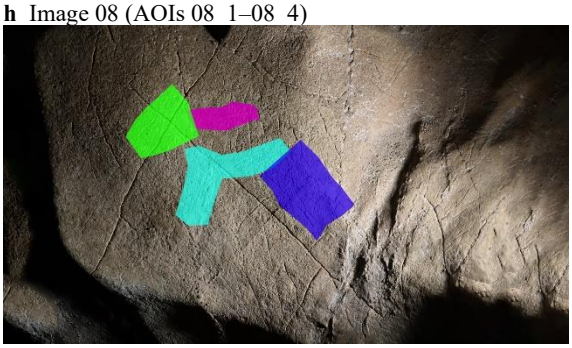
f Image 06 (AOIs 06_1–06_4)



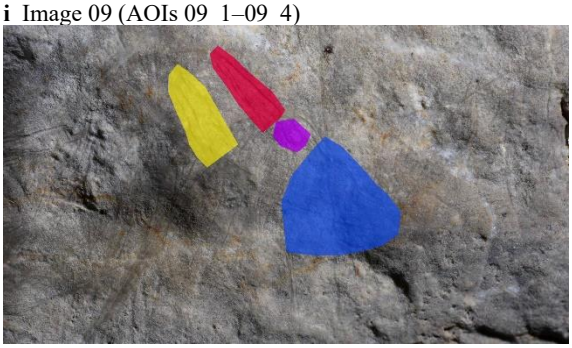
g Image 07 (AOIs 07_1–07_4)



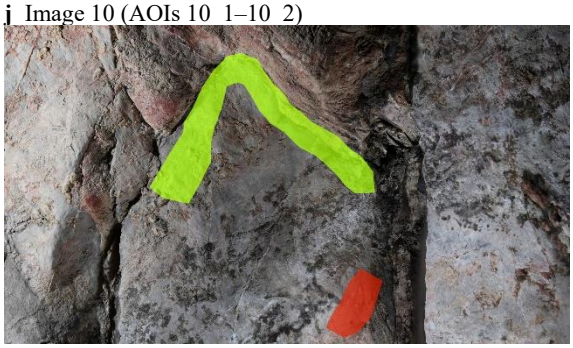
h Image 08 (AOIs 08_1–08_4)



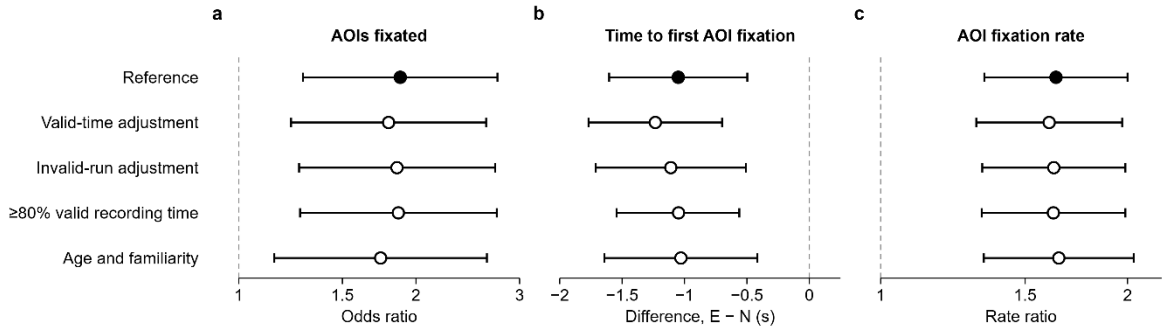
i Image 09 (AOIs 09_1–09_4)



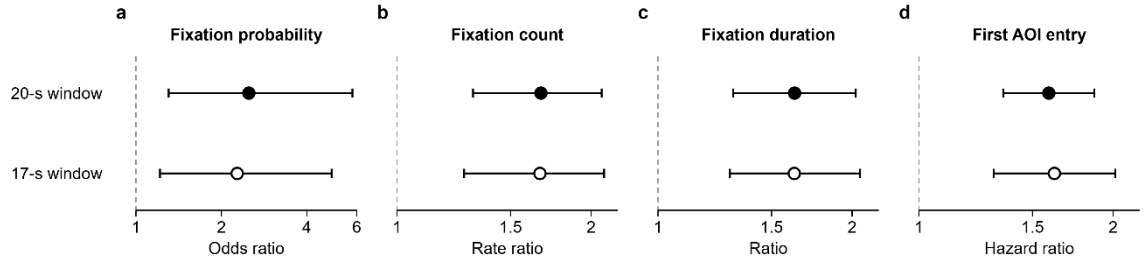
j Image 10 (AOIs 10_1–10_2)



Supplementary Fig. 1 | Experimental stimuli and areas of interest. a–j, The ten photographs of Upper Palaeolithic art (images 01–10) with the 37 AOIs used in the analyses overlaid; frames are taken from the stimulus videos and show no gaze data. Colours identify individual AOIs and carry no quantitative information. Archaeological descriptions of each AOI are given in Supplementary Table 1.



Supplementary Fig. 2 | Sensitivity of the primary eye-tracking outcomes to recording quality and covariates (20-s window). a, AOIs fixated (odds ratio, quasi-binomial GEE). **b**, Restricted time to first AOI fixation (difference between group means, experts minus non-experts). **c**, AOI fixation assignment rate (rate ratio, negative-binomial GEE). Filled circles, reference estimates; open circles, sensitivity variants; bars, 95% CIs. Ratio axes are logarithmic; dashed lines mark no group difference. The ≥80% valid-recording restriction changes the presentations analysed, not the participants included (n = 14 per group). Estimates are given in Supplementary Table 5.



Supplementary Fig. 3 | Group effects adjusted for AOI area. a, Fixation probability (odds ratio). **b**, Assigned fixation count (rate ratio). **c**, Assigned fixation duration (ratio). **d**, First AOI entry (hazard ratio). Estimates compare experts with non-experts after adjustment for image-centred log AOI area, in the 20-s (filled circles) and 17-s (open circles) windows; the two windows are not independent. Bars, 95% CIs (crossed participant/AOI bootstrap in **a–c**; image-stratified Cox model in **d**). Axes are logarithmic; dashed lines mark a ratio of 1. n = 28 participants; 1,008 (**a**) or 1,036 (**b–d**) AOI records. Estimates are given in Supplementary Table 4.