



**HAL**  
open science

# Critical information thresholds underlying generic and familiar face categorisation at the same face encounter

Genevieve L Quek, Bruno Rossion, Joan Liu-Shuang

► **To cite this version:**

Genevieve L Quek, Bruno Rossion, Joan Liu-Shuang. Critical information thresholds underlying generic and familiar face categorisation at the same face encounter. *NeuroImage*, 2021, 243, pp.118481. 10.1016/j.neuroimage.2021.118481 . hal-03437754

**HAL Id: hal-03437754**

**<https://hal.science/hal-03437754>**

Submitted on 20 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Critical information thresholds underlying generic and familiar face categorisation at the same face encounter

Genevieve L. Quek<sup>a,b,\*</sup>, Bruno Rossion<sup>c,d,e</sup>, Joan Liu-Shuang<sup>c</sup>

<sup>a</sup> Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

<sup>b</sup> School of Psychology, The University of Sydney, Sydney, Australia

<sup>c</sup> Institute of Research in Psychology (IPSY), University of Louvain, Louvain, Belgium

<sup>d</sup> Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

<sup>e</sup> Université de Lorraine, CHRU-Nancy, Service de Neurologie, Lorraine F-54000, France

## ARTICLE INFO

### Keywords:

Face recognition  
Face categorisation  
Spatial frequency  
Time  
EEG  
Frequency-tagging  
Familiar face recognition  
Spatial filtering  
Face processing

## ABSTRACT

Seeing a face in the real world provokes a host of automatic categorisations related to sex, emotion, identity, and more. Such individual facets of human face recognition have been extensively examined using overt categorisation judgements, yet their relative informational dependencies during the same face encounter are comparatively unknown. Here we used EEG to assess how increasing access to sensory input governs two ecologically relevant brain functions elicited by seeing a face: Distinguishing faces and nonfaces, and recognising people we know. Observers viewed a large set of natural images that progressively increased in either image duration (experiment 1) or spatial frequency content (experiment 2). We show that in the absence of an explicit categorisation task, the human brain requires less sensory input to categorise a stimulus as a face than it does to recognise whether that face is familiar. Moreover, where sensory thresholds for distinguishing faces/nonfaces were remarkably consistent across observers, there was high inter-individual variability in the lower informational bound for familiar face recognition, underscoring the neurofunctional distinction between these categorisation functions. By *i*) indexing a form of face recognition that goes beyond simple low-level differences between categories, and *ii*) tapping multiple recognition functions elicited by the same face encounters, the information minima we report bear high relevance to real-world face encounters, where the same stimulus is categorised along multiple dimensions at once. Thus, our finding of lower informational requirements for generic vs. familiar face recognition constitutes some of the strongest evidence to date for the intuitive notion that sensory input demands should be lower for recognising face category than face identity.

## 1. Introduction

Faces hold exceptional status in the human brain, conveying a great deal of meaningful social information that is recognised nearly effortlessly by neurotypical adults. Yet the relative ease of face recognition belies the complex and multifaceted nature of this key human faculty, which in fact comprises a heterogeneous set of processes that culminate in distinct categorisations of a face's category (i.e., recognising a face as a face), its sex, emotion, familiarity, identity, and beyond (Young and Bruce, 2011). Remarkably, these various high-level recognition functions appear to be evoked automatically at every face encounter, such that to see a face is to almost instantaneously 'recognise' it in a multitude of ways.

In contrast to the multifaceted nature of face recognition/categorisation<sup>1</sup> in the real world, there is a long tradition in face research of studying these various aspects in isolation, by tasking observers with categorising face images along a single dimension (e.g., familiarity) at a time (Gobbini and Haxby, 2007; Hill et al., 1995; Ekman, 1993; Rhodes et al., 1989). Within this modular framework, empirically relating the different face recognition functions to one another necessitates comparisons across distinct observer tasks and face encounters (e.g., contrasting response times, or RTs, for recognising faces amongst objects with RTs for recognising familiar faces amidst unfamiliar ones) (Crouzet et al., 2010; Besson et al., 2017; Barragan-Jason et al., 2013; Barragan-Jason et al., 2012; Thorpe et al.,

<sup>1</sup> The term "recognition" here refers to the production of a selective (i.e., discriminant) response to a given sensory input, a response that can be reproduced (i.e., generalized) across variable viewing conditions. In this sense, recognition is essentially a categorisation function (see Rossion & Retter, 2020); the two terms are used interchangeably here.

\* Corresponding author.

E-mail address: [genevieve.lauren.quek@gmail.com](mailto:genevieve.lauren.quek@gmail.com) (G.L. Quek).

1996; Grill-Spector and Kanwisher, 2005; Schyns et al., 2002). Since each categorisation task is necessarily associated with its own specific goals, stimulus images, target-distractor overlap, observer strategies, and even participant groups, general experimental factors hamper the identification of distinctive characteristics and informational requirements associated with the different forms of face categorisation. Thus, the cross-task comparative approach can offer limited insight into how the brain achieves the manifold automatic face categorisations that arise naturally when we see a face in the real world.

Recently, a new wave of research has emerged aimed at understanding how the brain extracts information along different face dimensions at the same face encounter (Nemrodov et al., 2016; Ambrus et al., 2019; Dobs et al., 2019; Ghuman et al., 2014). In contrast to second-order comparisons of face recognition functions (which are drawn across different tasks/face encounters), this approach investigates the different classes of categorisation reflected in the exact same neural response elicited by seeing a face, typically by applying multivariate pattern analysis (MVPA) techniques to high temporal resolution electro/magnetoencephalographic data (EEG, MEG) (Carlson et al., 2013). For example, a recent MEG study used this approach to examine the temporal unfolding of familiarity, sex, and age categorisations of the same faces. Contrasting the time course of decoding associated with each dimension as reflected in the same neural response showed that the age and sex of a face were categorised earlier than its identity (Dobs et al., 2019). Where this multivariate approach focuses near-exclusively on contrasting the relative onset and duration of categorical representations that follow a face presentation, cognitive processes can differ not only in their temporal unfolding, but also in the amount of sensory evidence they require to proceed (VanRullen, 2011). To date, this latter possibility has received little explicit exploration in the context of multifaceted face categorisation. Thus, where there have been many modular investigations of how evidence accumulates to support performance on explicit categorisation tasks (e.g., manipulations of image duration (Grill-Spector et al., 2000; Tanskanen et al., 2007; Or and Wilson, 2010; Näsänen et al., 2006), spatial resolution (Quek et al., 2018; Ramon et al., 2015), visibility (Ales et al., 2012), and so on), the relative informational dependencies of different forms of categorisation that arise during the same face encounter remain comparatively unknown.

In the current study, we aimed to characterise how increasing access to sensory face input influences two ecologically relevant recognition functions evoked at every face encounter: **generic face categorisation** (i.e., recognising that a visual stimulus is a face, as opposed to another type of object) and **familiar face categorisation** (i.e., recognising that a face is one you have encountered before). We took care to avoid tasking participants with an explicit categorisation judgement (which would evoke task-specific strategies), instead using high-density EEG to track implicit neural measures of the two brain functions of interest as observers viewed a large number of widely variable, unsegmented images from various natural categories. We constrained sensory evidence by parametrically varying either image viewing time (i.e., stimulus presentation duration, Expt. 1), or spatial frequency content (i.e., image resolution, Expt. 2) (Quek et al., 2018). By measuring the effect of increasing temporal/spatial exposure on implicit neural indices of generic and familiar categorisation evoked by the same face encounters, we identified the minimal amount of sensory input<sup>2</sup> capable of driving each function as it occurs in the real world, where unexpectedly encountering a face automatically provokes multiple categorisations at once.

<sup>2</sup> Our focus on *informational* dependencies should not imply that face categorisation is governed solely by physical characteristics of visual input. Indeed, to any naïve system (human or artificial), there is no stimulus-level information that distinguishes a 'familiar' face from an 'unfamiliar' one. Here we assume the integration of sensory information with other semantic/memory processes to be inherent to both generic and familiar face recognition.

We compared these spatiotemporal thresholds at the individual participant level, with the goal of explicitly relating the informational requirements of generic and familiar face categorisation within the same observer. Since a tacit assumption underlying many influential theories of human face recognition is that recognising that a stimulus is a face necessarily precedes recognising that face as familiar (Burton et al., 1999; Bruce and Young, 1986), the strong prediction here is that sensory input demands should be lower for generic face recognition. On this possibility, individual observer thresholds would be similarly distributed regardless of face familiarity, yet reliably offset from familiar face recognition thresholds. On the other hand, however, while this *basic-before-subordinate* directionality is certainly intuitive, it remains untested at the level of individual face encounters. By providing this granularity, the present study stands to reveal whether the sensory input diagnostic for an observer to recognise a given visual stimulus as a face also enables them to recognise whether that face is familiar. If generic and familiar face recognition are indeed tightly coupled within each observer, generic face categorisation thresholds will vary as a function of face familiarity, producing familiar and unfamiliar distributions that are at least partially non-overlapping.

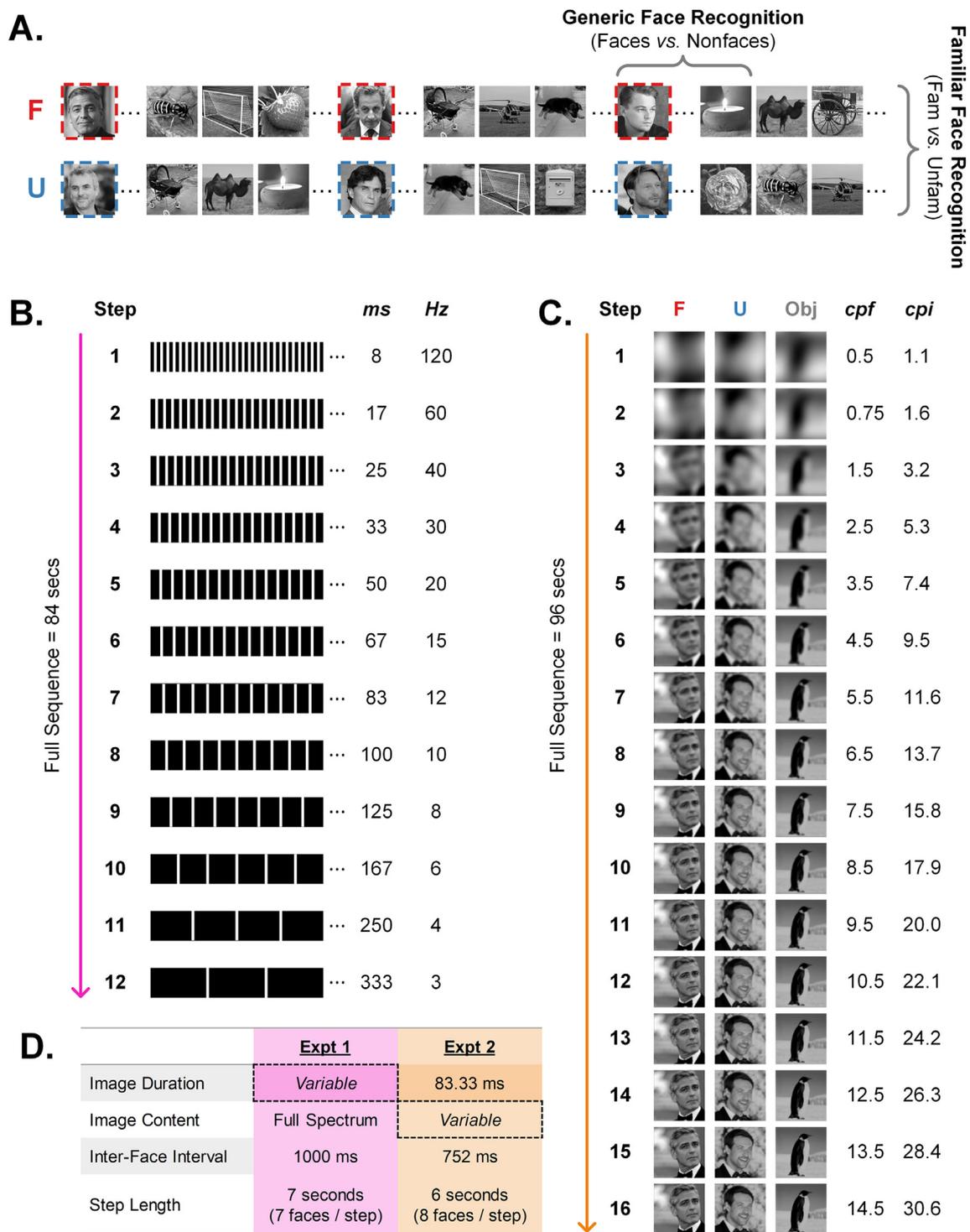
## 2. Materials and methods

### 2.1. Participants

We tested independent samples of 25 participants each in Expts. 1 and 2. All gave written informed consent in accordance with UCLouvain BioEthics committee guidelines and were monetarily compensated. All were right-handed, with normal or corrected-to-normal vision, and did not report any psychiatric or neurological history. All were of French-speaking Belgian background or verified to be familiar with its culture. For each experiment separately, we excluded participants with poor behavioural performance (e.g., task performance < 2.5 SD from group average, see below) or those who blinked excessively during EEG recording (i.e., mean blinks / second > 2.5 SD from group average). The group-level accuracy and blink statistics were calculated on the full sample and both exclusion criteria were applied in parallel. The final sample consisted of 22 participants in Expt. 1 (13 females, mean age = 22.27 yrs  $\pm$  1.98) and 21 participants in Expt. 2 (12 females, mean age = 21.81 yrs  $\pm$  1.65).

### 2.2. Protocol and design

To quantify the amount of sensory input required for successful generic and familiar face recognition, we adapted an EEG frequency-tagging paradigm in which observers view long lasting sequences of rapidly presented images belonging to many different natural categories (e.g., plants, animals, buildings, vehicles, etc.) with faces embedded at strict periodic intervals (Fig. 1A). Stimulating the visual system in this way is known to yield separable electrophysiological indices of *i*) general visual processing (i.e., processing common to faces and objects, measurable at the frequency of image presentation), and *ii*) face-selective visual processing (i.e., the differential response to faces vs. objects, measurable at the frequency of face presentation) that are identifiable at the level of individual observers (Quek et al., 2018; Rossion et al., 2015; G.L. Quek et al., 2018; Retter and Rossion, 2016). For our purposes here, the face-selective visual response provides an index of generic face recognition (i.e., recognising a face as a face), as it can only arise if the neural response evoked by faces in the sequence is both consistently similar across different face exemplars, yet consistently different to the responses evoked by nonface images. We obtained a corresponding index of familiar face recognition by computing the difference between face-selective responses elicited by sequences containing either highly *Familiar* (F) or *Unfamiliar* (U) faces (Fig. 1A; Figure S1A). Importantly, face exemplars varied widely in



**Fig. 1.** Design overview. **A.** The basic frequency-tagging paradigm consists of a rapid stream of various object categories interspersed with Familiar (F) or Unfamiliar (U) faces at strictly periodic intervals (Rossion et al., 2015). **B.** Sequences in Expt. 1 contained full spectrum images whose duration increased every 7 seconds. The inter-face interval was always 1000 ms for a total of 84 face presentations/sequence. **C.** In Expt. 2, image duration was fixed at 83.33 ms, with image resolution increasing every 6 seconds. The inter-face interval was always 752 ms, for a total of 127 faces/sequence (cpf = cycles per median face width; cpi = cycles per image). **D.** Summary of the differing design properties in Expts. 1 & 2.

pose, lighting, expression, background, etc. (Figure S1B), such that specific low-level visual features could not systematically occur at the face presentation frequency. This ensures that neither recognition index (i.e., the face-selective response itself and the familiarity effect obtained via subtraction) can be driven by image-level differences between categories (Rossion et al., 2015; Gao et al., 2018). In two sep-

arate groups of observers, we tracked both indices as a function of parametric increase in either image presentation duration (Expt. 1, Fig. 1B) or spatial frequency content (Expt. 2; Fig. 1C), in both cases identifying the minimal informational input (i.e., threshold) required for *i*) successful generic face recognition, and *ii*) successful familiar face recognition.

### 2.2.1. Expt 1: Increasing image duration

Sequences contained full spectrum face and object images whose presentation duration began at 8.33 ms (i.e., 120 Hz), increasing every 7 s across 12 steps to reach 333.33 ms (i.e., 3 Hz). A face appeared every 1000 ms (i.e., 1 Hz face presentation rate), such that the number of intervening objects between faces decreased with each duration step. The first/last sequence steps each lasted an extra second wherein the global image contrast gradually ramped up/down. These fade in/out periods were excluded from response quantification. Participants had no explicit task during the sequence itself, and were simply told to pay close attention to the face and object images as they appeared one by one. At the start of the experiment, we ran a single *Familiar* practice sequence to explain the experiment to participants, and to introduce the basic identity recognition task that followed each sequence (see below). The experiment proper consisted of 24 sequences (12 *Familiar*, 12 *Unfamiliar*), presented in pseudo-random order across two counterbalanced blocks.<sup>3</sup>

### 2.2.2. Expt 2: Increasing image content

Each 96 second sequence contained face/nonface images with a fixed image duration of 83.33 ms (i.e., 12 Hz, a rate known to elicit face categorisation responses at the individual observer level (Retter and Rossion, 2016)). Spatial frequency (SF) content of the images parametrically increased every 6 s such that initially blurry images progressively sharpened over the course of the full sequence (Fig. 1C; see our previous work for a movie of a similar sequence of parametrically increasing SF content (Quek et al., 2018)). A 6 second fade-in period preceded the initial step (excluded from analysis). Unlike Expt. 1, here there were always exactly 8 objects, or 752 ms, between faces (i.e., face presentation rate = 1.33 Hz). As in Expt. 1, participants had no explicit task during the sequence itself. There were 24 sequences (12 *Familiar* and 12 *Unfamiliar*) that appeared in a pseudo-random order across two counterbalanced blocks; participants saw one *Familiar* practice sequence prior to the experiment proper.

### 2.3. Post-sequence identity recognition task

To encourage participants to attend to the faces in the image sequences, both experiments included a simple 2AFC identity recognition decision after each sequence (totalling 24 2AFC responses across the full experiment). Here participants saw one 3 second display containing a probe face and a foil face, whose position was counterbalanced across sequences (see Figure S1C). Within the display duration, participants had to indicate using the arrow keys whether the left or right person had appeared during the preceding sequence. Both the probe and the foil held the same familiarity status as the preceding sequence, such that participants could not respond based on a greater overall sense of familiarity. Additionally, we emphasised that the task pertained to the *identities* shown, not the specific images (which were always novel). Note that this task was not designed to be a sensitive measure of overt face recognition; we included it simply to encourage observers to pay attention to the faces in the sequence. Nevertheless, performance on this simple 2AFC task validated our familiarity manipulation, showing that probe identification rates were significantly higher for familiar vs. unfamiliar identities in both Expt. 1 (Fam =  $0.98 \pm 0.01$ ; Unfam =  $0.79 \pm 0.03$ ;  $t(21) = 6.87$ ,  $p < .0001$ ) and Expt. 2 (Fam =  $0.98 \pm 0.60$ ; Unfam =  $0.60 \pm 0.03$ ,  $t(20) = 11.39$ ,  $p < .0001$ ). A corresponding effect was found for response times in Expt. 1 (Fam =  $1.49 \pm 0.10$  secs; Unfam =  $2.01 \pm 0.10$  secs;  $t(21) = -7.05$ ,  $p <$

<sup>3</sup> Prior to the practice trial, both experiments contained an additional four sequences in which there was no parametric variation (i.e., observers saw full spectrum images presented at 12 Hz with a fixed face frequency of 1.5 Hz). These additional sequences (containing a different set of face images) pertained to a separate investigation and are not reported further here.

.0001) and Expt. 2 (Fam =  $1.52 \pm 0.11$  secs; Unfam =  $2.21 \pm 0.12$  secs,  $t(20) = -5.09$ ,  $p < .0001$ ).

### 2.4. Stimuli and display

Stimuli were 200 greyscale images of various nonface visual categories (e.g., animals, plants, structures, vehicles, objects, etc., Rossion et al., 2015; Retter and Rossion, 2016) and 240 greyscale images of faces. All image subjects were unsegmented, i.e., embedded in their natural backgrounds (see Fig. 1A). Specific face images were selected based on a separate stimulus pre-screening experiment run prior to Expts 1 & 2 as follows: We paired each of 13 celebrities considered highly recognisable to a French-speaking Belgian population with a comparatively unknown foreign celebrity of similar age and appearance. For each of the resulting 26 identities, we sourced 30 individual exemplar images online (i.e., 780 images, each  $256 \times 256$  pixels), ensuring the sets varied widely in terms of background, lighting, facial expression, age, pose, etc. (Figure S1B). We used a browser-based experimental platform (<https://www.testable.org>) to present 50 French-speaking Belgian participants aged between 18 and 42 years (18 males, mean age =  $23 \text{ years} \pm 3.62$ ) with a 390 trial 2AFC naming task containing these images (i.e., each participant saw half the exemplar images for each identity). Each trial contained a brief fixation cross, followed by a central target face image (e.g., image of George Clooney) with a target name and lure name below (e.g., “GEORGE CLOONEY” — “BRAD PITT”, target name position randomised across trials). The face image disappeared after 500 ms, while the names remained onscreen. Participants had a maximum of 3 s from display onset to click on the name that matched the face image on that trial. Both names were always drawn from the same Familiarity category (e.g., a famous face always appeared with two famous names), to prevent participants from responding based on a sense of familiarity alone. The five identities with the highest recognition rates were Danny Boon, Nicolas Sarkozy, George Clooney, Leonardo DiCaprio, and Emmanuel Macron ( $M = 96.55\%$ ,  $SD = 0.91$ ). Recognition rates for their corresponding matched identities (respectively, Alfonso Cuarón, Thomas Kretschmann, Kirill Safonov, Najib Amhali, Humberto Zurita) were significantly lower,  $t(4) = 27.68$ ,  $p < .0001$ , trending close to chance ( $M = 60.51\%$ ,  $SD = 3.47$ ) (see Figure S2 in supplemental material). Observers were also significantly slower to identify the *Unfamiliar* identities ( $M = 1488$  ms,  $SD = 62$  ms) compared to the *Familiar* ones ( $M = 1216$  ms,  $SD = 30$  ms),  $t(4) = -15.88$ ,  $p < .0001$ . Since the familiar identities in this group were both *i)* highly recognisable, and *ii)* very distinct from their unfamiliar counterparts, we selected these 10 face identities to use in the main experiments (see Figure S1A), and kept the remaining identities aside to use as foil faces during the post-sequence 2AFC identity recognition task. We narrowed the image set for each of these selected 10 identities to a representative 24 exemplars, taking care to exclude exemplars whose individual recognition rate fell below the uniformly high group rate. We divided the resulting 120 *Familiar* and 120 *Unfamiliar* faces into three subsets of 40, each containing 8 exemplars per identity. During the two experiments, each of the three face subsets served as the face stimuli for four sequences (two in each testing block).

All object and face stimuli were sized  $256 \times 256$  pixels and equalised in terms of mean luminance and contrast; this finalised image set comprised the stimuli used in Expt. 1. For Expt. 2, we took the additional step of generating spatially filtered versions of all images at 16 increasing low-pass filter cut-off values ranging from 1.05 to 30.6 cycles per image (cpi), corresponding to 0.5 to 14.5 cycles per face (cpf; estimated using the median width of faces within each image) (see Fig. 1C). Notably, this range encompasses SF bands previously implicated in processing both face category (Quek et al., 2018) and face identity (i.e., ~8–12 cpf) (Näsänen, 1999). The filter cut-off values were spaced to maximise resolution in cpf with the goal of identifying more precise thresholds at the individual subject level. We used custom Java software to display the finalised stimuli on a 120 Hz BenQ LED monitor with  $1920 \times 1080$

resolution in a dimly lit room. The viewing distance was 50 cm, such that images spanned a visual angle of  $\sim 10^\circ$ . All stimuli and instructions appeared on a grey uniform background; during the sequence presentation a small central fixation cross remained overlaid on the images.

### 2.5. Post-sequence 2AFC task

Stimuli for the 2AFC post-sequence identity recognition task were 24 unique probe + foil combinations (12 *Familiar* and 12 *Unfamiliar*, see Figure S1C in supplemental material). Each of the experimental 10 identities appeared as the probe at least twice, and was always paired with a novel, never-before-seen identity as the foil. Note that the actual images used for the experimental and 2AFC task were completely distinct, i.e., the probe was always a completely novel image of a previously-seen identity.

### 2.6. EEG acquisition and analysis

We used a BioSemi ActiveTwo system with standard 10–20 system electrode locations and additional intermediate positions to acquire 128-channel scalp EEG (512 Hz sample rate). We monitored eye movements using electrodes at the outer canthi of both eyes, and above/below the right eye. Individual electrode offsets were held below  $\pm 50 \mu\text{V}$ . During testing, digital triggers were sent via a parallel port to mark the start of each stimulation sequence and all behavioural responses. The experimenter manually initiated each sequence's recording after the EEG trace showed no muscular/ocular artefact for at least 5 s. To maintain observer comfort, we encouraged participants to rest their eyes between trials, initiating the next sequence once after they indicated they were ready to proceed. We imposed a longer rest-break after every six sequences.

### 2.7. EEG preprocessing

We analysed EEG data offline using Letswave5 (<https://www.letswave.org/>) running on MATLAB R2012b (MathWorks, MA, United States). We realigned the continuous EEG data to remove abrupt signal offsets that resulted from pausing the recording, then de-trended and removed the DC component from the data. Next we applied a band-pass filter with cut-offs at 0.05 Hz and 125 Hz (4th order zero-phase Butterworth filter), followed by a multi-notch filter remove electrical noise carried at 50, 100, and 150 Hz (FFT filter, width = 0.5). Data was downsampled to 256 Hz for easier handling and storage, and segmented according to stimulation sequences, with two extra seconds before and after the sequence (Expt. 1 = 86 s; Expt. 2 = 102 s). For each participant, we used independent component analysis (ICA) with a square mixing matrix to remove a single component corresponding to eyeblinks (identified through visual inspection of component waveforms and topographical distributions). We interpolated artefact-ridden channels with the average of the 3 neighbouring channels (less than 5% of channels were corrected for each observer) and re-referenced the cleaned data to the average of all 128 scalp channels. We cropped the preprocessed data to exclude the fade-in and fade-out periods for each sequence (final epoch lengths were 84 s in Expt. 1; 96 s in Expt. 2). We then averaged each participant's *Familiar* and *Unfamiliar* segments separately, before chunking these conditional averages into separate epochs for each duration/SF step (Expt. 1 =  $12 \times 7$  s epochs, Expt. 2 =  $16 \times 6$  s epochs). Finally, we applied a Fast Fourier Transformation (FFT) to each epoch to extract 0–128 Hz frequency amplitude spectra for each combination of participant, condition, and step (frequency resolution in Expt. 1 = 0.14 Hz, in Expt. 2 = 0.16 Hz). We considered conditional group means within a predefined bilateral occipitotemporal (OT) region-of-interest (ROI), averaging across electrode sites previously shown to be involved in both generic and familiar face categorisation (P8, PO8, P10, PO10, PO12, P7, PO7, P9, PO9, PO11, see insets on Figs. 2& 6). All group-level and individual response profiles are noise-corrected amplitudes, obtained by

subtracting the mean noise value from the signal, separately for each condition and each step.

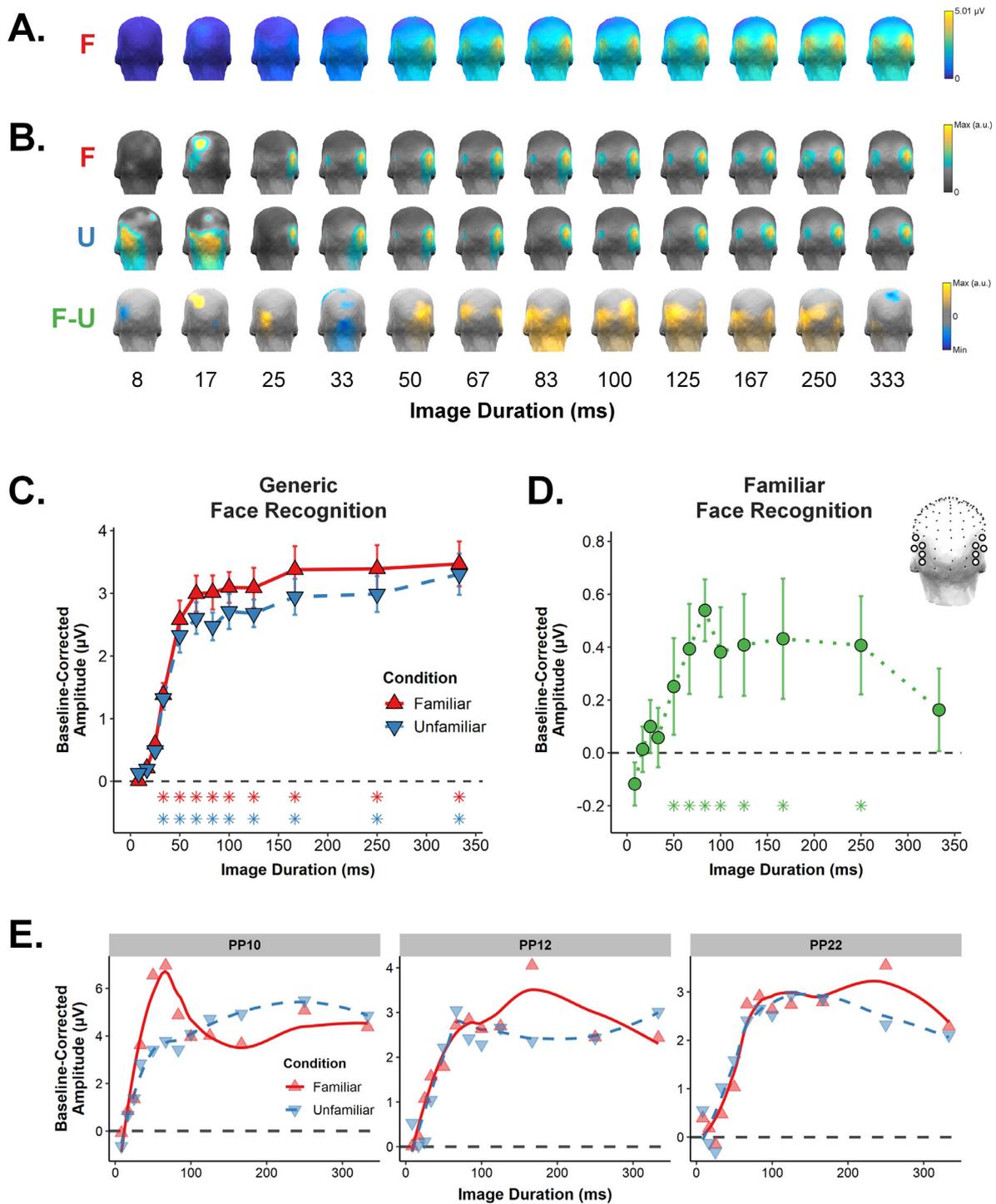
### 2.8. Threshold analysis

We identified recognition thresholds at both the group and individual level by comparing the magnitude of the relevant signal against an empirical noise distribution generated using a bootstrap procedure. For generic face recognition, the signal estimate for each step/condition combination comprised the summed amplitude values on all face frequency harmonics up to 30 Hz (Expt. 1: 30 harmonics = 1–30 Hz; Expt. 2: 22 harmonics = 1.33–29.26 Hz), excluding harmonics of the image presentation frequency. To compute a noise estimate, we summed the amplitude values on an identical number of randomly selected noncritical frequencies (i.e., excluding the face presentation and image presentation frequencies). To take account of the 1/f profile of the EEG spectrum, the frequency range for the noise calculation included 3 extra frequency bins either side of the face categorisation response range (Expt. 1 = 0.58–30.42 Hz; Expt. 2 = 0.85–29.74 Hz). For each combination of condition/step, we generated a noise distribution by repeating the random sampling procedure 10,000 times, and considered the corresponding generic face categorisation response significant if it exceeded the top 1% of this distribution (i.e.,  $p < 0.01$ ). The first step meeting this significance criterion was taken as the threshold for generic face recognition (i.e., *Gen-Threshold*). To determine a corresponding threshold for familiar face recognition (i.e., *Fam-Threshold*), we applied the above procedure to the difference amplitude spectrum (i.e., *Familiar*–*Unfamiliar*), taking the same 99th percentile cut-off. The direction of this subtraction pinpoints the step at which the face-selective response elicited by familiar faces was significantly larger than that elicited by unfamiliar faces.

We compared the resulting distributions of individual observer thresholds using two nonparametric tests: the Wilcoxon matched-pairs signed rank test, where the  $H_0$  is equivalent central tendencies, and the two-sample Kolmogorov-Smirnov (K-S) test, which contrasts the distributions' empirical cumulative distribution functions (i.e., takes into account their global shape). To circumvent ties in the data (differences of zero that preclude the calculation of exact  $p$ -values), we added a very small amount of jitter to the test vectors ( $\pm < 0.01$  ms or cpf), and repeated this process 10,000 times. In both tests, differences were deemed significant if the mean  $p$ -value obtained across these 10,000 iterations was  $< 0.05$ ; we report these mean  $p$ -values.

### 2.9. Lateralisation analysis

Hemispheric differences at threshold points were examined only for observers for whom both *Gen-Threshold* and *Fam-Threshold* could be defined (Expt. 1:  $n = 20$ ; Expt. 2:  $n = 18$ ). For each observer, we isolated the duration/SF content step corresponding to their individually-defined threshold for generic or familiar face recognition. We then averaged across the data in these individually-defined steps in two ways: To inspect the scalp topographies at *Gen-Threshold*, we averaged each participant's *Familiar* and *Unfamiliar* data and calculated a group mean from the resulting averages. To inspect the scalp topographies at *Fam-Threshold*, we performed the *Familiar* – *Unfamiliar* subtraction for each participant and averaged across the resulting differences. Next we quantified the degree of lateralisation at each threshold point by calculating the difference between corresponding right and left electrode sites within our *a priori* OT ROI (see above), comparing these lateralisation indices at *Gen-Threshold* and *Fam-Threshold* using a two-tailed paired  $t$ -test. Lastly, to examine how potential lateralisation differences arose, we further decomposed the responses at each threshold by averaging neural activity separately for *Familiar* and *Unfamiliar* conditions.



**Fig. 2.** Group-level data for Expt. 1. *F* = Familiar, *U* = Unfamiliar. **A.** Baseline-corrected amplitude scalp topographies show the face-selective response for the Familiar condition emerging as a function of increasing duration. **B.** Normalised scalp topographies as a function of image duration for all conditions. **C.** The *Familiar* and *Unfamiliar* generic face categorisation response profiles within the OT ROI. **D.** The familiar face recognition response profile within the same ROI (right inset). Error bars are SEM, asterisks correspond to image durations eliciting a significant response ( $p < .01$ , one-tailed). **E.** Face recognition response profiles for three example participants, shown with local polynomial regression fits (see Figure S3 for all individual profiles).

### 3. Results

#### 3.1. Expt. 1: Increasing image duration

In Expt. 1, we examined how neural measures of generic and familiar face categorisation evolved as a function of increasing image duration over 12 incremental steps. At the group-level, inspection of the scalp

topographies (Fig. 2A) and response profiles within our predefined bilateral occipitotemporal (OT) ROI revealed that the generic face categorisation responses for *Familiar* and *Unfamiliar* faces emerged gradually, reaching significance at the same image duration of 33.33 ms (Fig. 2C). Inspection of the normalised topographical maps indicated stable activation of (right) OT channels throughout all supra-threshold steps (Fig. 2B), validating our *a priori* selection of those ROI electrodes

(Retter and Rossion, 2016; Quek and Rossion, 2017). In contrast, the familiar face recognition response profile hovered around the noise baseline at the shortest image durations before reaching significance at a slightly longer presentation duration (50 ms) than was observed for generic face recognition (Fig. 2D). Corresponding scalp topographies (bottom row, Fig. 2B) implicated similar bilateral OT regions as for generic face recognition. At the group-level, this familiar face recognition response increased to a local peak at 83 ms image duration, then plateaued until the last duration step, suggesting a sustained differentiation between *Familiar* and *Unfamiliar* faces. However, many individual participant response profiles in fact exhibited a transient discrimination between *Familiar* and *Unfamiliar* faces (see examples in Fig. 2E), with little difference between conditions at the longer image durations (e.g., >167 ms, for all participant profiles, see Figure S3 in supplemental materials).

To more precisely characterise the relative informational dependencies of generic and familiar face categorisation, we focused on the distributions of individual-level thresholds for each of these processes (Fig. 3). Individual observer thresholds for generic face recognition were distributed very similarly in the *Familiar* and *Unfamiliar* conditions; both distributions were narrow and peaked over 33–50 ms image duration. Both Wilcoxon and Kolmogorov-Smirnov (K-S) tests revealed no significant difference between *Gen-Threshold* distributions for *Familiar* and *Unfamiliar* faces (Wilcoxon  $p = 0.704$ ; K-S  $p = .833$ ). In contrast, thresholds for familiar face recognition were distributed much more broadly, peaking over longer durations of 83–100 ms. Nonparametric tests indicated that both central tendency and shape differed significantly between the *Fam-Threshold* and *Gen-Threshold* (averaged across *Familiar* and *Unfamiliar*) distributions (Wilcoxon  $p = .010$ ; K-S  $p = .005$ ), suggesting that familiar face recognition not only necessitated longer temporal exposure than generic face recognition, but that the former is much more dependant on the individual processing efficiency of each observer.

An interesting possibility to consider is whether observers vary reliably in terms of their minimal required image duration supporting both generic and familiar face recognition. However, we observed no systematic relationship between individual *Gen-Threshold* and *Fam-Threshold* values ( $r_{\text{pearson}} = 0.09$ ,  $p = .70$ ). As can be seen in Fig. 4A, observers with more efficient generic face recognition did not also tend to exhibit more efficient familiar face recognition. Instead, the two thresholds appeared to vary independently, such that individuals with very similar generic face recognition thresholds displayed markedly different familiar face recognition thresholds.

Interestingly, while the evolving generic and familiar face categorisation responses were both consistently located over OT channels (Fig. 2B), the former appeared to be more strongly right-lateralised, suggesting there may be different underlying neural regions associated with processing critical sensory input required for each recognition function. This pattern was particularly evident in the scalp topographies corresponding to the (individually-defined) threshold points (Fig. 5A), where the lateralisation index (i.e., right ROI – left ROI) was significantly stronger at *Gen-Threshold* than at *Fam-Threshold*,  $t(1,19) = 2.51$ ,  $p < .02$  (Fig. 5B). Further decomposing this pattern revealed that, at the shortest durations supporting the distinction between faces and non-faces, responses were similarly right-lateralised for both *Familiar* and *Unfamiliar* faces (Fig. 5C). By contrast, at the shortest durations supporting *Familiar/Unfamiliar* face differentiation, *Familiar* faces evoked a more bilateral face-selective response than *Unfamiliar* faces.

### 3.2. Expt. 2: Increasing image content

In Expt. 2, we examined neural measures of generic and familiar face recognition as a function of increasing spatial frequency (SF) content. At the group-level, the face categorisation response emerged at the same coarse image resolution for both *Familiar* and *Unfamiliar* faces (i.e., *Gen-Threshold* = 3.5 cpf or 7.4 cpi) and increased steadily before stabilising around 10.5 cpf (Fig. 6B). As was the case in Expt. 1, inspection of the

normalised scalp topographies (Fig. 6A) indicated consistent activation of lateral occipitotemporal channels across increasing image content, suggesting that similar OT neural populations were engaged regardless of the spatial frequency content of the face images. Although the group-level face recognition response (Fig. 6C) also rose significantly above noise at 3.5 cpf, this difference between *Familiar* and *Unfamiliar* responses did not stabilise until a slightly higher image resolution (i.e., 5.5 cpf, or 11.6 cpi). Just as in Expt. 1, the *Familiar* and *Unfamiliar* face responses for individual observers were somewhat dissociated from the group-level average response profiles (see Figure S4 in supplemental materials), once again bolstering our approach to focus on threshold differences at the individual-level.

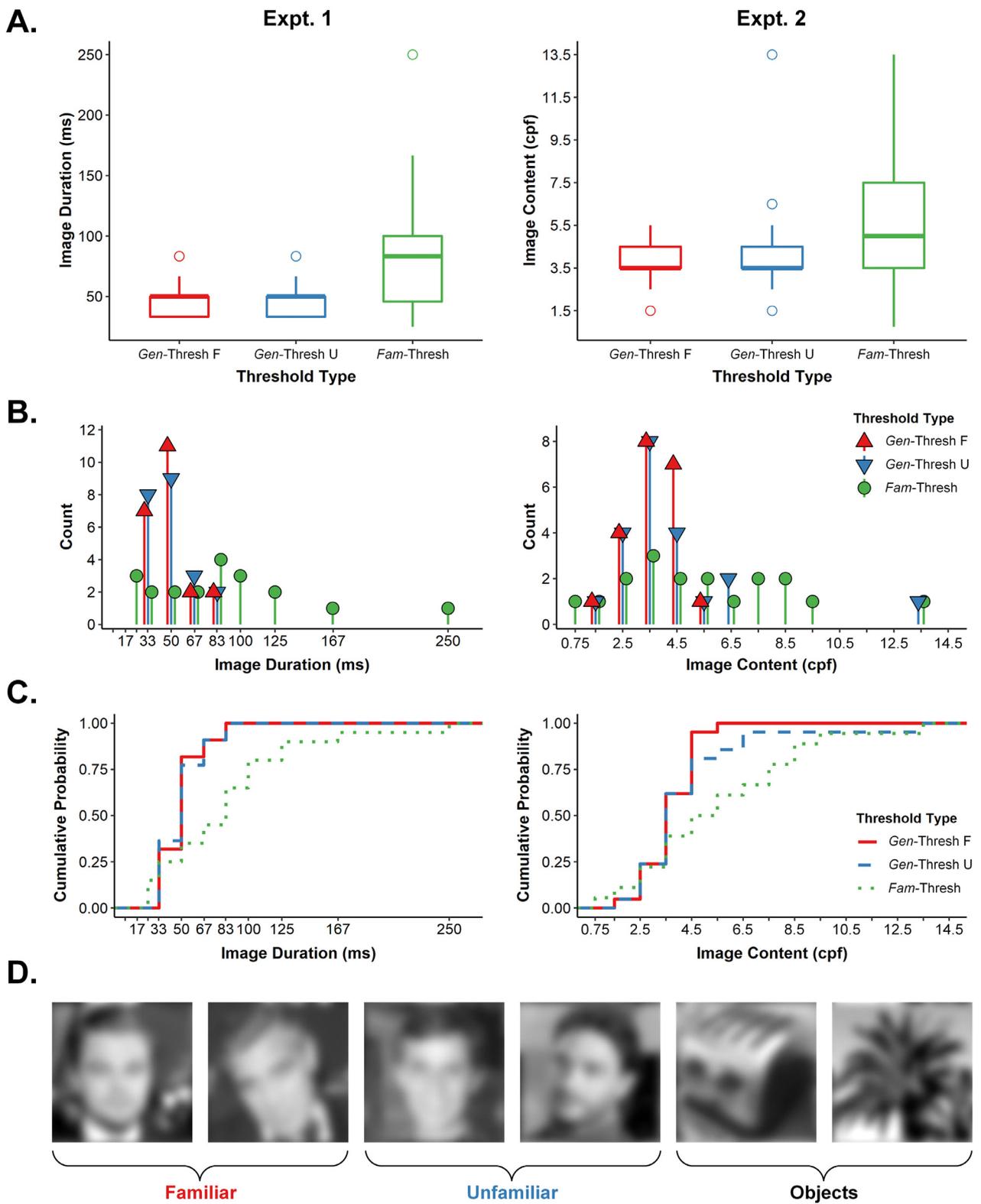
Fig. 3 (right column) shows the individual threshold distributions for generic and familiar face recognition in Expt. 2. Results mirrored the pattern observed in Expt. 1: *Gen-Threshold* was very similarly distributed for the *Familiar* and *Unfamiliar* conditions, peaking in both cases over a narrow range of 2.5–4.5 cpf (5.3–9.5 cpi) (Wilcoxon  $p$ -value = 0.572; K-S  $p$ -value = 0.803). Interestingly, while the *Fam-Threshold* distribution was centred over the same cpf range as *Gen-Threshold* (Wilcoxon  $p$ -value = 0.109), its shape was noticeably different (K-S  $p$ -value = 0.036), being much wider and characterised by a long rightward tail. These results point to greater inter-observer variability in the image resolution needed to perceive the familiarity status of a face than to perceive its category: While coarse information does indeed appear to be sufficient for some observers to reliably recognise familiar faces, still others require much finer image detail to make this distinction successfully.

As for Expt. 1, we found no evidence of general inter-individual differences affecting the efficiency of both generic and familiar face categorisation processes, in that *Gen-Threshold* and *Fam-Threshold* values did not vary systematically with individual observers ( $r_{\text{pearson}} = 0.15$ ,  $p = .52$ ). Instead, participants for whom generic face categorisation was possible based on extremely coarse visual input required widely varying levels of increased resolution to recognise face familiarity. Finally, inspection of the topographical distribution at the individual observer thresholds revealed a similar profile as obtained in Expt. 1 (Fig. 5D): Where scalp topographies at *Gen-Threshold* were evidently right-lateralised, we observed a more bilateral topography at *Fam-Threshold*, although the difference between these lateralisation indices did not reach statistical significance,  $t(1,17) = 1.56$ ,  $p = .14$  (Fig. 5E). Once again, this pattern again appeared to be driven by more distributed hemispheric engagement when viewing *Familiar* faces compared to *Unfamiliar* ones (Fig. 5F).

## 4. Discussion

Encountering a face in the real world provokes a barrage of functional categorisations at once. In what feels like the same instant, the observer knows that the stimulus before them is indeed a face and not another type of object, that it is female and not male, that it is well-known to them, that its expression is happy, and so on. In a bid to increase our understanding of how the brain achieves these manifold categorisations at the same face encounter, here we provide a systematic investigation of the relative informational dependencies underlying two ecologically relevant brain functions that arise each time we see a face – generic face recognition and familiar face recognition.

Using visual periodicity to isolate the selective neural response to unsegmented faces presented amidst a wide variety of nonface stimuli, we identified individual observer thresholds for generic and familiar face recognition within two parametric manipulations of sensory input. In Expt. 1, manipulating image duration showed that exposures of just 33–50 ms enabled nearly all observers to consistently distinguish faces from a host of other categories (e.g., animals, plants, buildings, vehicles, etc.). In contrast, the temporal exposure required to recognise whether the faces were familiar was both higher on average and much more variable ( $M = 83$  ms, range = 25–250 ms). Manipulating image resolution in Expt. 2 yielded similar findings: Nearly all participants recognised faces amongst nonface stimuli based on extremely coarse visual



**Fig. 3.** Distributions of individual *Gen-Thresh* and *Fam-Thresh* values in Expt. 1 (left column) and Expt. 2 (right column), represented as **A.** box-and-whisker plots, **B.** Frequency counts, and **C.** Empirical cumulative distribution functions. **D.** Examples of *Familiar*, *Unfamiliar* and *Object* stimuli at the median *Gen-Thresh* for Expt. 2 (3.5 cpf).

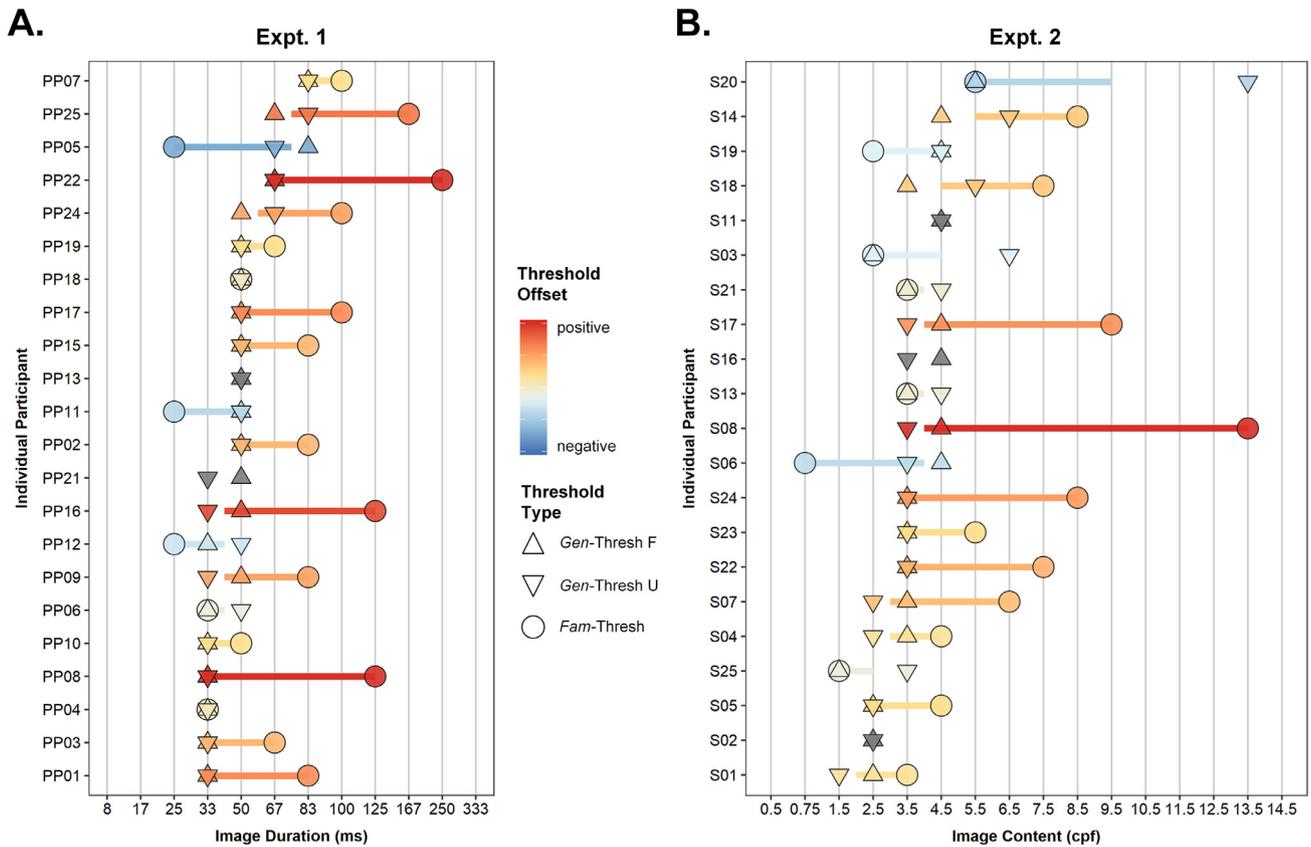


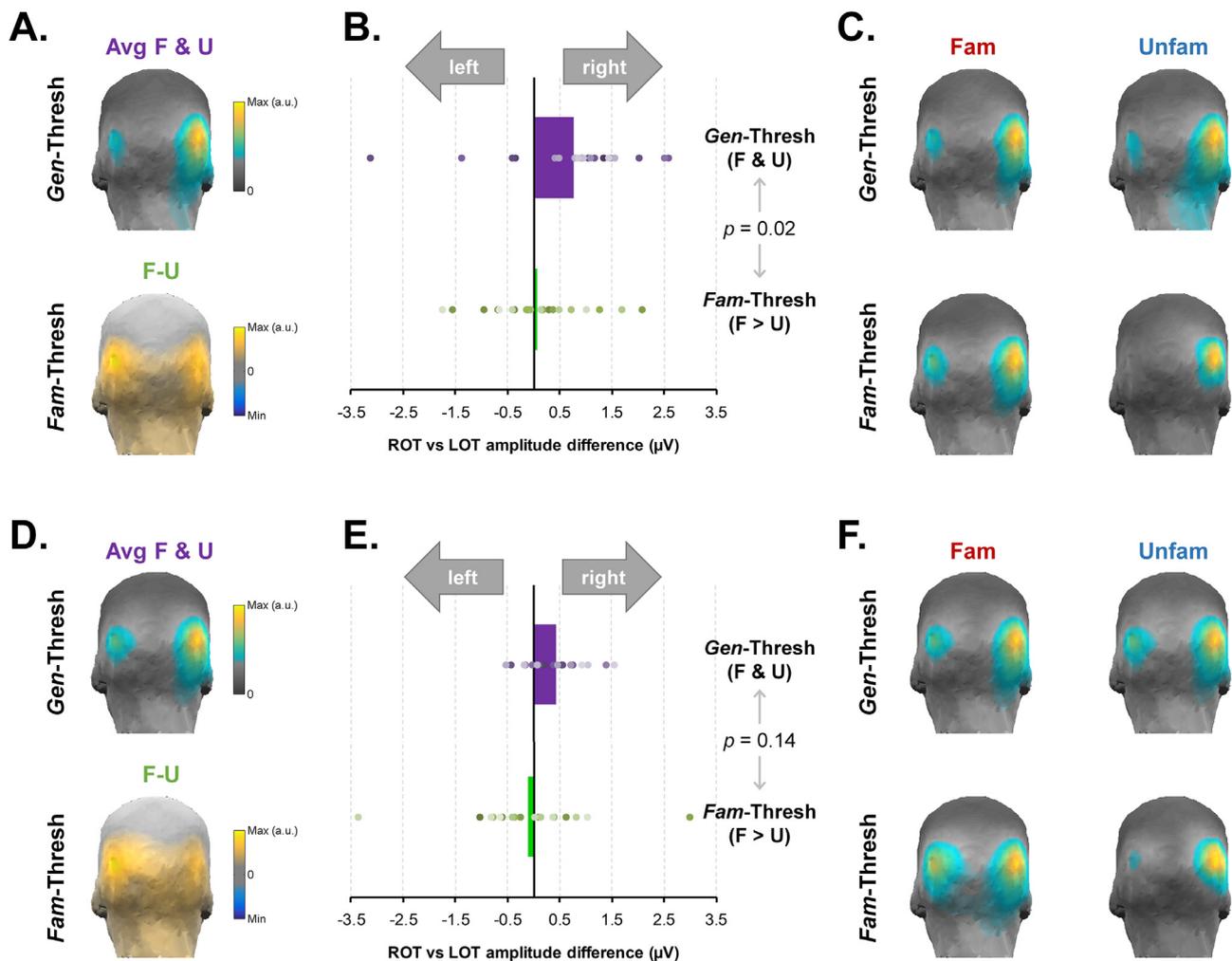
Fig. 4. Individual threshold values for Expt. 1 (A) and Expt. 2 (B), ordered by decreasing *Gen-Thresh* value (averaged across F and U). Colours show the degree and direction of offset between *Gen-Thresh* and *Fam-Thresh* (grey points = subjects with no identifiable *Fam-Thresh*).

input (i.e., just 2.5–4.5 cpf), but they exhibited considerable variability in how much finer resolution they required to recognise face familiarity (range = 1.5–13.5 cpf). In both experiments, the scalp distribution of neural responses was right-lateralised to a greater extent at the generic categorisation threshold than at the familiar face categorisation threshold. Conditional decomposition at individual familiarity thresholds showed that while unfamiliar face presentations engaged predominantly right OT regions, familiar face presentations evoked a much more bilateral pattern of activation. Notably, neither manipulation of face information revealed a systematic offset in generic and familiar face categorisation across individual participants (i.e., observers who categorised faces vs. objects at short durations were not also able to recognise familiar faces at comparatively short durations). Taken together, the two lines of evidence presented here indicate that during a given face encounter, observers need less sensory evidence to successfully recognise that stimulus as a face than they do to recognise whether the face is familiar. In other words, generic face recognition is both more efficient, less variable, and processed in a more spatially constrained neural network than familiar face recognition.

To some extent, the observation that face ‘detection’ should precede ‘identification’ from an informational standpoint appears rather intuitive. To date, however, evidence for this claim has comprised second-order comparisons of information thresholds obtained using explicit categorisation tasks, where observers judge highly simplified and/or uniform faces along a single dimension at a time. Since informational minima identified under this ‘modular’ framework must certainly in part reflect stimuli- and task-specific observer strategies, comparing thresholds across different explicit recognition tasks is less than ideal. Our study makes an important advance in this regard, tracking the informational requirements of two different recognition functions as elicited by the same face encounters, in the absence of any explicit recognition task.

Here, face images appeared amidst a wide variety of nonface categories, such that observers were obliged to categorise each face in the sequence at both the generic and familiar level (i.e., *Is that a face?* AND *Do I know that face identity?*). We emulated the high variability that the visual system readily contends with in everyday face recognition by employing a wide set of naturalistic face images (see Fig. 1 & Figure S1) which guaranteed that specific low-level features would not systematically occur at the 1 Hz frequency where we quantified the response (Rossion et al., 2015; Gao et al., 2018) (as might conceivably occur with full frontal, closely cropped face images). In this way, we can be confident that our sensory input thresholds pertain to a form of face categorisation that goes beyond simple low-level differences between categories, and that these informational minima bear greater relevance to real-world face encounters (where the same stimulus is evidently categorised along multiple dimensions at once) than those identified by studies focused on just one aspect of recognition at a time. As such, our finding of lower informational requirements for generic vs. familiar face recognition constitutes some of the strongest evidence to date for the intuitive notion that sensory input demands should be lower for recognising face category than face identity.

At the same time, however, it remains the case that for a subset of our observers, the same informational content gave rise to successful face recognition at both the generic and familiar levels (Tanaka, 2001) (see Fig. 4). While this finding obviously requires replication before drawing any strong conclusions, it does somewhat temper the claim that in order to know whether a face is familiar or not, you must presumably already know that the stimulus is a face (but see Grill-Spector and Kanwisher, 2005). This *basic-before-subordinate* notion (Rosch et al., 1976) underlies many classic theoretical accounts in which face recognition is claimed to unfold in a serial fashion from visual analysis (e.g., structural encoding of features) all the way up to the retrieval of high-level



**Fig. 5.** Response lateralisation at *Gen-* and *Fam-*Threshold in Expt.1 (top row) and Expt. 2 (bottom row). **A. & D.** Mean normalised topographies at the individual observer values of *Fam-*Threshold and *Gen-*Threshold (averaged across *Familiar/Unfamiliar*). **B. & E.** Differences in baseline-corrected amplitudes between the right and left OT ROIs for each threshold type. Bars reflect the group mean; dots are individual observers. **C. & F.** Decomposition of the mean normalised topographies at threshold into the *Familiar* and *Unfamiliar* conditions. Scalp activation at *Fam-*Threshold was comparatively more bilateral for *Familiar* faces than *Unfamiliar* ones.

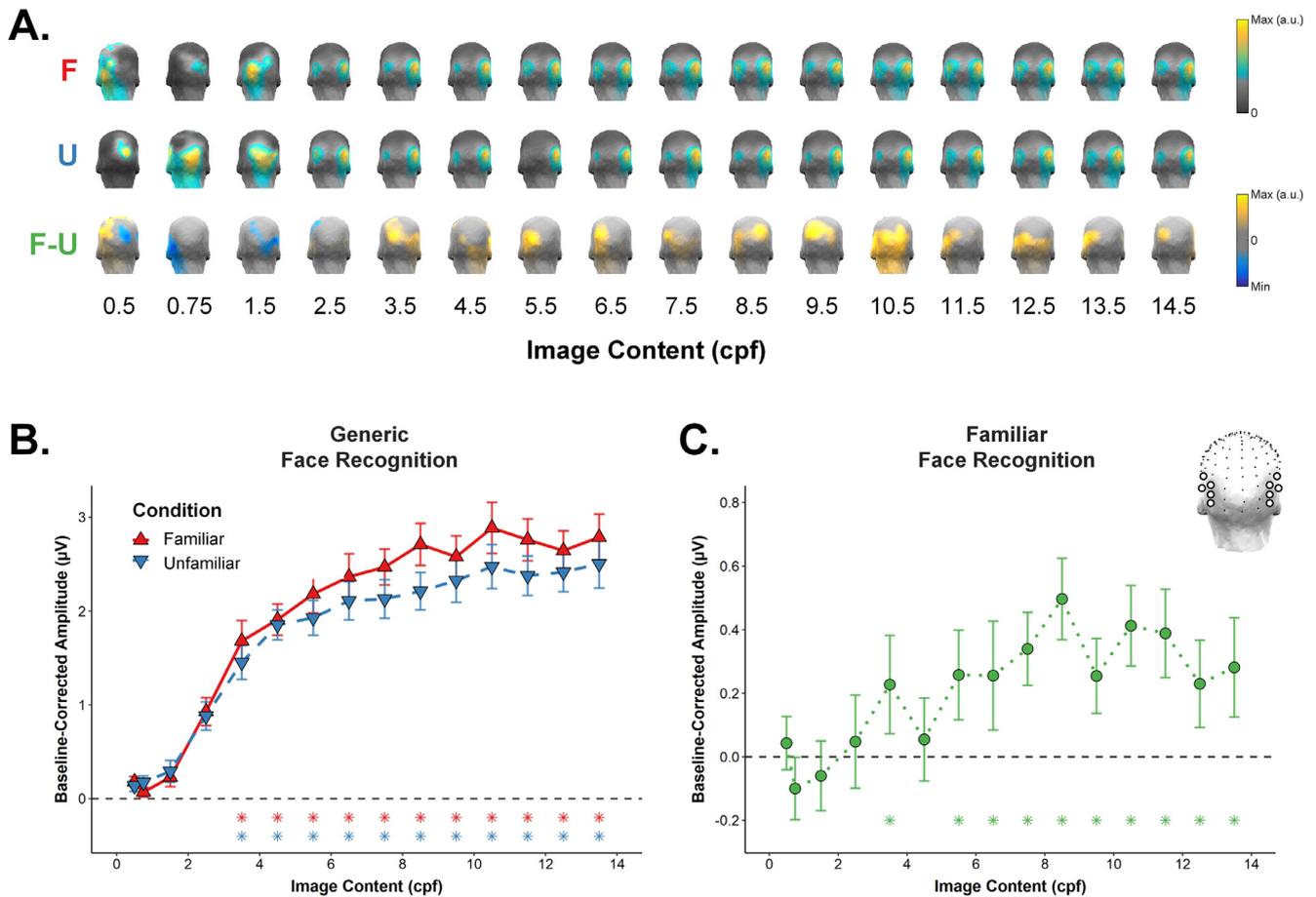
semantic associations (e.g., the name and context associated with that identity) (Burton et al., 1999; Bruce and Young, 1986). By starting with the premise that a face is present in foveal vision, these models tacitly imply that low-level detection (e.g., figure-ground segmentation) and recognition of face category have already taken place. The fact that this assumption does not hold for all observers in our experiments suggests generic face recognition is a nontrivial stage function that should be explicitly incorporated into extant face recognition models.

Our approach of indexing two dimensions of face recognition prompted by the same face encounter aligns with the field's recent shift toward acknowledging the inherently multifaceted nature of real-world face recognition. Several recent studies have used time-resolved multivariate decoding to probe the overlapping time-courses of distinct face categorisations reflected within the same neural response (i.e., decoding “which information is available when” (Nemrodov et al., 2016; Ambrus et al., 2019; Dobs et al., 2019; Ghuman et al., 2014)). In contrast, our approach of quantifying the differential response to faces amongst nonfaces regardless of exact timing (e.g., early/fast vs. late/slow) enabled us to inspect the temporal/spatial sensory input requirements for generic and familiar face recognition independently of any differences in their representational time-courses. In this way, our findings bring an complementary perspective to the origin of putative processing speed differences between various dimensions of face recognition: Namely, if two components of face recognition differ in

terms of how soon they can be achieved (Besson et al., 2017; Barragan-Jason et al., 2012) or decoded (Dobs et al., 2019; Ghuman et al., 2014) after face onset, this might not strictly reflect a difference in their relative speed of processing *per se*, but could equally arise if the processes unfold at the same speed, but with one requiring comparatively more evidence (i.e., signal-to-noise) than the other, and therefore being completed later (VanRullen, 2011).

#### 4.1. The minimum presentation duration supporting generic face recognition vs. familiar face recognition

Although our study is the first to characterise how increasing image duration affects different forms of recognition arising at the same face encounter, many extant studies have considered the minimum viewing time required for isolated aspects of face recognition (Grill-Spector et al., 2000; Tanskanen et al., 2007; Or and Wilson, 2010). One such modular study that presented degraded synthetic faces at various brief durations found that observers required just 32 ms image exposure to reach 75% accuracy for detecting whether a face was present/absent, but needed nearly double that time to reach the same threshold when identifying which of two facial identities had appeared (Or and Wilson, 2010). Our Expt. 1 findings accord well with these results in suggesting that sensory input demands are higher for more complex forms of face recognition (on average, 33–50 ms for generic face recognition vs. 83 ms for



**Fig. 6.** Group-level data for Expt. 2. **A.** Scalp topographies normalised to highlight the peak response locations across cpf values. **C.** Conditional mean face categorisation response profiles within the OT ROI (shown as right inset). **D.** The familiar face categorisation response profile within the same ROI. Error bars are SEM, asterisks indicate SF cutoffs at which a significant response was identified ( $p < .01$ , one tailed).

familiar face recognition). Notably, however, the same brief exposure durations that enable observers in that previous study to differentiate between faces and very simple distractors (i.e., scrambled face/oval outline) were also sufficient for our participants to distinguish faces from a rich variety of nonface stimuli (i.e., animals, plants, man-made artefacts, buildings, etc.). This highlights the efficiency of generic face recognition mechanisms, which appear capable of extracting highly complex visual information contained in natural images as well as simplistic features at very short image durations. In a similar vein, our critical image durations supporting successful generic face recognition neatly overlap those reported by a recent frequency-tagging investigation of this same face function that used full-colour images (instead of the greyscale images we used here) and a completely orthogonal observer task (Retter et al., 2020) (i.e., ~33–50 ms vs. ~33–42 ms). Together with the reliable lower bound in individual observer thresholds in our Expt. 1, this consistency of image duration thresholds across differences in task and stimulus characteristics underscores the robustness and automaticity of generic face recognition in the human brain. Put differently, it appears there is indeed a hard limit to the minimum sensory input the human visual system requires to discern faces from other natural categories – one that does not seem to be lowered by attentional facilitation or richer image signal (Fabre-Thorpe, 2011).

In contrast to the extremely brief exposures enabling successful generic face recognition, comparatively longer image durations have been implicated in identity-level face recognition. One study showed that 6-way identity discrimination was no better than chance for (masked) image durations of 17 ms and 33 ms, but increased markedly from 50 ms upward (Tanskanen et al., 2007). Elsewhere, the tempo-

ral processing capacity for face identification (one image per identity) has been estimated at 10 Hz, with 100 ms exposure providing sufficient inspection time for observers to distinguish recently learned faces from novel distractor ones (Näsänen et al., 2006). Yet since the duration threshold reported by this study corresponded to an arbitrary performance level (79%), it is likely that this 100 ms estimate represents neither the minimal nor the optimal image duration for face identification. Indeed, no study to date has conclusively established the minimum viewing time required for successful recognition of highly familiar (i.e., famous) faces. The data presented here indicate that this temporal threshold is, on average, very low (i.e., 50–83 ms), and that the informational dependency of familiar face recognition varies rather widely across individuals.

What gives rise to the wide distribution of familiarity thresholds in our data? From our perspective, this variability is unlikely to have resulted from differing degrees of famous identity familiarity in our participant sample, for two reasons. First, our separate stimulus prescreening experiment demonstrates that recognition rates for these specific *Familiar* identities are very much at ceiling in our test population (see Figure S2 in Supplemental materials). Second, we took the extra step of excluding participants who performed <2.5 SD below average on the post-sequence 2AFC task (1 subject in Expt 1; 2 subjects in Expt 2). Although clearly not designed to be a sensitive index of (overt) familiar face recognition, removing poor performers on this simple task does to some extent guard against the possibility that globally “poor recognizers” account for the right tail of our familiar face recognition threshold distribution. Moreover, we found no compelling evidence for a systematic relationship between participants’ neural thresholds for

familiar face recognition and their performance on the post-sequence identity recognition task (see Figure S5 in supplemental materials). As such, a clear target for follow-up investigation are the factors driving high variability in sensory input requirements for recognising identity, which could include differences in the way observers encode specific visual features of certain identities, or in the strength of our semantic associations for those people.

One important question that could arise in considering our data is whether we can reasonably draw conclusions about the informational requirements of familiar face recognition given that our familiarity index is based on a subtraction of signals that reflect face vs. nonface categorisation. Crucial to note here is that while the *presence* of a significant response at the face presentation frequency acts as a functional marker of successful generic face categorisation (since this 1 Hz signal will only arise if the neural response elicited by faces differs consistently from that elicited by nonfaces), the face-selective response itself inherently contains much more information than just perceptual category. Onsetting around 100 ms post stimulus, this complex response lasts around 420 ms and unfolds over four component time-windows from posterior to anterior occipitotemporal regions (Retter and Rossion, 2016; Jonas et al., 2016). As such, indexing familiarity representations by subtracting the quantified face-selective responses for familiar and unfamiliar faces is not only valid, but specifically advantageous insofar as it ensures we examine how Familiarity modulates only face-selective neural activity. To the best of our knowledge, ours is the first study to achieve this. At the same time, however, it should be noted that the viewing time threshold we report here (i.e., 50–83 ms) pertains to recognising face familiarity in the context of visually dissimilar distractors (akin to unexpectedly encountering a face in the environment). Distinguishing familiar and unfamiliar faces directly within a periodicity based design (e.g., UU-UUFUUUF...) (Yan and Rossion, 2020) would likely demand a longer minimum exposure duration due the high stimulus similarity and resulting increased masking (akin to recognising a familiar face in a crowd of people).

#### 4.2. The lowest spatial frequencies supporting generic face recognition vs. familiar face recognition

Alongside image duration manipulations, face recognition research also has a long history of impoverishing faces by selectively removing content at certain spatial scales (i.e., spatial frequencies, SFs) (Ruiz-Soler and Beltran, 2006; Morrison and Schyns, 2001) – a manipulation thought to correspond to testing the distance at which faces can be recognised (Loftus and Harley, 2005). Previous studies using explicit categorisation tasks have reported diagnostic SF bandwidths for various aspects of face categorisation, to date largely focussing on higher-level forms of recognition (e.g., identity, emotional expression, gender, etc.). In contrast, investigations of critical SFs for recognising a face as a *face* have been much rarer (Harel and Bentin, 2009). Our Expt. 2 results make an important contribution in this regard, establishing the relative involvement of SF content for both the basic and finer recognition of a face within the same individuals, at the same face encounters.

In terms of generic face recognition, the ultra-coarse information threshold in Expt. 2 (i.e., 2.5 – 4.5 cpf) is consistent with prior studies that have emphasised the importance of low SFs for this function (Owsley and Sloane, 1987; Goffaux et al., 2003), albeit using more simplistic indices of this process (e.g., contrasting faces with just a single other category, such as cars). To date, only one other study has manipulated SF content for generic face recognition in the context of multiple naturalistic visual categories (Quek et al., 2018). In our previous work, we demonstrated that the brain can achieve this function based on images containing less than 5 cycles per image (corresponding to < 2 cpf). Together with the current results, these findings suggest that discriminating faces from nonfaces in natural environments depends on detecting the global structure of a face, rather than individual facial features

which are not yet evident at the coarse spatial resolution of threshold images (see Fig. 4D).

In contrast to generic face recognition, the contribution of different spatial scales to familiar face recognition has received comparatively greater attention, with general agreement that the critical SF bandwidth supporting identity-related face tasks comprises midrange values between 8 and 16 cpf (Näsänen, 1999; Gold et al., 1999). However, these studies have largely focused on identifying the *optimal* SF range for familiar face recognition, comparing performance for different bandpass-filtered versions of the exact same face stimuli (typically just one full-frontal image per identity, presented without any external features). In contrast, we focused here on identifying the *minimal* amount SF content capable of driving successful familiar face recognition. We filtered images with a progressively increasing low-pass cutoff that enabled us to track the impact of cumulatively-integrated SF information on generic and familiar face recognition, mimicking coarse-to-fine information processing in naturalistic vision (Quek et al., 2018). This approach diverges from other methodologies that are optimised for probing the related (but distinct) issue of which specific features within a given SF range are diagnostic for functional categorisations of a face (Gosselin and Schyns, 2001; Dakin and Watt, 2009; Pachai et al., 2018). Where the latter techniques are classically used in conjunction with both explicit (overt) categorisation tasks and/or highly normalised, full-frontal view faces, our approach enabled us to capture naturally-arising categorisations of unsegmented faces, preserving an important aspect of recognizing faces in natural environments, i.e., figure-ground segmentation. Unexpectedly, this approach revealed a degree of overlap in the coarse spatial resolutions (i.e., ~3.5 cpf) capable of driving both successful generic and familiar face recognition in some individuals. Importantly, however, the fact that familiar face recognition *can* proceed based on highly degraded visual input does not imply that this process is optimally subserved by such limited image information. Rather, in the same way that finer-scale information improves the recognition of the stimulus as a face (Halit et al., 2006), the addition of high SF details likely serves to refine face identity representations, accentuating the difference between familiar and unfamiliar faces (Sergent, 1986). The rising familiar recognition response profiles in our own data suggest this is indeed the case.

#### 4.3. Broader implications and future directions

At a broader level, the current findings can be contextualised within the coarse-to-fine visual processing framework (Morrison and Schyns, 2001; Sergent, 1986; Goffaux et al., 2011; Hegdé, 2008), wherein limiting the temporal exposure of full-spectrum (i.e., unfiltered) images essentially serves to constrain the spatial scales the observer is able to extract while the image is onscreen. In this way, our two manipulations of face information and resulting findings potentially represent two sides of the same coin: upon encountering a face, the human visual system may rapidly extract its (coarse) global structure that suffices to distinguish it from other object categories, with more fine-grained sensory cues that facilitate familiarity recognition being more gradually accumulated. Note, however, that the threshold offset between the two recognition functions should not imply that these phenomena represent different points on the same general evidence accumulation profile. That is, we do not consider familiar face recognition to be a simple extension of processes supporting generic face recognition, particularly since it is clear that the former necessarily involves access to an observer's long-term memory representations of that identity. Rather, processes supporting the two functions may operate in parallel, and could potentially even be sub-served by partly different neural substrates, as their dissociation in cases of acquired prosopagnosia alludes to (Rossion et al., 2011). Our data hint at this possibility, insofar as distinguishing faces from nonfaces mainly activated right-lateralised posterior sites, while recognising face familiarity appeared to recruit a more bilateral network. In particular, at the individual threshold points for familiar face recognition, familiar

faces seemed to engage both the left and right OT regions, where unfamiliar faces predominantly activated only right OT. However, given the limited spatial resolution of EEG, it is clear that future studies combining the current approach with intracerebral recordings or fMRI (Gao et al., 2018) will be important to clarify this outstanding issue. In the same vein, high spatial resolution measures could also be used to determine the overlap in neural populations driving the selective response to familiar and unfamiliar faces at the generic recognition threshold. That is, while the *magnitude* of the face-selective response evoked during familiar and unfamiliar sequences is similar enough to result in identical generic face categorisation identical thresholds, the neural populations that give rise to the two responses we measure here with EEG could well be distinct (e.g., already tightly clustered for familiar and more diffuse for unfamiliar).

Finally, although we might have expected to find a systematic relationship between an individual observer's sensory input thresholds, in fact observers in both experiments exhibited highly consistent thresholds for generic face recognition compared with more variable thresholds for familiar face recognition. To the best of our knowledge, this relative difference in individual variance is a novel empirical observation with important implications for the neurofunctional basis of each process. That generic face recognition proceeds based on extremely impoverished information underlines the "hardwired" nature of this function – a process that both develops early (de Heering and Rossion, 2015) and appears robust to brain lesions that otherwise severely impair familiar face recognition (Busigny et al., 2010). In contrast, the maturation of face identity processing, including familiar face recognition, is still much debated (Crookes and McKone, 2009; Germine et al., 2011) given the wide inter-individual variability in this function (Behrmann and Avidan, 2005; Russell et al., 2009). Related to this point, a small subset of our sample actually exhibited a lower sensory threshold for successful familiar face recognition than generic face recognition. While these "negative" offsets should be interpreted with caution (since they may reflect signal-to-noise fluctuations associated with quantifying responses at the single-subject level), it is nevertheless interesting to speculate whether some individuals might detect the familiarity of a stimulus regardless of its visual category. Ascertaining whether these patterns represent a true functional phenomenon will necessitate probing the functional relevance of the neural thresholds reported here, by directly relating them to behavioural outcomes.

## 5. Conclusion

Effective social behaviour in the real world depends on our ability to efficiently categorise a face along multiple dimensions (e.g., sex, emotion, identity, etc.) at once. How the human brain achieves these manifold aspects of categorisation within individual face encounters is not yet well understood. Here we characterized the relative informational dependencies – quantified in space and time – of two critical recognition functions evoked at every face encounter: generic and familiar face recognition. We show that the sensory evidence required by the human brain to distinguish faces from other categories is systematically lower than that required to recognise faces we know, and that informational thresholds supporting familiarity recognition vary widely across observers. These findings underscore the neurofunctional distinctions between these two recognition functions - generic face recognition being robust in terms of its minimal sensory input requirements and low inter-individual variability, relative to the more demanding and idiosyncratic processes of familiar face recognition.

## Author contributions

Research design: GLQ & JL-S;

Data collection: JL-S & GLQ;

Data analysis: JL-S & GLQ;

Data interpretation & manuscript preparation: GLQ, JL-S, & BR

## Acknowledgments

This work was supported in part by a European Research Council grant awarded to BR [grant no: [facesvpep 284025](#)], an FSR-FNRS post-doctoral grant awarded to JLS [grant no: [FC 91608](#)], a co-funded initiative by the University of Louvain and the Marie Curie Actions of the European Commission awarded to GLQ [grant no: [F211800012](#)], and the European Union's Horizon 2020 research and innovation programme via a Marie Skłodowska-Curie Individual Fellowship awarded to GLQ [grant no: 841909]. We thank A. Conte for developing the Java stimulation code and V. Goffaux for her help in generating the low-pass filtered images in Expt. 2.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2021.118481](#).

## References

- Ales, J.M., Farzin, F., Rossion, B., Norcia, A.M., 2012. An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *J. Vis.* 12. doi:[10.1167/12.10.18](#), 18-18.
- Ambrus, G.G., Kaiser, D., Cichy, R.M., Kovacs, G., 2019. The Neural Dynamics of Familiar Face Recognition. *Cereb. Cortex* doi:[10.1093/cercor/bhz010](#).
- Barragan-Jason, G., Lachat, F., Barbeau, E.J., 2012. How Fast is Famous Face Recognition? *Front. Psychol.* 3, 454. doi:[10.3389/fpsyg.2012.00454](#).
- Barragan-Jason, G., Besson, G., Ceccaldi, M., Barbeau, E.J., 2013. Fast and Famous: looking for the Fastest Speed at Which a Face Can be Recognized. *Front. Psychol.* 4, 100. doi:[10.3389/fpsyg.2013.00100](#).
- Behrmann, M., Avidan, G., 2005. Congenital prosopagnosia: face-blind from birth. *Trends Cogn. Sci.* 9, 180–187.
- Besson, G., et al., 2017. From face processing to face recognition: comparing three different processing levels. *Cognition* 158, 33–43.
- Bruce, V., Young, A., 1986. Understanding face recognition. *Br. J. Psychol.* 77, 305–327. doi:[10.1111/j.2044-8295.1986.tb02199.x](#).
- Burton, A.M., Bruce, V., Hancock, P.J.B., 1999. From pixels to people: a model of familiar face recognition. *Cogn. Sci.* 23, 1–31. doi:[10.1016/S0364-0213\(99\)80050-0](#).
- Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., Rossion, B., 2010. Holistic perception of the individual face is specific and necessary: evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia* 48, 4057–4092.
- Carlson, T., Tovar, D.A., Alink, A., Kriegeskorte, N., 2013. Representational dynamics of object vision: the first 1000ms. *J. Vis.* 13, 1.
- Crookes, K., McKone, E., 2009. Early maturity of face recognition: no childhood development of holistic processing, novel face encoding, or face-space. *Cognition* 111, 219–247.
- Crouzet, S.M., Kirchner, H., Thorpe, S.J., 2010. Fast saccades toward faces: face detection in just 100ms. *J. Vis.* 10. doi:[10.1167/10.4.16](#), 16-16.
- Dakin, S.C., Watt, R.J., 2009. Biological "bar codes" in human faces. *J. Vis.* 9. doi:[10.1167/9.4.2](#), 2.1-10.
- de Heering, A., Rossion, B., 2015. Rapid categorization of natural face images in the infant right hemisphere. *eLife* 4, e06564.
- Dobs, K., Isik, L., Pantazis, D., Kanwisher, N., 2019. How face perception unfolds over time. *Nat. Commun.* 10, 1258. doi:[10.1038/s41467-019-09239-1](#).
- Ekman, P., 1993. Facial expression and emotion. *Am. Psychol.* 48, 384–392. doi:[10.1037/0003-066X.48.4.384](#).
- Fabre-Thorpe, M., 2011. The Characteristics and Limits of Rapid Visual Categorization. *Front. Psychol.* 2. doi:[10.3389/fpsyg.2011.00243](#).
- Gao, X., Gentile, F., Rossion, B., 2018. Fast periodic stimulation (FPS): a highly effective approach in fMRI brain mapping. *Brain Struct. Funct.* 223, 2433–2454. doi:[10.1007/s00429-018-1630-4](#).
- Germine, L.T., Duchaine, B., Nakayama, K., 2011. Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition* 118, 201–210. doi:[10.1016/j.cognition.2010.11.002](#).
- Ghuman, A.S., et al., 2014. Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* 5, 5672. doi:[10.1038/ncomms6672](#).
- Gobbini, M.I., Haxby, J.V., 2007. Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi:[10.1016/j.neuropsychologia.2006.04.015](#).
- Goffaux, V., Gauthier, I., Rossion, B., 2003. Spatial scale contribution to early visual differences between face and object processing. *Brain Res. Cogn. Brain Res.* 16, 416–424.
- Goffaux, V., et al., 2011. From Coarse to Fine? Spatial and Temporal Dynamics of Cortical Face Processing. *Cereb. Cortex* 21, 467–476. doi:[10.1093/cercor/bhq112](#).
- Gold, J., Bennett, P.J., Sekuler, A.B., 1999. Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Res.* 39, 3537–3560. doi:[10.1016/S0042-6989\(99\)00080-2](#).
- Gosselin, F., Schyns, P.G., 2001. Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res.* 41, 2261–2271. doi:[10.1016/S0042-6989\(01\)00097-9](#).
- Grill-Spector, K., Kanwisher, N., 2005. Visual recognition - As soon as you know it is there, you know what it is. *Psychol. Sci.* 16, 152–160.

- Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* 3, 837.
- Halit, H., De Haan, M., Schyns, P., Johnson, M., 2006. Is high-spatial frequency information used in the early stages of face detection? *Brain Res.* 1117, 154–161.
- Harel, A., Bentin, S., 2009. Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1264–1273. doi:10.1037/a0013621.
- Hegd , J., 2008. Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84, 405–439.
- Hill, H., Bruce, V., Akamatsu, S., 1995. Perceiving the sex and race of faces: the role of shape and colour. *Proc. R. Soc. Lond. B. Biol. Sci.* 261, 367–373.
- Jonas, J., et al., 2016. A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4088–E4097. doi:10.1073/pnas.1522033113.
- Loftus, G.R., Harley, E.M., 2005. Why is it easier to identify someone close than far away? *Psychon. Bull. Rev.* 12, 43–65. doi:10.3758/BF03196348.
- Morrison, D.J., Schyns, P.G., 2001. Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychon. Bull. Rev.* 8, 454–469. doi:10.3758/bf03196180.
- N s nen, R., Ojanp a, H., Tanskanen, T., P allysaho, J., 2006. Estimation of temporal resolution of object identification in human vision. *Exp. Brain Res.* 172, 464–471. doi:10.1007/s00221-006-0354-5.
- N s nen, R., 1999. Spatial frequency bandwidth used in the recognition of facial images. *Vision Res.* 39, 3824–3833. doi:10.1016/S0042-6989(99)00096-6.
- Nemrodov, D., Niemeier, M., Mok, J.N.Y., Nestor, A., 2016. The time course of individual face recognition: a pattern analysis of ERP signals. *Neuroimage* 132, 469–476. doi:10.1016/j.neuroimage.2016.03.006.
- Or, C.C.F., Wilson, H.R., 2010. Face recognition: are viewpoint and identity processed after face detection? *Vision Res.* 50, 1581–1589. doi:10.1016/j.visres.2010.05.016.
- Owsley, C., Sloane, M.E., 1987. Contrast sensitivity, acuity, and the perception of 'real-world' targets. *Br. J. Ophthalmol.* 71, 791–796.
- Pachai, M.V., Bennett, P.J., Sekuler, A.B., 2018. The Bandwidth of Diagnostic Horizontal Structure for Face Identification. *Perception* 47, 397–413. doi:10.1177/0301006618754479.
- Quek, G.L., Rossion, B., 2017. Category-selective human brain processes elicited in fast periodic visual stimulation streams are immune to temporal predictability. *Neuropsychologia* 104, 182–200. doi:10.1016/j.neuropsychologia.2017.08.010.
- Quek, G.L., Liu-Shuang, J., Goffaux, V., Rossion, B., 2018a. Ultra-coarse, single-glance human face detection in a dynamic visual stream. *Neuroimage* 176, 465–476. doi:10.1016/j.neuroimage.2018.04.034.
- Quek, G.L., Nemrodov, D., Rossion, B., Liu-Shuang, J., 2018b. Selective Attention to Faces in a Rapid Stream: hemispheric Differences in Enhancement and Suppression of Category-selective Neural Activity. *J. Cogn. Neurosci.* 30, 393–410. doi:10.1162/jocn\_a.01220.
- Ramon, M., Vizioli, L., Liu-Shuang, J., Rossion, B., 2015. Neural microgenesis of personally familiar face recognition. *Proc. Natl. Acad. Sci.* 112, E4835–E4844. doi:10.1073/pnas.1414929112.
- Retter, T.L., Rossion, B., 2016. Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia* 91, 9–28. doi:10.1016/j.neuropsychologia.2016.07.028.
- Retter, T.L., Jiang, F., Webster, M.A., Rossion, B., 2020. All-or-none face categorization in the human brain. *Neuroimage* 213, 116685. doi:10.1016/j.neuroimage.2020.116685.
- Rhodes, G., Tan, S., Brake, S., Taylor, K., 1989. Expertise and configural coding in face recognition. *Br. J. Psychol.* 80 (Pt 3), 313–331. doi:10.1111/j.2044-8295.1989.tb02323.x.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P., 1976. Basic objects in natural categories. *Cogn. Psychol.* 8, 382–439. doi:10.1016/0010-0285(76)90013-X.
- Rossion, B., Dricot, L., Goebel, R., Busigny, T., 2011. Holistic Face Categorization in Higher Order Visual Areas of the Normal and Prosopagnosic Brain: toward a Non-Hierarchical View of Face Perception. *Front. Hum. Neurosci.* 4. doi:10.3389/fnhum.2010.00225.
- Rossion, B., Torfs, K., Jacques, C., Liu-Shuang, J., 2015. Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *J. Vis.* 15. doi:10.1167/15.1.18, 15.11.18.
- Ruiz-Soler, M., Beltran, F.S., 2006. Face perception: an integrative review of the role of spatial frequencies. *Psychol. Res.* 70, 273–292.
- Russell, R., Duchaine, B., Nakayama, K., 2009. Super-recognizers: people with extraordinary face recognition ability. *Psychon. Bull. Rev.* 16, 252–257.
- Schyns, P.G., Bonnar, L., Gosselin, F., 2002. Show me the features! Understanding recognition from the use of visual information. *Psychol. Sci.* 13, 402–409. doi:10.1111/1467-9280.00472.
- Sergent, J. in *Aspects of Face Processing* (eds Hadyn D. Ellis, Malcolm A. Jeeves, Freda Newcombe, & Andy Young) 17–33 (Springer Netherlands, 1986).
- Tanaka, J.W., 2001. The entry point of face recognition: evidence for face expertise. *J. Exp. Psychol. Gen.* 130, 534.
- Tanskanen, T., N s nen, R., Ojanp a, H., Hari, R., 2007. Face recognition and cortical responses: effect of stimulus duration. *Neuroimage* 35, 1636–1644. doi:10.1016/j.neuroimage.2007.01.023.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381, 520–522.
- VanRullen, R., 2011. Four Common Conceptual Fallacies in Mapping the Time Course of Recognition. *Front. Psychol.* 2. doi:10.3389/fpsyg.2011.00365.
- Yan, X., Rossion, B., 2020. A robust neural familiar face recognition response in a dynamic (periodic) stream of unfamiliar faces. *Cortex* 132, 281–295. doi:10.1016/j.cortex.2020.08.016.
- Young, A.W., Bruce, V., 2011. Understanding person perception. *Br. J. Psychol.* 102, 959–974. doi:10.1111/j.2044-8295.2011.02045.x.