Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Archaeological expertise shapes the visual exploration and recognition of Palaeolithic art

Olivia Rivero Vila

oliviariver@usal.es

University of Salamanca

Maria Beato

University of Salamanca

Mar Suarez

University of Las Palmas de Gran Canaria

Ana M. Mateo-Pellitero

University of Salamanca

Alicia Alvarez-Martinez

University of Salamanca https://orcid.org/0009-0008-7453-3236

Javier Eseverri

University of Salamanca https://orcid.org/0009-0008-1013-0681

Xabier Eguilleor-Carmona

University of Salamanca

Miguel García-Bustos

University of Salamanca

Article

Keywords:

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11772932/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

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.

[*]Corresponding author. E-mail: oliviariver@usal.es

Abstract

Much of Europe's Upper Palaeolithic art consists of faint, fragmentary engravings entangled with the natural relief of the rock, so recognising it is a perceptual problem as much as an interpretive one. We asked whether expertise in this tradition changes how people look at it. Fourteen specialists in Palaeolithic art and 14 non-experts viewed ten photographs of predominantly engraved Upper Palaeolithic motifs while their eye movements were recorded, then reported what they saw. Experts fixated a larger share of archaeologically defined areas of interest, reached them about one second sooner and fixated them more often, differences that held after accounting for area size and across sensitivity analyses. Experts also correctly identified the depicted animals nearly four times as often (50.9% versus 13.7%). Recognising Palaeolithic imagery thus relies on learned perceptual skill, suggesting that seeing these images, not only making them, was perhaps part of what Palaeolithic visual traditions transmitted.

Images are among the oldest cultural technologies of *Homo sapiens*. In Europe, Upper Palaeolithic art spans more than 20,000 years and an exceptional diversity of techniques and styles, and it has long been treated as a privileged window onto symbolic behaviour, visual culture and the cognitive capacities of early modern humans.[1–3] Research on this record has focused overwhelmingly on the people who made the images. Yet images are also things that must be seen, and decades of work on perceptual learning show that what observers are able to see depends on what they have learned to look for.[4] If the capacities that support cultural practices are themselves partly acquired through those practices,[5] then the perception of Palaeolithic imagery, and not only its production, becomes a question about human learning. On the production side, technical, experimental and cognitive approaches have established that Palaeolithic image-making involved skill, practice and apprenticeship. Variability in execution quality, technical procedures and graphic control has been interpreted as evidence for the coexistence of more and less skilled image-makers,[6–8] and experiments show that producing Palaeolithic-style images requires specific visuomotor routines that must be learned and transmitted.[9,10]

Recognition, however, poses its own difficulties. Although Palaeolithic art is popularly imagined as instantly legible, much of the corpus is in fact very hard to perceive. Many images are located in deep, unlit and poorly accessible parts of caves; many are fragmentary, partially preserved or superimposed; and many exploit, and blend into, the natural morphology of their support. Engraving, which accounts for a substantial part of the record, was frequently used to produce motifs that are small, fine and of very low visibility. Identifying such images is therefore not a passive visual act but requires separating intentional marks from natural features, completing partial forms and inferring what a set of lines depicts. These demands raise the possibility that reliable recognition depends on domain-specific perceptual skills acquired through training and experience.

In many other domains, expertise is known to reshape eye-movement behaviour, including fixation patterns, attentional strategies and visual memory.[11–14] Gaze differences between experts and novices have been documented in chess,[15,16] radiology and medical image interpretation[17,18] and other forms of professional vision.[19] Trained observers do not merely know more; they look differently. Eye tracking makes these differences measurable: the number, duration and spatial distribution of fixations, and the time taken to first fixate task-relevant regions, index how observers select and search visual information,[20–22] although the specific signatures of expertise depend on the task and stimuli.[12,13]

Within cognitive archaeology, eye tracking has recently been applied to the perception of the archaeological record itself, including prehistoric pottery,[23] Lower Palaeolithic stone tools[24] and Palaeolithic rock art.[25] Silva-Gago and colleagues showed that viewers of Palaeolithic art direct their attention preferentially to the depicted figures, especially animal heads, rather than to salient features of the cave wall, and that incomplete figures may be perceived as complete through Gestalt organisation and top-down processing.[25] A complementary line of work used virtual-reality reconstructions of decorated caves with the art digitally removed and found that the natural topography of the walls drew attention to the very features later used for figurative depictions, supporting a role for pareidolia in the genesis of cave art.[26,27] These studies show how closely the perception of Palaeolithic imagery is bound to its support, but none has asked whether expertise in this graphic tradition changes how it is seen.

Here we tested whether specialists in Palaeolithic art explore and recognise Palaeolithic imagery differently from non-experts. Using a stimulus set dominated by low-visibility engravings, we compared the two groups on the proportion of archaeologically defined areas of interest (AOIs) they fixated, the time they took to first fixate any AOI and the rate at which they fixated AOIs, and we assessed their accuracy in identifying the depicted animals. We found that experts covered more of the archaeologically relevant graphic information,

reached it sooner and fixated it more often, and that they identified the depicted taxa far more accurately; within participants, identification tended to be more accurate when AOIs were fixated more often, but this trial-level association was imprecise and not robust across analyses.

Results

Experimental design

Fourteen experts in Palaeolithic art (mean rock-art research experience 26.7 ± 12.4 years) and 14 non-experts without domain-specific training took part (Table 1). Both groups comprised 12 men and 2 women and were of similar age (54.1 ± 14.6 and 53.6 ± 10.6 years). Each participant viewed ten unpublished photographs of Upper Palaeolithic art from four Spanish caves (La Garma, Atxurra, Micolón and El Pindal) for 20 s each while their gaze was recorded, and after each image reported whether and what they had recognised (Methods). One stimulus was painted in red and nine were predominantly engraved; the set spanned anatomically incomplete figures, very fine low-visibility engravings and panels with multiple superimposed or closely associated representations, several of which also contained natural fissures, relief features or bear claw marks (Table 2 and Supplementary Fig. 1). Thirty-seven AOIs (1–7 per image) were delineated on archaeological criteria around the graphic elements most informative for identifying each motif. All 28 participants contributed all ten presentations. Because every participant fixated the single AOI of image 01, that image carried no information for the coverage model, which was therefore estimated from the other nine images (252 presentations).

Experts cover more of the relevant graphic information, sooner and more often

Experts and non-experts differed on all three primary measures of visual exploration (Table 3 and Fig. 1). Experts fixated a greater proportion of AOIs than non-experts (84.5 ± 6.1%

versus 75.6 ± 6.9%; odds ratio (OR) = 1.88, 95% confidence interval (CI) 1.29–2.75; Wald $t(26) = 3.42$; $P_{Holm} = 0.0021$; Fig. 1a). They also reached the first AOI sooner (restricted time to first AOI fixation 1.33 ± 0.46 versus 2.38 ± 0.88 s; difference = −1.05 s, 95% CI −1.60 to −0.50; Welch $t(19.7) = -3.96$; $P_{Holm} = 0.0016$; Fig. 1b) and fixated AOIs at a higher rate (1.14 ± 0.16 versus 0.71 ± 0.23 AOI fixation assignments per second of exposure; rate ratio (RR) = 1.64, 95% CI 1.34–2.00; Wald $t(26) = 5.04$; $P_{Holm} = 9 \times 10^{-5}$; Fig. 1c). Expressed as standardised differences between participant means, these effects were large (Hedges' g = 1.33, 1.45 and 2.12, respectively).

Experts also accumulated more AOI-assigned fixation time per second of exposure (ratio = 1.65, 95% CI 1.35–2.01; $q_{BH} = 7.5 \times 10^{-5}$), but mean fixation duration did not differ between groups (ratio = 1.03, 95% CI 0.93–1.14; $q_{BH} = 0.53$; Supplementary Table 3). The expert advantage therefore reflected more numerous and more widely distributed fixations on archaeologically relevant regions, rather than longer individual fixations.

Expertise effects are not explained by area size, timing or recording quality

Because the probability of fixating a region scales with its size, we fitted complementary AOI-level models that included each AOI's image-centred log area as a covariate. Group differences persisted: compared with non-experts, experts showed a higher probability of fixating an AOI (OR = 2.50, 95% CI 1.30–5.79; $q_{BH} = 0.025$), more assigned fixations (RR = 1.67, 95% CI 1.31–2.08; $q_{BH} = 0.0010$), longer assigned fixation duration (ratio = 1.63, 95% CI 1.31–2.03; $q_{BH} = 0.0013$) and a higher rate of first AOI entry (hazard ratio = 1.59, 95% CI 1.35–1.87; $q_{BH} = 1.4 \times 10^{-5}$; Supplementary Table 4 and Supplementary Fig. 3). These models account for AOI area but not for every aspect of the spatial configuration of the stimuli.

Excluding the first 3 s of viewing, when gaze is most strongly driven by stimulus onset, preserved the direction and approximate magnitude of all three primary effects

(Supplementary Table 2). Restricting the analyses to high-quality recordings and adjusting for valid recording time, prolonged tracking loss, age and prior familiarity with the stimuli yielded concordant estimates (Supplementary Table 5 and Supplementary Fig. 2), and no single participant drove the results (Supplementary Table 6). Non-experts more often reported visual problems (13 versus 9 of 14), but adjusting for visual problems, or restricting the analyses to the 22 participants who reported them, left all three effects and the recognition difference essentially unchanged (for example, coverage OR = 1.69, 95% CI 1.18–2.41; Supplementary Table 11). The same held when the two experts who mistakenly believed they had seen some of the images were excluded (coverage OR = 1.71, 95% CI 1.18–2.47; recognition difference = 36.3 percentage points, 95% CI 18.0–54.6; Supplementary Table 11).

Experts identify the depicted animals far more accurately

Taxonomic identification was scored for the eight stimuli whose preserved graphic elements were diagnostic enough to allow it. Of 224 responses, 172 could be scored as strictly correct or incorrect (90 from experts and 82 from non-experts), of which 46 and 10, respectively, were correct (Table 4). The remaining responses typically described anatomical, figurative or technical features without committing to a taxon, and were not necessarily recognition failures. On average, experts correctly identified 50.9% of their scorable responses and non-experts 13.7% (difference = 37.2 percentage points, 95% bootstrap CI 22.4–51.7; Hedges' g = 1.77; studentised t = 4.83, two-sided permutation test, $P = 9 \times 10^{-5}$; Fig. 2). These participant-level estimates weight all participants equally and are conditional on the fixed set of eight images.

The single painted stimulus behaved differently from the engravings. Image 01, a headless bison painted in red, was correctly identified by 11 of 14 experts (78.6%) and 6 of 14 non-experts (42.9%) (84.6% (11/13) and 50.0% (6/12) of scorable responses; Supplementary Table 7). This one image accounted for 60% (6/10) of all correct identifications made by non-

experts, and every participant fixated its single AOI. Thus, for non-experts, successful identification was concentrated on the most conspicuous, high-contrast representation, whereas experts also identified the low-visibility engravings.

In exploratory hierarchical models that separated within- and between-participant variation, identification tended to be more accurate on trials with more frequent AOI fixation, but this trial-level association was imprecise and not robust across analyses (Supplementary Methods 1 and Supplementary Tables 9 and 10). The within-participant association with AOI fixation rate was positive in all six reported conditions (odds ratio per s.d. 1.34–2.01), but its 95% credible interval excluded 1 only in the reference (1.93, 95% CrI 1.004–3.82) and 17-s models, and the corresponding conditional logistic estimates were imprecise (for example, 1.47, 95% CI 0.78–4.37). AOI coverage showed no association (odds ratios 0.98–1.31). Complementary recognition indices and image-exclusion analyses are reported in Supplementary Table 8.

Discussion

Expertise in Palaeolithic art was associated with a distinctive way of looking at Palaeolithic imagery. Compared with non-experts, specialists explored a larger share of the archaeologically informative regions of each image, reached them about a second earlier and fixated them more often, and they identified the depicted animals roughly four times as accurately. Archaeological expertise is therefore expressed not only in how Palaeolithic images are interpreted but also in how the visual evidence on which those interpretations rest is gathered.

The nature of the stimuli makes these differences especially informative. Most images were engravings with extremely fine or discontinuous lines, low contrast and close integration with the natural relief of the rock; others contained incomplete animals or crowded compositions; and natural fissures and animal scratches further blurred the distinction between intentional

marks and the surface on which they were made. Under these conditions, finding the image is itself a substantial part of the task. Experts' broader and earlier coverage of archaeologically defined regions, together with their higher fixation rate and greater cumulative fixation time on those regions, indicates a more extensive search for, and examination of, the anthropogenic traces. Because these differences survived adjustment for AOI area, they are unlikely to be an artefact of region size, and because mean fixation duration did not differ, they reflect where and how often experts looked rather than slower processing of each fixation.

The pattern resembles findings from art viewing. Art historians and untrained viewers attend to different aspects of figurative and abstract paintings,[28] expertise-related differences are particularly apparent during the initial inspection of complex compositions[29] and art-history students show different gaze patterns and greater accuracy when judging style, date and medium.[14] Most closely, participants with formal art education make more fixations on task-relevant areas of art-historical images without longer mean fixations,[30] the same signature that we observed. Gaze also changes with learning: students' viewing of photographs changed after a three-month course in photographic production,[31] and historians explore challenging sequences of historical images more systematically and interpret them more accurately than novices.[32] Given that our experts had on average more than a quarter of a century of rock-art research experience, perceptual learning through prolonged engagement with this graphic tradition is a plausible source of the differences we observed, although a cross-sectional comparison cannot reconstruct how they developed.

These results extend previous experimental work on the perception of Palaeolithic art. Silva-Gago and colleagues showed which elements of Palaeolithic images attract attention and how incomplete representations are organised perceptually.[25] By introducing expertise as a variable and by using a stimulus set dominated by low-visibility engravings, we show that the

extent, timing and frequency of exploration of archaeologically relevant regions depend on who is looking. Together with evidence that cave topography guides attention towards features later used in figurative depictions,[26,27] this suggests that both the making and the reading of Palaeolithic images are shaped by an interaction between the properties of the rock and the knowledge of the observer.

The painted bison illustrates the role of graphic visibility. Despite lacking its head, the most diagnostic part of an animal depiction, it was identified by most experts and by a substantial minority of non-experts, and it accounted for most of the correct identifications made by non-experts. Its chromatic contrast and conspicuity probably made it accessible without specialised experience, in contrast to the engravings, whose fine incisions on irregular surfaces were rarely identified by non-experts. However, this stimulus also differed from the others in composition and position in the fixed presentation sequence, so the role of chromatic visibility in recognition should be tested directly, for example by comparing painted and engraved versions of comparable motifs.

The weakness of the trial-level association between gaze and identification deserves emphasis. Fixating a region is evidence of attention to it but not of recognising what it depicts;[33] experts may fixate the same lines as non-experts and still extract different information from them. The convergence of gaze and identification differences at the group level, without a one-to-one link at the level of individual images, is consistent with expertise operating jointly on search and on interpretation, the latter drawing on knowledge of graphic techniques, anatomical conventions and the relationship between marks and support.

These findings bear on a broader question about the societies that produced these images. The difficulty of distinguishing fine anthropogenic marks from natural features, and of recognising fragmentary or crowded figures, need not have been exclusive to present-day observers. Archaeological evidence for different levels of technical mastery and for

apprenticeship among Magdalenian engravers indicates that image-making involved learned skills.[6–9] It is plausible that recognising some of these images also depended on familiarity with their conventions, techniques and supports, acquired through exposure and participation in artistic practice. This remains a hypothesis rather than a conclusion of the present experiment: contemporary specialists cannot be treated as proxies for Palaeolithic viewers, and evidence of learning in production does not establish equivalent learning in perception. Nevertheless, our results identify perceptual expertise as a candidate component of Palaeolithic visual traditions, alongside the motor and technical skills needed to create images, and they offer an experimental approach to studying it.

Several limitations constrain these conclusions. The comparison was observational and the sample was small, providing adequate power only for large effects. Visual problems were more common among non-experts, although the results were robust to adjustment for them. Both groups were predominantly male and middle-aged, and the experts were drawn from a specialist research community working mainly on Iberian and Franco-Cantabrian art, so generalisation to other populations, other expert communities and other graphic traditions remains to be established.[34] Stimuli were ten two-dimensional photographs presented in a fixed order, which cannot reproduce the lighting, scale, viewpoint and three-dimensionality of direct observation in caves. Taxonomic identification captures only one dimension of recognition; responses were coded by the research team without blinding to group or an independent second coder; and the prioritisation of outcome measures was not preregistered. Future work should use preregistered designs, larger and more diverse samples, randomised stimulus orders, longitudinal tracking of trainees and immersive or in situ viewing conditions.

In sum, recognising Palaeolithic art is not an automatic, bottom-up response to figurative marks but a skilled perceptual activity. Specialists see more of the relevant evidence, sooner, and make more of it. Combining archaeological characterisation of the stimuli with eye

tracking and identification responses provides a basis for investigating how visual expertise is acquired, and for asking whether learning to see was part of how Palaeolithic graphic traditions were sustained.

Methods

Participants

The sample comprised 28 adults: 14 experts and 14 non-experts in Palaeolithic art (Table 1). Experts had extensive professional research experience in Palaeolithic rock art (mean 26.7 ± 12.4 years, range 9–47). Non-experts had little or no such experience (0.3 ± 1.1 years, range 0–4) but comparable professional experience in other research fields (27.0 ± 11.9 years; available for 13 participants). Experts were recruited from the small community of researchers who specialize in Palaeolithic art, and were defined by sustained professional research experience in this field; non-experts were researchers from other fields without training in rock art. Because the pool of available experts is very limited, the groups could not be matched or stratified by age or sex beyond the similar composition shown in Table 1. Each group comprised 12 men and 2 women, as reported by the participants. None of the authors took part as a participant. No sex- or gender-based analyses were performed because of the small number of women, which reflects the composition of the available expert pool. Visual correction was available to 23 participants and worn during the task by 22; self-reported visual problems are summarised in Table 1. No participants were excluded from the analyses. No statistical method was used to predetermine sample size. With 14 participants per group, a two-sided comparison at $\alpha = 0.05$ had 80% power to detect standardised group differences of approximately $d \geq 1.10$ in participant-level outcomes (sensitivity power analysis for a two-sample *t*-test); smaller effects could not be reliably detected.

Ethics

The study was approved by the Research Ethics Committee (Comité de Ética de la Investigación) of the University of Salamanca (registration no. 1156) and was conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent.

Stimuli

Ten unpublished photographs of Upper Palaeolithic art from four Spanish caves (La Garma Galería Inferior, Atxurra, Micolón and El Pindal) were selected from the research team's archaeological documentation (Table 2 and Supplementary Fig. 1). The set comprised one red-painted representation and nine predominantly engraved images, selected to span different degrees of perceptual complexity, from anatomically incomplete figures and very fine low-visibility engravings to panels containing multiple superimposed or closely associated representations. Several images also contained non-anthropogenic surface marks, including natural fissures, relief features and bear claw marks.

Apparatus and procedure

Participants were tested individually in sessions of approximately 30 min. Eye movements were recorded with a Tobii Pro Spectrum eye tracker (Tobii Technology) at 150 Hz. Stimuli were presented on a 1,920 × 1,080-pixel monitor with an active display area of approximately 528 × 297 mm; participants were seated centrally in front of the screen and no chin rest was used. Before each image, a five-point calibration was performed, followed by a four-point validation, and was repeated when the initial result was unsatisfactory.[21] Validation accuracy (mean angular offset across the four validation points) and precision (root-mean-square sample-to-sample deviation) were exported from Tobii Pro Lab for each of the 280 recordings and averaged within participants. Accuracy was 0.84 ± 0.42° in experts and 0.77 ± 0.72° in non-experts (Welch $t(20.9) = 0.31$, $P = 0.76$; median across recordings, 0.50°), RMS

precision was 0.30 ± 0.18° and 0.30 ± 0.32° (P = 0.99), and validation data loss was 4.5 ± 3.7% and 5.0 ± 6.2% (P = 0.83) (Table 1). The proportion of valid recording time was high in both groups (experts, 98.9 ± 1.1%; non-experts, 97.6 ± 3.5%; 0 and 3 of 140 presentations, respectively, below 80%).

In an identification-directed viewing task, the ten photographs were presented in the same fixed order to all participants, each for 20 s and preceded by a central fixation cross for 1.5 s. Participants were instructed to examine each image, try to identify the depicted motif or motifs and keep their gaze on the image throughout its presentation. After each image, they indicated whether they had recognised any motif and, if so, identified it; responses were recorded in writing and on audio for transcription and coding. After the task, participants were asked whether they had previously seen any of the images, recognised the sites or found them similar to representations they knew. All photographs were unpublished and had not been shown to the participants. Two experts nevertheless stated that they recognised some of them; because the images had never been made public, these statements were attributed to confusion with other representations, and both participants were retained in all primary analyses. Demographic and professional information was collected at the end of the session.

Gaze data were processed with the velocity-threshold identification (I-VT) filter in Tobii Pro Lab (version 25.7) with a threshold of 30° s^{-1}; the resulting fixations were analysed without reclassification.

Areas of interest

Thirty-seven AOIs (1–7 per stimulus) were delineated manually on archaeological criteria around anthropogenic graphic elements considered particularly informative for identifying the represented motifs, including anatomical features and figurative contours (Table 2). Natural rock features unrelated to the representations were not defined as AOIs, and overlaps between AOIs were avoided wherever possible. The AOIs are therefore a selection of

diagnostic elements rather than an exhaustive inventory of the visible information. The archaeological characterisation of each AOI, including cases in which the attribution of individual elements is uncertain, is given in Supplementary Table 1, and the AOI overlays are shown in Supplementary Fig. 1. Three minor polygon overlaps remained in images 08 and 09; fixations falling within overlapping regions were assigned to each region concerned. A fixation within an AOI was interpreted as evidence of gaze allocation to that region, not necessarily of recognition of the motif.

Eye-tracking measures

Three primary measures characterised regional coverage, initial orientation and fixation frequency.[19,21,22,35] The proportion of AOIs fixated was the number of AOIs receiving at least one fixation divided by the number of AOIs in the stimulus, an index of coverage and search exhaustiveness whose interpretation depends on the number, size and arrangement of AOIs.[35,36] Time to first AOI fixation was the time from the onset of the observation window to the first fixation within any AOI.[19,35,36] The AOI fixation assignment rate was the total number of fixation assignments to AOIs divided by the actual exposure time in seconds;[22,35] because a fixation in an overlapping region could be assigned to more than one AOI, this measure is a rate of assignments rather than of unique fixations.[21]

Complementary measures were mean and cumulative fixation duration, AOI visit counts and durations, time to first AOI entry and revisits.[21,35,36] The prioritisation of the three primary measures was decided after exploratory analyses and was not preregistered. The full 20-s presentation was the primary analysis window; a 17-s window excluding the first 3 s was used to assess sensitivity to initial stimulus-driven orientation. Actual recorded exposure durations were used for time-normalised measures, and observations from the two overlapping windows were treated as dependent.

Response coding

Open-ended responses were transcribed and evaluated against the established archaeological interpretation of each representation, distinguishing taxonomic identification from descriptions of anatomical or figurative features, graphic techniques or characteristics of the support. Audio recordings were transcribed and the transcriptions checked against the audio where necessary. Responses were then coded by the research team using predefined scoring rules, and taxonomic, anatomical and figurative attributions were adjudicated by the principal investigator (O.R.), a specialist in Palaeolithic art. Coding was not blind to group membership, and no independent second coding was performed, so inter-rater reliability could not be estimated. Taxonomic identification was assessed for eight stimuli (images 01, 02, 03, 06, 07, 08, 09 and 10). Images 04 and 05 were excluded because their visible graphic elements lacked sufficient diagnostic features; although image 05 is archaeologically attributed to a goat, the fragment presented does not preserve diagnostic anatomy. Responses to these two images were retained for descriptive analyses (Supplementary Methods 3). Group comparisons used only binary-scored responses (0, incorrect; 1, correct). Mixed identifications receiving partial credit and responses without an unequivocal taxonomic identification were retained separately, and the original 0/0.5/1 scoring was used in complementary analyses. For each participant, the proportion of correct identifications was calculated with binary-evaluable responses as the denominator.

Statistical analysis

Analyses accounted for repeated observations by clustering on participant and treated the ten images as fixed effects; group comparisons were observational. Generalized estimating equations (GEE)[37] were used for coverage and fixation frequency, with an independence working correlation and the bias-reduced sandwich covariance estimator of Mancl and DeRouen,[38] which corrects the downward bias of robust standard errors with few clusters.

The proportion of AOIs fixated was analysed with a quasi-binomial GEE with logit link, weighting each image equally. Time to first AOI fixation was analysed with a restricted-time approach, with a common horizon given by the shortest recorded exposure in each window; presentations without an AOI fixation before the horizon were assigned the horizon as a restricted time rather than an observed latency. Participant-level mean restricted times, weighting images equally, were compared with Welch's *t*-test. AOI fixation assignments were analysed with a negative-binomial GEE with log link and log actual exposure duration as an offset, retaining assignments to overlapping AOIs.[35] Wald tests for GEE contrasts used a *t* reference distribution with 26 degrees of freedom. Frequentist analyses were run in Python 3.12 with statsmodels 0.15.0, SciPy 1.18.1, NumPy 2.5.4 and pandas 3.0.6; Bayesian models used PyMC 6.3.2 and ArviZ 1.3.0 (Supplementary Methods 1).

Holm correction[39] controlled the family-wise error rate across the three primary comparisons in the 20-s window, with the 17-s window used for sensitivity analysis. Estimates are reported with 95% CIs and unadjusted and adjusted *P* values; all tests were two-sided. Benjamini–Hochberg correction[40] was applied within each window to separate exploratory families: exploration measures (AOI visit rate, cumulative fixation duration per exposure and mean fixation duration), four AOI-area effects and four area-adjusted group effects. Standardised effect sizes (Hedges' *g*, with small-sample correction) were computed from participant means with pooled standard deviations.

Complementary AOI-level models included image-centred log polygon area as a covariate for fixation probability, fixation count, assigned duration and time to first AOI entry. Inference used crossed participant/AOI bootstrap intervals for the first three outcomes and an image-stratified Cox model for first AOI entry; group and area effects were corrected within separate four-comparison Benjamini–Hochberg families (Supplementary Methods 2 and Supplementary Table 4).

Sensitivity analyses examined recording quality, prolonged tracking loss, age and prior familiarity with the stimuli. Two further sets of analyses addressed self-reported visual problems (adding a participant-level indicator to each model, or restricting the sample to the 22 participants who reported problems) and reported familiarity (excluding the two experts who stated that they recognised some images, or only their presentations of those images) and eye-tracker validation accuracy (adding the accuracy of the validation preceding each image as a covariate, or excluding the 31 presentations with accuracy above 1.5°); these used the same model specifications as the primary analyses (Supplementary Table 11). Excluding presentations with less than 80% valid recording time was a sensitivity restriction rather than a primary inclusion criterion.[36] Bootstrap resampling preserved repeated observations within participants, and participant-omission analyses assessed the influence of individual participants.

Taxonomic recognition was analysed as each participant's proportion of correct responses among binary-evaluable responses to the eight eligible images, with participants weighted equally. Groups were compared with two-sided studentised participant-label permutation tests (99,999 permutations), and pointwise 95% percentile CIs were obtained from 19,999 group-stratified bootstrap resamples preserving complete participant profiles. Inference was conditional on binary evaluability and the fixed image set. Exploratory gaze–recognition associations were assessed with hierarchical logistic models for AOI coverage and fixation assignments that separated within- and between-participant variation and included participant and image effects; sensitivity analyses examined recording quality, image exclusion, regularization and hypothetical outcomes for unscored responses, and a conditional logistic model provided an additional check (Supplementary Methods 1 and Supplementary Tables 9 and 10). Data distributions for the Welch test were assumed to be normal but this was not formally tested.

Use of generative AI

ChatGPT (OpenAI) was used for English-language editing and proofreading. All suggested revisions were reviewed by the authors, who take full responsibility for the content.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Pseudonymized participant- and presentation-level eye-tracking measures, AOI-level measures, response codings and the source data for all figures and tables are available in Zenodo at https://doi.org/10.5281/zenodo.24310850. Participants' ages, years of experience and verbatim interview responses are not shared publicly because, given the small size of the expert community, they could allow participants to be identified; they are available from the corresponding author on reasonable request, subject to the conditions of the ethics approval. The stimulus photographs are unpublished and are therefore not publicly released; they are available from the corresponding author on request. Additional methodological details and analyses are provided in the Supplementary Information.

Code availability

The Python analysis code and the scripts used to generate the figures are available in Zenodo at https://doi.org/10.5281/zenodo.24310850.

References

1. Conkey, M. W. New approaches in the search for meaning? A review of research in 'Paleolithic art'. *J. Field Archaeol.* **14**, 413–430 (1987).
2. Clottes, J. *What Is Paleolithic Art? Cave Paintings and the Dawn of Human Creativity* (Univ. Chicago Press, 2016).
3. Guthrie, R. D. *The Nature of Paleolithic Art* (Univ. Chicago Press, 2005).
4. Goldstone, R. L. Perceptual learning. *Annu. Rev. Psychol.* **49**, 585–612 (1998).
5. Heyes, C. *Cognitive Gadgets: The Cultural Evolution of Thinking* (Harvard Univ. Press, 2018).
6. Fritz, C. *La Gravure dans l'Art Mobilier Magdalénien: Du Geste à la Représentation* (Documents d'Archéologie Française Vol. 75, Éditions de la Maison des Sciences de l'Homme, 1999).
7. Rivero, O. Master and apprentice: evidence for learning in Palaeolithic portable art. *J. Archaeol. Sci.* **75**, 89–100 (2016).
8. Rivero, O. in *The Prehistoric Apprentice: Investigating Apprenticeship, Know-How and Expertise in Prehistoric Technologies* (ed. Klaric, L.) 275–294 (The Dolní Věstonice Studies Vol. 24, Czech Academy of Sciences, Institute of Archaeology, 2018).
9. Rivero, O. & Garate, D. Motion and gesture: analysing artistic skills in Palaeolithic art. *J. Archaeol. Method Theory* **27**, 561–584 (2020).
10. Rivero, O. et al. Experimental insights into cognition, motor skills, and artistic expertise in Paleolithic art. *Sci. Rep.* **14**, 18029 (2024).
11. Antes, J. R. & Kristjanson, A. F. Discriminating artists from nonartists by their eye-fixation patterns. *Percept. Mot. Skills* **73**, 893–894 (1991).
12. Vogt, S. & Magnussen, S. Expertise in pictorial perception: eye-movement patterns and visual memory in artists and laymen. *Perception* **36**, 91–100 (2007).

13. Francuz, P., Zaniewski, I., Augustynowicz, P., Kopiś, N. & Jankowski, T. Eye movement correlates of expertise in visual arts. *Front. Hum. Neurosci.* **12**, 87 (2018).
14. Sharvashidze, N. & Schütz, A. C. Task-dependent eye-movement patterns in viewing art. *J. Eye Mov. Res.* **13**, 12 (2020).
15. Charness, N., Reingold, E. M., Pomplun, M. & Stampe, D. M. The perceptual aspect of skilled performance in chess: evidence from eye movements. *Mem. Cognit.* **29**, 1146–1152 (2001).
16. Sheridan, H. & Reingold, E. M. Expert vs. novice differences in the detection of relevant information during a chess game: evidence from eye movements. *Front. Psychol.* **5**, 941 (2014).
17. Kundel, H. L. & Nodine, C. F. Interpreting chest radiographs without visual search. *Radiology* **116**, 527–532 (1975).
18. Reingold, E. M. & Sheridan, H. in *The Oxford Handbook of Eye Movements* (eds Liversedge, S. P., Gilchrist, I. D. & Everling, S.) 528–550 (Oxford Univ. Press, 2011).
19. Gegenfurtner, A., Lehtinen, E. & Säljö, R. Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educ. Psychol. Rev.* **23**, 523–552 (2011).
20. Henderson, J. M. Human gaze control during real-world scene perception. *Trends Cogn. Sci.* **7**, 498–504 (2003).
21. Holmqvist, K. et al. *Eye Tracking: A Comprehensive Guide to Methods and Measures* (Oxford Univ. Press, 2011).
22. Brunyé, T. T., Drew, T., Weaver, D. L. & Elmore, J. G. A review of eye tracking for understanding and improving diagnostic interpretation. *Cogn. Res. Princ. Implic.* **4**, 7 (2019).

23. Criado-Boado, F. et al. Coevolution of visual behaviour, the material world and social complexity, depicted by the eye-tracking of archaeological objects in humans. *Sci. Rep.* **9**, 3985 (2019).
24. Silva-Gago, M., Ioannidou, F., Fedato, A., Hodgson, T. & Bruner, E. Visual attention and cognitive archaeology: an eye-tracking study of Palaeolithic stone tools. *Perception* **51**, 3–24 (2022).
25. Silva-Gago, M., García-Diez, M., Bruner, E., Martínez, L. M. & Criado-Boado, F. Figure recognition and visual attention patterns behind the observation of Palaeolithic art. *J. Cult. Cogn. Sci.* **9**, 225–241 (2025).
26. Wisher, I., Pettitt, P. & Kentridge, R. The deep past in the virtual present: developing an interdisciplinary approach towards understanding the psychological foundations of Palaeolithic cave art. *Sci. Rep.* **13**, 19009 (2023).
27. Wisher, I., Pettitt, P. & Kentridge, R. Conversations with caves: the role of pareidolia in the Upper Palaeolithic figurative art of Las Monedas and La Pasiega (Cantabria, Spain). *Camb. Archaeol. J.* **34**, 315–338 (2024).
28. Pihko, E. et al. Experiencing art: the influence of expertise and painting abstraction level. *Front. Hum. Neurosci.* **5**, 94 (2011).
29. Rosenberg, R. & Klein, C. in *Art, Aesthetics, and the Brain* (eds Huston, J. P., Nadal, M., Mora, F., Agnati, L. F. & Cela-Conde, C. J.) 79–108 (Oxford Univ. Press, 2015).
30. Korkmaz, A., Gülsecen, S. & Mihaylov, G. Visual evaluation strategies in art image viewing: an eye-tracking comparison of art-educated and non-art participants. *J. Eye Mov. Res.* **19**, 14 (2026).
31. Ishiguro, C., Yokosawa, K. & Okada, T. Eye movements during art appreciation by students taking a photo creation course. *Front. Psychol.* **7**, 1074 (2016).

32. Puurtinen, M., Kaakinen, J. K. & Jarodzka, H. Visual expertise in history: historians' and novices' visual processing and interpretation of historical images. *J. Expertise* **9** (2026).
33. Just, M. A. & Carpenter, P. A. A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* **87**, 329–354 (1980).
34. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
35. Hooge, I. T. C. et al. The fundamentals of eye tracking part 6: working with areas of interest. *Behav. Res. Methods* **58**, 65 (2026).
36. Godwin, H. J., Hout, M. C., Alexdóttir, K. J., Walenchok, S. C. & Barnhart, A. S. Avoiding potential pitfalls in visual search and eye-movement experiments: a tutorial review. *Atten. Percept. Psychophys.* **83**, 2753–2783 (2021).
37. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).
38. Mancl, L. A. & DeRouen, T. A. A covariance estimator for GEE with improved small-sample properties. *Biometrics* **57**, 126–134 (2001).
39. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
40. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Acknowledgements

This research was conducted within the ArtMindHuman project (PID2021-125166OB-I00).

Funding

This work was supported by the project 'Creation and perception in Anatomically Modern Humans: analysis of the biological, cognitive and social skills linked to the production of Palaeolithic art (ArtMindHuman)', PID2021-125166OB-I00, funded by MCIN/AEI/10.13039/501100011033 and by FEDER, A way of making Europe (PI: O.R.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

O.R. conceived the study, conducted the investigation and formal analysis, wrote the original draft, prepared the visualizations, supervised the work and acquired funding. M.S.B. conceived the study, conducted the investigation, curated the data, reviewed and edited the manuscript and supervised the work. M.S. developed the methodology and software, curated the data and reviewed and edited the manuscript. A.M.M.-P. developed the methodology, curated the data, provided resources and prepared the visualizations. A.A.-M., J.E., X.E.-C. and M.G.-B. developed the methodology, curated the data, provided resources and reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper.

Correspondence and requests for materials should be addressed to Olivia Rivero.

Figure legends

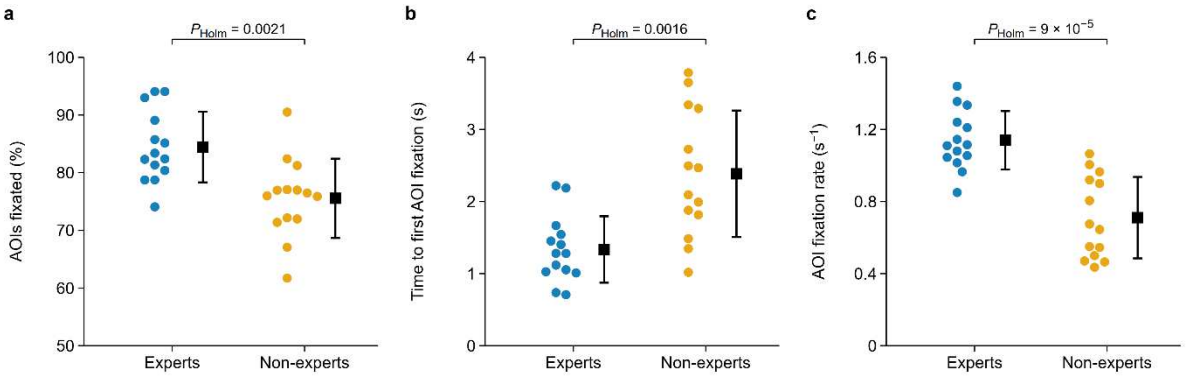


Fig. 1 | Experts explore archaeologically defined areas of interest more extensively, sooner and more often. a, Proportion of AOIs fixated. **b**, Restricted time to first AOI fixation; presentations without an AOI fixation were assigned the common observation horizon, so values should not be read as uncensored latencies. **c**, AOI fixation assignment rate (assignments per second of actual exposure, including multiple assignments to overlapping AOIs). Points show each participant's mean across ten equally weighted images (n = 14 experts and 14 non-experts); black markers and whiskers show group means ± s.d. Group contrasts are from a quasi-binomial GEE (a), Welch's t-test (b) and a negative-binomial GEE (c), all two-sided and Holm-adjusted (P_{Holm} = 0.0021, 0.0016 and 9 × 10^{-5}, respectively); full statistics are given in Table 3.

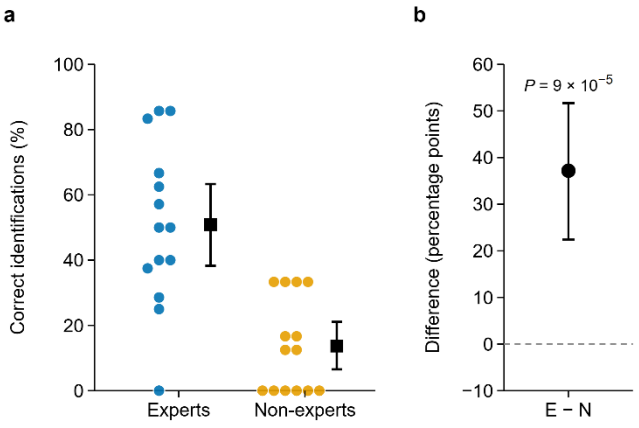


Fig. 2 | Experts identify the depicted animals more accurately than non-experts. a, Proportion of correct taxonomic identifications among binary-evaluable responses to the

eight eligible stimuli (01, 02, 03, 06, 07, 08, 09 and 10); points show individual participants (n = 14 per group) and black markers and whiskers show group means and pointwise 95% bootstrap CIs. **b**, Expert minus non-expert difference in mean accuracy (37.2 percentage points) with its 95% bootstrap CI (19,999 group-stratified resamples). The difference was tested with a two-sided studentised participant-label permutation test (99,999 permutations; $P = 9 \times 10^{-5}$). Estimates weight participants equally and are conditional on binary evaluability and the fixed set of eight images.

Tables

Table 1 | Demographic characteristics, professional experience, visual status and eye-tracker validation of participants

Variable	Experts (n = 14)	Non-experts (n = 14)	Total (N = 28)
Demographic characteristics			
Age, years (mean ± s.d.)	54.1 ± 14.6	53.6 ± 10.6	53.9 ± 12.5
Age range, years	32–82	38–71	32–82
Male, n (%)	12 (85.7)	12 (85.7)	24 (85.7)
Female, n (%)	2 (14.3)	2 (14.3)	4 (14.3)
Professional experience			
Rock-art experience, years (mean ± s.d.)	26.7 ± 12.4	0.3 ± 1.1	—
Rock-art experience, range (years)	9–47	0–4	—
Experience in other research fields, years (mean ± s.d.)	—	27.0 ± 11.9*	—
Visual characteristics			
Visual correction available, n (%)	11 (78.6)	12 (85.7)	23 (82.1)
Visual correction worn during the task, n (%)	10 (71.4)	12 (85.7)	22 (78.6)
Self-reported visual problems, n (%)	9 (64.3)	13 (92.9)	22 (78.6)
Eye-tracker validation			
Accuracy, degrees (mean ± s.d.)	0.84 ± 0.42	0.77 ± 0.72	0.81 ± 0.58
Precision (RMS), degrees (mean ± s.d.)	0.30 ± 0.18	0.30 ± 0.32	0.30 ± 0.26
Validation data loss, % (mean ± s.d.)	4.5 ± 3.7	5.0 ± 6.2	4.7 ± 5.0

Rock-art experience refers to professional or research experience in the study of rock art. *Available for 13 non-expert participants. Validation values are participant means over the ten pre-image validations (four points each); group differences were not significant (Welch tests, all $P \geq 0.76$). A dash indicates that a pooled value is not reported or that the measure is not applicable to that group.

Table 2 | Archaeological characteristics of the ten stimuli

Image	Site	Archaeological content	Graphic characteristics	AOIs (n)
01	La Garma	Headless bison facing left	Red painting incorporating natural rock relief	1
02	Atxurra	Multiple animals, including two bison, a horse and a goat	Engraved figures and additional graphic elements	7
03	La Garma	Bison facing right	Engraving with associated remains of red pigment	3
04	La Garma	Fragmentary animal representations and graphic elements, including aurochs/caprine and taxonomically indeterminate components	Engraved elements with uncertain anatomical and figurative relationships	5
05	La Garma	Fragment attributed to a goat, facing right	Engraved animal representation	3

06	La Garma	Horse facing left	Engraving associated with a red pigment stain	4
07	La Garma	Animal interpreted as a wolf or bear, facing left	Engraving and traces of red pigment	4
08	Micolón	Goat facing left and a separate vulva representation	Engraved figurative and symbolic elements	4
09	La Garma	Animal head interpreted as a goat or hind, facing right	Engraving with ambiguous anatomical features	4
10	El Pindal	Bison facing right and downwards	Engraving incorporating natural rock relief, with traces of red pigment	2
Total				**37**

Descriptions refer to the broader content of each stimulus and are not exhaustive inventories of the graphic elements within the AOIs, which were characterised separately (Supplementary Table 1). In image 07, red pigment traces lie outside the AOIs; in image 08, the vulva has no corresponding AOI; in image 10, red pigment is present within AOI 10_1 but not AOI 10_2.

Table 3 | Primary eye-tracking outcomes in the complete (20-s) observation window

Outcome	Experts	Non-experts	Effect	95% CI	*P*	P_{Holm}
Proportion of AOIs fixated (%)	84.5 ± 6.1	75.6 ± 6.9	OR 1.88	1.29, 2.75	0.0021	0.0021
Restricted time to first AOI fixation (s)	1.33 ± 0.46	2.38 ± 0.88	Δ −1.05	−1.60, −0.50	7.9×10^{-4}	0.0016
AOI fixation assignment rate (s^{-1})	1.14 ± 0.16	0.71 ± 0.23	RR 1.64	1.34, 2.00	3.0×10^{-5}	9.0×10^{-5}

Group values are means ± s.d. of participant means over ten equally weighted images (n = 14 per group). Effects compare experts with non-experts: OR, coverage odds ratio (quasi-binomial GEE); Δ, difference in restricted time (Welch's t-test); RR, assignment-rate ratio (negative-binomial GEE with log actual exposure offset). GEE models cluster on participant and adjust for image. All tests are two-sided; Holm adjustment was applied across the three contrasts. Restricted non-events retain the common horizon and should not be interpreted as latencies.

Table 4 | Response categories in the eight-image taxonomic assessment

Category	Experts	Non-experts
Eligible presentations	112	112
Correct taxonomic identifications (score 1)	46	10
Incorrect identifications (score 0)	44	72
Binary-evaluable subtotal	**90**	**82**
Mixed alternatives with partial credit	1	2
Broadly compatible, unscored	1	0
Unscored for strict taxonomic identification	20	28

Only scores of 0 and 1 entered the binary analysis; the subtotal is not an additional category. Unscored responses could include recognition of anatomical, figurative or technical features and were not necessarily identification failures. Images 04 and 05 were excluded from taxonomic scoring.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf