1

2

All-or-none visual categorization in the human brain

4

Talia L. Retter[1,2], Fang Jiang[2], Michael A. Webster[2] & Bruno Rossion[1,3,4]

6

1 Psychological Sciences Research Institute, Institute of Neuroscience,

University of Louvain, Belgium

2 Department of Psychology, Center for Integrative Neuroscience,

University of Nevada, Reno, USA

3 Université de Lorraine, CNRS, CRAN - UMR 7039, F-54000 Nancy, France

4 Université de Lorraine, CHRU-Nancy, Service de Neurologie, F-54000, France

13

Corresponding author:

Talia Retter

1664 N Virginia St.

Mail Stop 296

Reno, Nevada 89557

+1 (775) 303-3472

tlretter@nevada.unr.edu

1

Abstract

Whether visual categorization, i.e., specific responses to a certain class of visual events across a wide range of exemplars, is graded or all-or-none in the human brain is largely unknown. We address this issue with an original frequency-sweep paradigm probing the evolution of responses between the minimum and optimal presentation times required to elicit both neural and behavioral face categorization responses. In a first experiment, widely variable natural images of nonface objects are progressively swept from 120 to 3 Hz (8.33 to 333 ms duration) in rapid serial visual presentation sequences; variable face exemplars appear every 1 s, enabling an implicit frequency-tagged face-categorization electroencephalographic (EEG) response at 1 Hz. In a second experiment, faces appear non-periodically throughout such sequences at fixed presentation rates, while participants explicitly categorize faces. Face-categorization activity emerges with stimulus durations as brief as 17 ms for both neural and behavioral measures (17 – 83 ms across individual participants neurally; 33 ms at the group level). The face-categorization response amplitude increases until 83 ms stimulus duration (12 Hz), implying graded categorization responses. However, a strong correlation with behavioral accuracy suggests instead that dilution from missed categorizations, rather than a decreased response to each face stimulus, may be responsible. This is supported in the second experiment by the absence of neural responses to behaviorally uncategorized faces, and equivalent amplitudes of isolated neural responses to only behaviorally categorized faces across presentation rates, consistent with the otherwise stable spatio-temporal signatures of face-categorization responses in both experiments. Overall, these observations provide original evidence that visual categorization of faces, while being widely variable across human observers, occurs in an all-or-none fashion in the human brain.

2

1 **Introduction**

2 A fundamental function of nervous systems is to organize the flurry of sensory inputs
3 generated by their rich, dynamic, and ambiguous environment. This organization requires
4 generating distinct responses to different sensory inputs (i.e., discrimination), but also
5 generating a common response to different but related sensory inputs (i.e., generalization).
6 This categorization function plays a fundamental role in responding adaptively to the
7 environment, and has been considered to be the foundation of many other cognitive
8 functions, allowing us to learn, memorize, act, and communicate through language, gesture,
9 and expression (Smith & Medin, 1981; Edelman, 1987; Murphy, 2002; Goldstone, Kersten &
10 Carvalho, 2018).

11 An outstanding issue is whether neural categorization of currently experienced stimuli
12 (i.e., perceptual categorization) is graded or all-or-none. That is, does our brain progressively
13 build categorization responses, or does categorization emerge all at once from an
14 accumulation of non-categorical sensory information? This issue is important insofar as
15 perceptual categorization relates to our subjective experience of the stimulus, or perceptual
16 awareness (e.g., del Cul et al., 2007; Quiroga et al., 2008; de Gardelle et al., 2011;
17 Bachmann, 2013; Sekar et al., 2013; Navajas et al., 2014; Windey et al., 2015).
18 Understanding whether neural categorization is all-or-none may constrain psychological,
19 philosophical, and computational constructs of conscious awareness, as well as advance
20 theoretical models of brain function.

21 Studies that have addressed this issue have generally relied on vision, the dominant
22 modality in humans and other primates, and used neural measures recorded during variable
23 stimulus viewing conditions. On one hand, a number of studies have found that neural
24 activity in visual cortical areas increases in an analog, continuous way with increases in
25 stimulus visibility (Kovács, Vogels & Orban, 1995; Vanni et al, 1996; Grill-Spector et al,

3

1   2000; Keysers et al, 2001; Bar et al., 2001; Moutoussis & Zeki, 2002; Bacon-Macé et al,

2   2005; Christensen et al., 2006), supporting a graded or progressive visual categorization

3   process. On the other hand, other studies have attempted to relate neural activity to explicit

4   behavioral reports and rather concluded in favor of all-or-none brain responses related to

5   perceptual awareness, more like discrete "hits" and "misses" (Del Cul et al., 2007; Quiroga et

6   al., 2008; Harris, Wu & Woldorff, 2011; Shafto & Pitts, 2015). The abundance of evidence

7   on both sides has been difficult to reconcile, largely hindering rather than inspiring a coherent

8   theoretical framework.

9   To address this fundamental yet unresolved issue in cognitive neuroscience (Windey

10   & Cleeremans, 2015), we introduce an approach that goes beyond previous efforts to address

11   this question in a number of important ways. First, going beyond previous studies that

12   compared to detection or identification of a limited set of specific stimuli without exemplar

13   generalization, we truly measure neural *categorization*, i.e., responses that are specific to a

14   certain class of stimuli across a wide range of variable exemplars. To do that, we use a large

15   set of widely variable natural images of faces. Faces are used as the visual stimuli to

16   categorize for a number of reasons. From the first minutes of life, the face is arguably the

17   most important visual stimulus for human ecology. Faces are complex multidimensional

18   stimuli that are ubiquitous in our visual environment and drive many of our behaviors. Their

19   initial categorization as "a face" (i.e., generic face categorization) unfolds into an extremely

20   rich social categorization of the individual, allowing to categorize people according to their

21   gender, ethnicity, emotional expression, age, attractiveness, and identity. In neurotypical

22   human adults, generic face categorization evokes extensive neural activity along the (ventral)

23   occipital temporal cortex (e.g., Sergent, Ohta & MacDonald, 1992; Allison et al., 1994; Puce

24   et al., 1995; Kanwisher, McDermott & Chun, 1997; Weiner & Grill-Spector, 2010; Zhen et

25   al., 2015; Jonas et al., 2016; Jacques et al., 2016; Gao et al., 2018). As shown with face

4

1   Pareidolia, generic face categorization in humans goes well beyond the identification of well-

2   defined objective physical features of real faces (Caharel et al., 2013; Omer et al., 2019),

3   providing an advantage over artificial systems in terms of generalization within natural and

4   degraded views and environments (Scheirer et al., 2014).

5   Second, we relate human's (implicit) neural and (explicit) behavioral responses always under

6   the same stimulation constraints, each face exemplar being immediately preceded and

7   followed by a nonface object, i.e. visually masked (Helmholtz, 1867; Enns & Di Lollo,

8   2000). Compared to studies that have used a backward- masking approach to probe

9   perceptual awareness (e.g., Del Cul et al., 2007; Quiroga et al., 2008; Sekar, Findley & Linas,

10   2012; Sekar et al., 2013), we use both backward- and forward-masking. Most importantly,

11   our stimuli are presented in a dynamic visual stream, i.e., variable faces embedded in a rapid

12   train of variable nonface object images for about 1 minute (Figure 1), similarly to rapid serial

13   visual presentation (RSVP) sequences (Potter & Levy, 1969; Keysers et al., 2001; 2005;

14   Potter et al., 2014). This paradigm, combined with neural frequency-tagging (Adrian &

15   Matthews, 1934; Regan, 1989; Norcia et al., 2015) is known to simultaneously measure the

16   two key aspects of generic face categorization, namely *discrimination* by measuring selective

17   responses to face stimuli embedded in rapid trains of non-face objects, and *generalization* by

18   measuring the common selective responses to widely variable exemplars of human faces

19   (Rossion et al., 2015). The neural face categorization responses obtained are devoid of low-

20   level confounds (i.e., amplitude spectrum; Rossion et al., 2015; Gao et al., 2018), and

21   associated with robust neural activity in face-selective regions of the ventral occipito-

22   temporal cortex as shown both with human intracerebral recordings (Jonas et al., 2016) and

23   functional magnetic resonance imaging (Gao et al., 2018).

24   Here, to probe the all-or-none or graded nature of face categorization, we increasingly

25   sweep stimulus duration in this paradigm, from 120 Hz stimulation rate (i.e., 8.33 ms SOA)

5

1   to 3 Hz (333 ms SOA). We then track the onset, optimum and interindividual variability of

2   the neural face categorization response. Importantly, in this case categorization is measured

3   implicitly, i.e., without requiring explicit categorization judgments from the observers,

4   another significant difference with previous studies. However, we also collect explicit

5   behavioral measures at different times under the same stimulation conditions, and relate the

6   two measures on an individual basis.

7   As we document below, our observations indicate that while general visual responses

8   can be tagged at presentation rates up to 120 Hz over the medial occipital cortex (Herrmann,

9   2001), face categorization emerges significantly as early as 60 Hz (17 ms SOA/duration) for

10   a few trials, then reliably at the group level at 30 Hz (i.e., 33 ms duration in a train of nonface

11   objects) to saturate at 12 Hz (83 ms duration) over occipito-temporal regions. However, there

12   is substantial interindividual variability in the presentation rates for face categorization

13   responses (emerging between 17 and 83 ms for different observers), and importantly, these

14   differences are correlated across behavioral and neural measures. Crucially, our findings

15   reveal that the apparent graded increase in the neural amplitudes of face-categorization

16   responses in fact reflects an all-or-none response occurring occasionally to faces at extremely

17   short viewing times and becoming more consistent, but not more evolved, as presentation

18   duration increases. The tight relationship with explicit behavior indicates that this all-or-none

19   face categorization response corresponds to perceptual awareness.

20   **<u>Materials and Methods</u>**

21   **Participants**

22   Sixteen adult participants at the University of Nevada, Reno took part in the experiment All

23   reported normal or corrected-to-normal visual acuity. Their mean age was 29 years (SE =

24   1.65; range = 22 to 44), ten identified as female and six as male, and 15 were right-handed.

25   The study was conducted in accordance with the Code of Ethics of the World Medical

1   Association (Declaration of Helsinki), with protocols approved by the University of Nevada

2   Institutional Review Board.

3   **Stimuli and Display**

4   Stimuli were 112 natural face images and 200 natural object images compiled internally (for

5   examples, see **Movie 1;** Quek et al., 2018; Gao et al., 2018). Facial identities were selected to

6   represent a broad range of characteristics (e.g., age, sex, expression, head orientation). Both

7   face and object images were edited from photographs taken under variable environmental

8   conditions and perspectives (in lighting, background, orientation, etc.). Each image was

9   cropped to a square with the target face or object centered or off-center, resized to 200 x 200

10   pixels, and saved in JPEG format. Stimulus variations were not further controlled because the

11   high amount of variability within both face and object categories reduces the diagnosticity of

12   these cues to any one category (Thorpe et al., 1996; Crouzet et al., 2010; Foldiak et al., 2004;

13   Rossion et al., 2015). The stimuli were presented on a cathode ray tube (CRT) monitor (NEC

14   AccuSync 120) with a 120 Hz screen refresh rate and a resolution of 800 x 600 pixels. At the

15   viewing distance of 80 cm, stimuli subtended 7.3° by 7.3° degrees of visual angle.

16   **Experiment 1 (EEG Frequency-Tagging) Procedure**

17   After the setup of the EEG system, participants were positioned in front of the monitor, with

18   access to a keyboard with which to start stimulation trials and give responses. The only

19   illumination in the room came from the testing and acquisition monitors. Participants were

20   informed that they would see series of natural images presented at a high frequency rate,

21   which would reduce incrementally throughout each trial. They were not given any

22   information about the types of images they would see. Instead, they were instructed to fixate

23   on a centrally presented fixation cross throughout the experiment, and to press on the space

24   bar when the cross briefly changed luminance, from black to white, 12 times per trial.

7

1   Each trial consisted of: 1) the cross to establish fixation lasting between 2-4 s; 2) a 63-

2   s sequence, consisting of nine, contiguous 7-s steps of stimulus presentation at 120, 60, 40,

3   30, 24, 20, 12, 6, and then 3 Hz; and finally, the fixation cross for another 2-4 s. Images were

4   presented continuously with a 100% duty cycle at maximum contrast (Retter et al., 2018).

5   Frequency rates followed integer multiples of the 120 Hz monitor frame rate, i.e. with 1

6   frame at 120 Hz (8.3 ms), 2 frames at 60 Hz (17 ms), 3 frames at 40 Hz (25 ms), etc.

7   Throughout each trial, image size varied randomly at each presentation between 95-105% of

8   the original in five steps.

9   Crucially, regardless of the frequency of image presentation throughout the sequence,

10   faces were embedded consistently at a rate of 1 Hz. Thus, for example, at 120 Hz, faces

11   appeared as every one out of 120 images, while at 6 Hz faces appeared as every one out of 6

12   images. The ratio of face to object images has been shown not to have an impact on the face-

13   categorization response, when as here the faces are separated by at least 400 ms (Retter &

14   Rossion, 2016a). The order of face stimuli and object stimuli was randomized for each

15   subject and trial repetition. Participants binocularly viewed 18 trial repetitions, leading to a

16   total testing time of about 30 minutes.

17   **Experiment 2 (Behavior and ERP) Procedure**

18   After Experiment 1, participants were asked two questions: 1) "What did you notice about the

19   images you just saw?"; and 2) "Did you notice any periodicity within the sequences?" In

20   response to the first, open-ended question, participants typically reported seeing a lot of

21   images, the rate of presentation decreasing, and that they noticed faces (14 participants),

22   animals (10 participants), non-living objects (8 participants), and plants/natural scenes (5

23   participants). Some participants reported one or a few specific images reoccurring: e.g., a

24   dog, elephant, clock, elderly woman, or car. Importantly, in response to the second question,

25   no participants reported noticing any periodicity of faces within the sequences.

8

1   Participants were then debriefed in regards to the first experiment, including being

2   informed that faces had appeared every 1 s throughout all of the sequences. They were then

3   given instructions for the second experiment, which instead involved an explicit face

4   detection task. Participants were asked to respond to human faces, which would appear non-

5   periodically, or not at all, within shorter (30 s) sequences, each at a single frequency

6   throughout, by pressing on the key "J" with the index finger of their dominant hand. The

7   faces appeared 5-7 times in five sequences per condition, and 0 times in one sequence per

8   condition. This resulted in a total of about 30 face presentations for each of the nine

9   frequency conditions (M = 30.4; SD = 2.89), for a total duration of about 40 minutes.

10   **EEG Acquisition**

11   EEG acquisition and analysis modeled protocols established in previous studies (e.g.,

12   Jacques, Retter & Rossion, 2016; Retter & Rossion, 2016a; Retter et al., 2018). A BioSemi

13   ActiveTwo system (BioSemi B.V., Amsterdam, The Netherlands) with 128 Ag-AgCl Active-

14   electrodes was used to record the EEG. Electrodes were organized in the default BioSemi

15   configuration, centered around nine standard 10/20 locations on the cardinal axes (for exact

16   position coordinates, see http://www.biosemi.com/headcap.htm). For ease of comparison

17   across studies, the default electrode labels were updated to closely match the conventional

18   10/5 system (for exact relabeling, see Rossion et al., 2015, Figure S2; Oostenveld &

19   Praamstra, 2001). Additionally, vertical and horizontal electrooculogram (EOG) were

20   recorded with four flat-type Active-electrodes: two above and below the right eye and two

21   lateral to the external canthi. The offset of all electrodes was set below 50 mV, relative to the

22   additional common mode sense (CMS) and driven right leg (DRL) electrode loop. The EEG

23   and EOG were digitized at a sampling rate of 2048 Hz and subsequently downsampled

24   offline to 512 Hz.

25   **Analysis**

9

1   Letswave 5, an open source toolbox (http://nocions.webnode.com/letswave) running over

2   MATLAB R2013b (The MathWorks, USA), was used for the data analysis.

3   *Preprocessing*

4   Data were bandpass filtered from 0.05 to 140 Hz with a fourth-order zero-phase

5   Butterworth filter. Sequences were coarsely segmented from 2 s before to 65 s after sequence

6   onset. To correct for blinks, an independent component analysis was applied for the six

7   participants who blinked more than 0.2 times/s on average throughout the testing sequences

8   (M = 1.20 blinks/s; SD = 0.22; as in, e.g., Retter & Rossion, 2016a). Channels with abrupt

9   deflections of $\pm$ 100 µV or above in more than one epoch were linearly interpolated with the

10 neighboring 3-4 channels (M = 0.75 channels; SE = 0.86 channels; range = 0-3 channels, i.e.,

11 up to no more than 3% of channels).

12 Data were re-referenced to the average of all 128 EEG channels. Each 7-s

13 presentation duration step was re-segmented in two ways: 1) beginning from -200 ms before

14 the first face stimulus onset, to capture face-categorization responses; and; 2) beginning

15 exactly at face onset, in order to isolate the stimulus-presentation response to the full 7 s of

16 each presentation duration step.

17 *Regions-of-interest*

18 To determine significance of the face-categorization responses, a ten-channel bilateral

19 occipito-temporal (OT) ROI was defined a priori based on previous studies with the generic

20 face categorization paradigm (e.g., Retter & Rossion, 2016b; Retter et al., 2018; see also

21 Dzhelyova & Rossion, 2014). To decompose the responses further, the average amplitude of

22 the right and left OT regions were examined separately (right: channels P10; P8; PO8; PO10;

23 PO12; and left: P9; P7; PO7; PO9; PO11). The bilateral ROI was verified to capture face-

24 categorization responses post-hoc (see Results). To provide a region-free assessment of

25 response significance, we also investigated responses across the average of all 128 EEG

10

1   channels. Finally, to identify and quantify face-categorization responses, we used a medial-

2   occipital (MO) region consisting of the 10 following channels: O2; POI2; I2; Iz; OIz; Oz;

3   POOz; O1; POI1; I1), as well as the 128-channel grand average.

4   *Frequency Domain Analysis*

5   The segmented data were averaged by presentation duration step in the time domain,

6   reducing non-phase locked (i.e., non-stimulus related) activity and thereby increasing the

7   signal-to-noise ratio. A fast Fourier transform (FFT) was applied to represent the data of each

8   channel in the temporal frequency domain, i.e., as a normalized amplitude spectrum (µV),

9   with a range of 0 to 256 Hz and a resolution of 0.14 Hz. To correct for variations in baseline

10   noise level around each frequency of interest (i.e., 1 Hz and its harmonics for face-

11   categorization responses, and the stimulus presentation rate and its harmonics), the average

12   amplitude of the neighboring six frequency bins were averaged (a range of 0.86 Hz) and

13   subtracted from each frequency bin. The harmonics of the face-categorization responses were

14   subsequently summed from the fundamental, 1 Hz, up to 20 Hz (similarly to Retter &

15   Rossion, 2016a; Retter et al., 2018), excluding harmonics that coincided with the stimulus

16   presentation frequency when present. The harmonics of stimulus-presentation responses were

17   summed from the fundamental up to 120 Hz for all presentation duration steps.

18   To assess significance of the face-categorization and stimulus-presentation responses

19   at each presentation duration condition at the group level, a Z-Score was computed on the

20   average of the bilateral OT ROI channels for face-categorization responses, and the average

21   of the MO ROI channels for the stimulus-presentation response (using the same baseline

22   frequency range as for the baseline-subtraction). Additionally, the same approach was used at

23   the individual participant level. Responses were considered to be significant when the Z-

24   Score exceeded 2.32 (p<.01, 1-tailed, i.e., predicting the signal was greater than the noise).

11

1   To describe the data further, the amplitude and scalp topographies were assessed

2   across presentation conditions. Scalp topographies were visualized in terms of both their

3   original amplitude and their normalized amplitude (McCarthy & Wood, 1985). To assess the

4   extent of right lateralization for face categorization, a lateralization index was computed

5   using the amplitudes of the right and left OT ROIs (R and L, respectively) as follows: (R-

6   L)/(R+L)*100 (as in Retter & Rossion, 2016a). Statistical tests on response amplitudes were

7   applied independently for face-categorization and stimulus-presentation responses in the form

8   of repeated measures analysis-of-variance tests (ANOVAs), with factors of *Region* and

9   *Condition*. A Greenhouse-Geisser correction was applied to the degrees of freedom in the

10   event that Mauchly's test of sphericity was significant. Data were grand-averaged across

11   participants for display purposes.

12   *Time Domain Analysis*

13   Segmented data were filtered more conservatively with a fourth-order, zero-phase

14   Butterworth low-pass filter at 30 Hz. A notch filter was applied to remove the periodic

15   responses to stimulus-presentation for each presentation condition at its fundamental and

16   harmonic frequencies up to 30 Hz. Data were then re-segmented as described previously for

17   face-categorization responses and averaged by condition. A baseline correction of voltage

18   offset was applied on the 200 ms immediately preceding face stimulus onset.

19   To test for statistical significance of the response from the baseline reference interval

20   (-200 to 0 ms) at each time point, bootstrapping was performed with a significance criterion

21   of p<.001, with 10,000 iterations. To reduce the likelihood of false positives, only groups of

22   at least 5 consecutive time bins (i.e., about 10 ms) were reported. Grand-averaging was

23   computed for display. Finally, to describe a face-categorization response independent of the

24   stimulus presentation duration, the data were averaged across all presentation-duration

25   conditions yielding significant frequency-domain face-categorization responses.

1   *Behavioral Face-Categorization Analysis (Experiment 2)*

2   Face-categorization responses were considered correct when occurring between 0.15

3   – 2 s after face presentation onset; responses occurring outside this range were considered

4   false positives. This criterion rejected 3.5% of potentially correct face-categorization

5   responses, and generated a response time distribution with a similar mean and median: 509

6   (SE = 117.5 ms) and 512 ms, respectively. False positives occurred on average at a rate of

7   0.24 per minute (SE = 0.087); they occurred maximally in a mid-frequency range, from 20 –

8   40 Hz, although this may be for several reasons (e.g., that participants found the task too

9   difficult to try at the highest frequencies, or that they actually did perceive faces sometimes in

10   the stream of non-face objects (pareidolia); **Supplemental Figure 3**). Note that this generous

11   window for correct responses was chosen in terms of the minimum and maximal expected

12   response times for any participant for any key press (the minimal time between two

13   consecutive face images was 2.6 s); however, application of a more conservative range from

14   0.2 – 1 s after face presentation only minimally affected the pattern of results across

15   participants or frequencies (e.g., **Supplemental Figure 3**, for the effect on false positives).

16   The total percent accuracy was defined as the percent of correct detection responses

17   minus the percent of false positives, relative to the total number of faces presented. This

18   modification of the raw accuracy guarded against the chance that a high rate of correct

19   responses occurred as a byproduct of a high rate of false positives; this was particularly

20   important given the differences in false alarms across presentation rates. Response time (RT)

21   was considered only for correct responses.

22   Additionally, inverse efficiency (IE) was computed by dividing RT by total percent

23   accuracy for each participant (Townsend & Ashby, 1983). This IE measure, combining

24   accuracy and RT, was used to rank participants in their face-categorization ability overall, by

25   using the average of trials from all presentation conditions together. Note that IE is thus used

1   only to rank participants' face-categorization *ability*, while in all other analyses accuracy is

2   used in order to examine *whether or not* faces were categorized, irrespective of the RT. The

3   relationship between behavioral face-detection ability and EEG responses was tested with

4   Pearson correlations between IE and EEG response amplitude. Differences in correlation

5   coefficient strength were calculated according to the asymptotic z-test of Lee and Preacher

6   (2013).

7   *Behavioral Fixation-Detection Analysis (Experiment 1)*

8   Responses to the detection of the fixation cross luminance change were analyzed with

9   the same parameters as for the face-categorization responses. Overall, participants' accuracy

10   was high for this easy task across all frequencies (93.4%; SE = 1.39%), with a mean correct

11   RT of 480.0 ms (SE = 16.27 ms). The rate of false positives was low: 0.15 per minute (SE =

12   0.038). The accuracy was higher at the highest frequencies (above 96% at 30 Hz and above;

13   relative to 91 – 93% at all lower frequencies), although this could be either because the

14   luminance changes of the cross were easier to see against a more rapid (i.e., less well-

15   processed) image stream, or as an order effect, with the higher frequencies consistently

16   appearing first in the sequences. Overall, the fixation cross task results differed greatly from

17   those of the face categorization task, emphasizing the differences in these tasks and

18   processes.

19   **Results**

20   The emergence of face-categorization responses (Section 1) was first determined at a neural

21   level from the EEG responses to periodic faces in Experiment 1 in the frequency domain, and

22   at a behavioral level from the face-categorization response accuracy to non-periodic faces in

23   Experiment 2. Then, these neural and behavioral face-categorization responses were

24   quantified across presentation rates from 120 to 3 Hz (Section 2). In Section 3, the

25   relationship between neural and behavioral face-categorization responses is characterized

14

1   across presentation rates, and explained in terms of neural responses (event-related potentials)

2   to behaviorally categorized vs. uncategorized faces in Experiment 2. A qualitative assessment

3   of face-categorization responses across presentation rates is presented in Section 4. Finally,

4   individual differences in generic face categorization are investigated in Section 5.

5   **1. Emergence of Face-Categorization Responses**

6   *1.1 Neural Responses (Experiment 1)*

7   At the individual participant level, significant summed-harmonic OT face-categorization

8   responses emerged between a range of 60 to 23 Hz (17 to 83 ms; **Figure 1**). On the high end,

9   a single participant had a significant response at 60 Hz, and two participants followed at 40

10 Hz. On the low end, a single participant had a significant response only up to 12 Hz, while

11 three participants had responses up to 20 Hz. The mean across participants was 29.4 Hz (SE
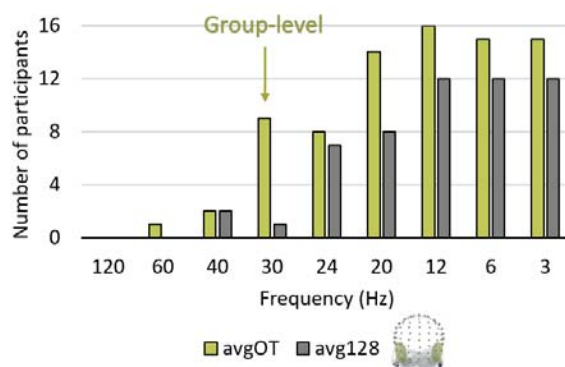
12 = 2.73 Hz), and the median 30.0 Hz.

Figure 1. The number of participants (out of 16 in total) with significant responses at each frequency rate (Z > 1.64; p< .05; Experiment 1). At the group-level, the response was first significant at 30 Hz over the occipito-temporal ROI (Z = 3.5, p =

19 *.0002).* **Key)** *avgOT = average across the left and right occipito-temporal ROIs; avg128 =*

20 *average of all 128 EEG channels.*

21   On average across observers, face-categorization responses appeared to first emerge

22 at 40 Hz (25 ms) over occipito-temporal channels (**Figure 2A**), although the response was

23 not significant until 30 Hz (33 ms) over the OT ROI (**Table 1A;** see also **Table 2A**). When

24 considering response significance from the average of all 128 EEG channels, the highest

25 significant face-categorization response was reduced to 20 Hz (50 ms). In contrast, the

1 stimulus-presentation responses were significant at all frequencies tested at the group level,

2 both at the MO ROI and the average of all 128 channels (**Table 1B**).

Frequency (Hz)	120	60	40	30	24	20	12	6	3
A. Face									
OT	0.11	-0.14	1.12	**3.50**	**5.09**	**6.77**	**10.37**	**8.87**	**6.92**
Avg128	0.25	0.04	0.25	0.55	1.46	**2.37**	**3.02**	**3.23**	**2.38**
B. Stimulus									
MO	**14.73**	**24.33**	**49.19**	**72.08**	**50.22**	**98.41**	**71.87**	**328.60**	**63.68**
Avg128	**18.16**	**18.68**	**22.51**	**53.25**	**30.01**	**74.36**	**47.62**	**144.46**	**35.99**

3 *Table 1. Group-level Z-scores, calculated at the occipito-temporal (OT) ROI for face-*

4 *categorization responses (**A**), the medial-occipital (MO) ROI for stimulus-presentation*

5 *responses (**B**), and the average of all 128 EEG channels for both face- and stimulus-*

6 *presentation responses. Significant responses are shown in bold (Z > 2.32, p<.01).*
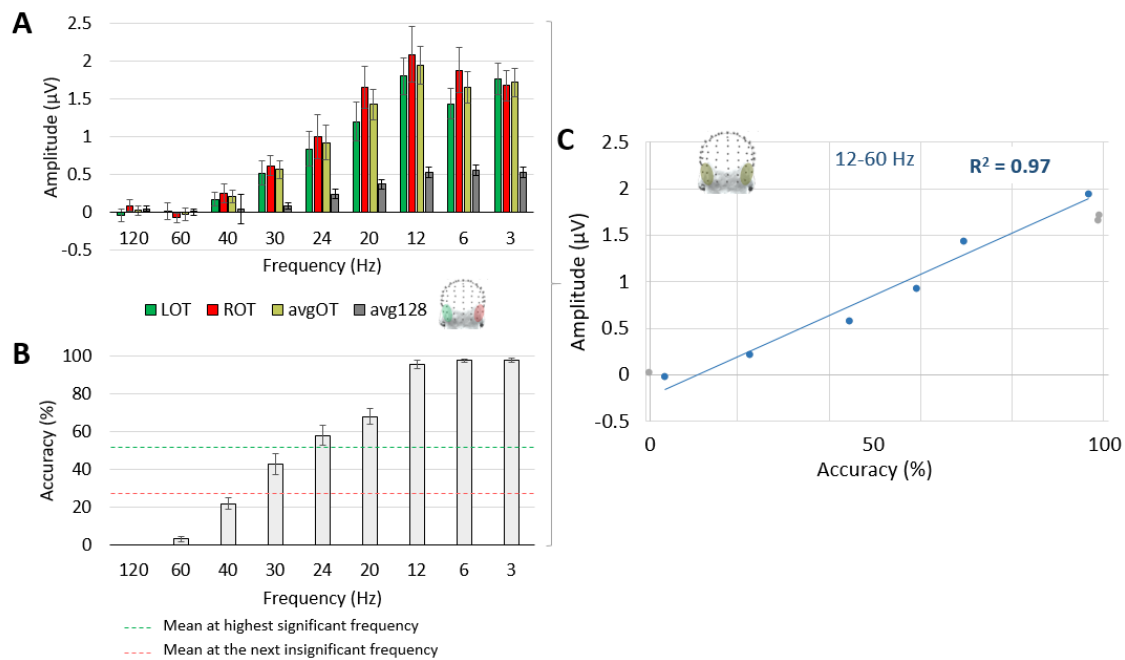
7 ***1.2 Behavioral Responses (Experiment 2)***

8 At the group level (**Figure 2B; Table 3A**), behavioral face-categorization accuracy,

9 corrected for false-positives, emerged at 60 Hz (17 ms; M = 3.03%; SE = 1.26%), $t_{15} = 2.41$,

10 p = .015, one-tailed, d = 0.85. At the higher frequency step, 120 Hz, accuracy did not differ

11 from 0% (-0.40%; SE = 0.50%); at the subsequent frequency step, 40 Hz, accuracy jumped to

12 22% (SE = 3.2%).

13 **2. Quantization of Face-Categorization Responses Across Presentation Rates**

14 ***2.1 Neural Responses (Experiment 1)***

15 The amplitude of face-categorization responses was quantified at four ROIs: over the

16 OT ROI, as well as its divisions into the right occipito-temporal (ROT) and left occipito-

17 temporal (LOT) ROIs, and the average of all 128 channels (**Figure 2A**). Across

18 quantification regions, the response amplitude neared zero at 120 and 60 Hz, and then

19 increased progressively until 12 Hz (83 ms; Figure 3B). After 12 Hz, it decreased but

20 remained significant until the slowest presentation rate, 3 Hz (333 ms; Table 1A). In

1   comparison, the amplitude of the stimulus-presentation response decreased continuously from

2   3 to 120 Hz, but always remained above zero (**Supplemental Figure 1**). The face-

3   categorization response was at least about 3 times larger over the OT than average 128-

4   channel region over all conditions below 60 Hz, indicating a relatively focal response (for

5   stimulus presentation, the MO ROI captured a lesser proportion of the response, on average a

6   little over 2 times larger than the 128-channel average).

7

8   *Figure 2. Behavioral face-categorization accuracy (Experiment 2) compared with neural*

9   *face-categorization response amplitude (Experiment 1). **A**) Baseline-subtracted summed-*

10 *harmonic amplitudes across different ROIs. **B**) Behavioral mean face-categorization*

11 *accuracy across frequency rates. The dashed green line at 52% accuracy represents the*

12 *mean accuracy corresponding to highest frequency yielding a significant face categorization*

13 *response for each participant; the dashed red line at 27% represents the mean accuracy*

14 *corresponding to the frequency one step higher for each participant, at which the face*

15 *categorization response was not significant. **C**) A linear relationship between accuracy*

16 *(Experiment 2) and amplitude (Experiment 1) between 12 and 60 Hz. The data points from*

1   *the remaining presentation rates (3, 6, and 120 Hz) are shown in light gray, but were not*

2   *included in this correlation. **Key)** LOT: left occipito-temporal ROI; ROT: right occipito-*

3   *temporal ROI; avgOT = average of these left and right occipito-temporal ROIs; avg128 =*

4   *average of all 128 EEG channels.*

5   For statistical analyses, the amplitude of face-categorization responses was compared

6   across the OT and MO ROIs (**Table 2A**). There was a significant main effect of *Region*, $F_{1,15}$

7    $= 19.5$, p $< .001$, $\eta_p^2 = 0.57$, reflecting the larger amplitude for the OT (M $= 0.94$, SE $= 0.12$)

8   than MO (M $= 0.56$, SE $= 0.06$) ROI in response to faces. There was also a main effect of

9   *Frequency,* $F_{3.3,49} = 5.94$, p $= .010$, $\eta_p^2 = 0.86$, with the low frequencies (12-3 Hz) being the

10   highest and peaking at 12 Hz (M $= 1.50$, SE $= 0.20$), and then declining sharply as the

11   frequencies continued to increase, with almost no amplitude at 120 and 60 Hz (M $\leq .02$ µV).

12   The interaction between these two factors, $F_{3.6,54} = 10.2$, p $= .002$, $\eta_p^2 = 0.91$, was significant:

13   most remarkably, there was no difference in amplitude across regions from 120-40 Hz (all

14   differences < 0.02 µV), but a separation at 30 Hz and all lower frequencies (all differences $>$

15   0.32 µV).

Frequency (Hz)	120	60	40	30	24	20	12	6	3
A. Face (µV)									
OT	0.03	-0.03	0.21	0.57	0.93	1.43	1.95	1.66	1.72
	(0.06)	(0.08)	(0.08)	(0.12)	(0.23)	(0.21)	(0.25)	(0.21)	(0.19)
MO	0.008	0.07	0.22	0.26	0.47	0.78	1.06	1.05	1.10
	(0.055)	(0.07)	(0.09)	(0.08)	(0.09)	(0.13)	(0.18)	(0.12)	(0.18)
B. Stimulus (µV)									
OT	0.04	0.05	0.13	0.22	0.24	0.35	0.59	1.62	2.40
	(0.005)	(0.006)	(0.01)	(0.02)	(0.03)	(0.05)	(0.06)	(0.13)	(0.18)
MO	0.03	0.06	0.16	0.33	0.45	0.64	1.11	2.05	3.35
	(0.004)	(0.008)	(0.01)	(0.03)	(0.04)	(0.08)	(0.12)	(0.19)	(0.27)

16   **Table 2.** *Mean response amplitude (µV), calculated at the occipito-temporal (OT) and*

17   *medial-occipital (MO) ROIs for both face-categorization (**A**) and stimulus-presentation (**B**).*

18   *One standard error, across participants, is included in parentheses.*

18

1   For comparison, the stimulus-presentation responses were also evaluated over the OT

2   and MO ROIs for statistical analyses (**Table 2B**). Here, the response was larger over the MO

3   (M = 0.91, SE = 0.07) rather than OT (M = 0.63, SE = 0.04) ROI, leading to a significant

4   main effect of *Region*, $F_{1,15}$ = 43.8, p < .001, η_p^2 = 0.75. The response also decreased from

5   low to high frequencies, but with a different trend than for the face-selective response

6   (compare Supplemental Figure 1 to Figure 2A). This produced a main effect of *Frequency,*

7     $F_{1.5,22}$ = 21.8, p < .001, η_p^2 = 0.96. The interaction between these two factors, $F_{1.8,26}$ = 21.8, p

8   = .001, η_p^2 = 0.92, was significant, as there was no amplitude difference across regions from

9   120-60 Hz (differences < 0.02 µV), but an increasing difference as the frequency decreased:

10   at 40 Hz up slightly to 0.035 µV, at 30 Hz, 0.11 µV, and at all lower frequencies > .20 µV.

11   *2.2 Behavioral Responses (Experiment 2)*

12   Behavioral face-categorization responses first emerged at 60 Hz, and accuracy increased

13   steadily up to 68% at 20 Hz (50 ms; **Table 3A**). Beyond 20 Hz, accuracy neared ceiling

14   between 12 to 3 Hz (83 to 333 ms; all accuracies above 95.6%). Overall, the differences in

15   accuracy across frequency rates were highly significant, $F_{2.2,33}$ = 217, p<.001, η_p^2 = 0.94.

Frequency (Hz)	120	60	40	30	24	20	12	6	3
A. Accuracy (%)	-0.4 (0.49)	3.0 (1.26)	21.7 (3.17)	42.9 (5.59)	57.9 (5.20)	68.0 (4.32)	95.6 (2.12)	97.8 (0.86)	98.0 (1.01)
B. RT (ms)	1435 (145)	650 (52.7)	557 (19.2)	529 (18.9)	505 (11.0)	503 (15.2)	498 (13.4)	495 (15.3)	499 (15.2)
C. IE	84.7 (260)	74.5 (20.8)	32.0 (4.85)	14.2 (1.91)	10.1 (1.16)	8.11 (0.81)	4.29 (0.26)	5.07 (0.18)	5.11 (0.18)

16   ***Table 3***. *Behavioral face-detection results: percent accuracy (**A**), response time (RT; **B**), and*

17   *inverse efficiency (IE; **C**). One standard error for each measurement is given in parentheses.*

18   Response time (RT) showed less variation across presentation rates (**Table 3B**): from

19   3 to 24 Hz, there were no significant RT differences across conditions, $F_{4,60}$ = 0.37, p = .83,

20   η_p^2 = 0.02, reflecting a stable face detection response at these rates. However, the response

21   was slower at shorter presentation durations, and greatly delayed at 120 Hz (although note

19

1   that only two participants had any (one or two) responses at this frequency). At the next

2   shortest presentation rate, 60 Hz, 10 participants had some correct detection responses,

3   contributing towards significant RT differences across frequency conditions from 3 to 60 Hz,

4    $F_{1.5,12} = 4.70$, p = .040, $\eta_p^2 = 0.37$. These differences persisted when removing 60 Hz, $F_{6,78} =$

5   6.84, p <.001, $\eta_p^2 = 0.35$, in this case driven by slower responses at 40 Hz.

6   Inverse efficiency (IE) was also calculated, as a way to capture an overall measure of

7   participants' performance (**Table 3C**). This measure, as well as response accuracy, will be

8   used in the following section to relate behavioral and EEG face-categorization responses.

9   **3. The Relationship Between EEG and Behavioral Face-Categorization Responses**

10   *3.1 The Correlation of Accuracy (Experiment 2) and Amplitude (Experiment 1)*

11   The face categorization accuracy (Experiment 2) and EEG amplitudes (Experiment 1)

12   seemed to show similar patterns across presentation rates (compare Panels B and C of Figure

13   1). Indeed, plotting the linear relationship between these two factors revealed a very strong

14   correlation across all nine presentation rates, $R^2 = 0.96$, p < .0001, as well as in the range

15   where amplitude consistently decreased, 60-12 Hz, $R^2 = 0.97$, p = .0002 (**Figure 2C**).

16   There are two possible interpretations of this observation: first, that the neural

17   amplitude at higher presentation rates is decreased in response to each face to the same extent

18   that the behavioral accuracy drops; or second, that the neural amplitude is decreased because

19   some faces were not categorized at all, in the same way that the behavioral accuracy reflects

20   the proportion of "hits" and "misses". That is, either graded or all-or-none neural face

21   categorization responses could explain this data. To fulfill our major goal of differentiating

22   these alternative accounts, in the following sections we compare neural responses to

23   behaviorally detected and undetected faces from Experiment 2.

24   *3.2 EEG Responses to Behaviorally Categorized or Uncategorized Faces (Experiment 2)*

20

1   We examined non-periodic EEG responses in the time domain from Experiment 2

2   separately for 1) behaviorally categorized faces and 2) behaviorally uncategorized faces.

3   Given that this experiment had relatively few trials per condition (about 30 on average,

4   containing a mix of categorized and uncategorized faces), we combined responses across a

5   range of presentation rates. Specifically, we used 12 to 60 Hz, in order to include a large

6   number of trials, balanced in terms of the number of categorized (M = 62, SE = 7.1) and

7   uncategorized (M = 60, SE = 5.7) faces across participants (after artifact rejection to remove

8   blink-contaminated trials). Note that this approach is supported by neural face-categorization

9   responses appearing qualitatively similar across presentation rates in Experiment 1 (see

10   Section 4). Note also that these face-categorization responses were confounded with

11   behavioral detection (pre-)motor activity, with response times occurring on average at 535 ms

12   over these presentation rates (SE = 15.1 ms). We thus focused the analyses on the first two

13   face-selective components, the P1-face (onset at about 100 ms, 144-ms peak latency) and N1-

14   face (202-ms peak latency), as defined in Experiment 1 (Section 4.3; and reported in previous

15   studies with this paradigm, see e.g., Rossion et al., 2015; Retter & Rossion, 2016a; Quek &

16   Rossion, 2017). Finally, note that although faces appeared periodically in Experiment 1 and

17   non-periodically in Experiment 2, comparing neural face categorization responses across

18   experiments is valid because such responses are immune to the periodicity of the stimulus

19   (Quek & Rossion, 2017).

20   For behaviorally categorized faces, significant neural responses (event related

21   potentials) were present during the windows of the P1-face and N1-face; for uncategorized

22   faces, little to no responses were apparent, exhibiting no significance (**Figure 3A**).

23   Examination of the P1-face and N1-face peak-time deflections together over the bilateral OT

24   ROI also showed a significant response to behaviorally categorized faces (M = 3.12 µV, SE =

25   0.37 µV), t_{15} = 8.47, p <.001, d = 3.0, but an insignificant response to behaviorally

21

1   uncategorized faces (M = 0.38 µV, SE = 0.30 µV), t_{15} = 1.24, p = .12, d = 0.44 (**Figure 3B**).

2   The topographies of these face-selective components for behaviorally categorized faces were

3   maximal over the OT region; little activation was observed across the scalp to behaviorally
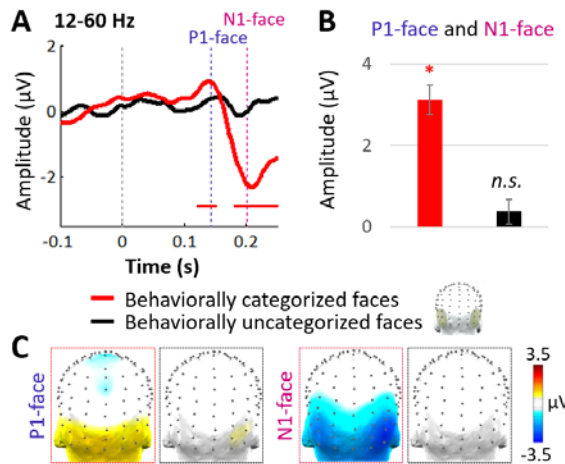
4   uncategorized faces (**Figure 3C**).

Figure 3. Time-domain responses (N = 16) across 12-60 Hz to all behaviorally categorized faces. A) Waveforms over the occipito-temporal ROI, to behaviorally categorized (in red) and uncategorized (in black) faces. Significant time points (p<.001) are indicated below for

12   *categorized faces (in red) and uncategorized faces (none). B) The amplitude across the P1-*

13   *face and N1-face components, as defined in Experiment 1 (see Figure 6A), at the occipito-*

14   *temporal ROI. Responses to behaviorally categorized faces (red) are significantly above*

15   *zero, while those to behaviorally uncategorized faces (black) are not significant. C) Posterior*

16   *topographies at the peak times of the P1-face and N1-face, for behaviorally categorized (red*

17   *outline) and uncategorized (black outline) faces. Component times, taken from Experiment 1,*

18   *are indicated in Panel A.*

19   **3.3 Behavioral Face Categorization Predicts EEG Face-Categorization Amplitude**

20   **(Experiment 2)**

21   All-or-none neural responses, as reported in the previous section, could thus explain the

22   gradient of response amplitudes reported in Experiment 1: this gradient reflects the

23   proportion of absent neural responses to uncategorized faces and full neural responses to

24   categorized faces. To further test this hypothesis, we examined whether the neural face-

1   categorization amplitude would be reduced at higher relative to lower rates in the *same*

2   *proportion* that the behavioral face-categorization accuracy was reduced.

3   To this end, we examined the neural responses to non-periodically presented faces in

4   Experiment 2 in the time domain, for behaviorally categorized and uncategorized faces

5   together. Specifically, we examined neural responses across two relatively high presentation

6   rates (24 and 30 Hz) with fewer correct detections, and compared them to two relatively low

7   presentation rates (12 and 20 Hz) with more correct detections, for nine participants meeting

8   a minimum behavioral accuracy threshold (see Methods). Again, we combined responses

9   across multiple rates given the limits from the number of face presentations in this

10   experiment. As before, we constrained the analysis to the first two face-selective components.

11   The amplitude of the neural responses to behaviorally categorized and uncategorized

12   faces together appeared reduced at 24-30 Hz, relative to 12-20 Hz, at the times of the P1-face

13   and N1-face; in contrast, the response to behaviorally categorized faces only did not appear to

14   differ across presentation rates (**Figure 4A**). This was evident over the bilateral occipito-

15   temporal ROI, which appropriately captured face-categorization responses for both

16   conditions (**Figure 4B**).
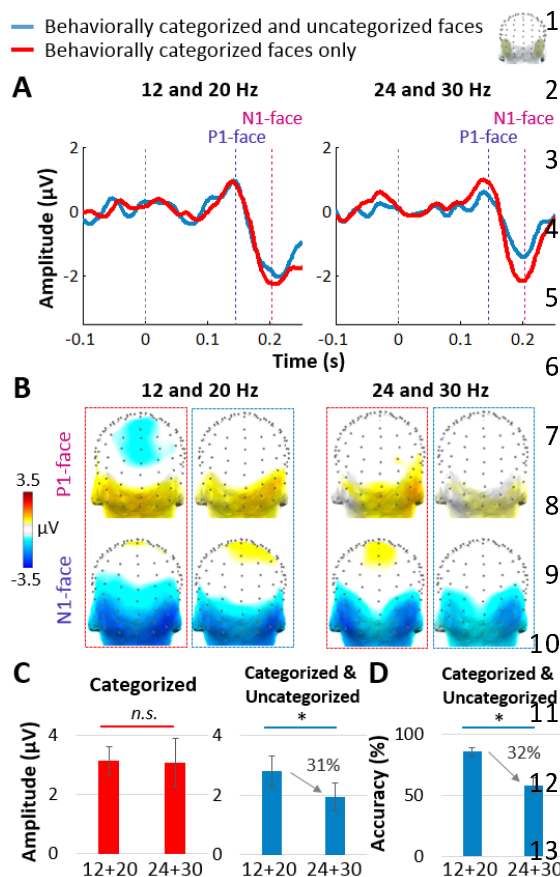
17

18

19

20

21

22

23

24

23

*Figure 4. Time-domain responses from Experiment 2 to behaviorally categorized and uncategorized faces together (blue), compared to categorized faces alone (red). There were fewer categorized faces at the higher (24 and 30 Hz), relative to lower (12 and 20 Hz) frequencies; results are shown for participants who had a minimum of 20 artifact-free face-categorizations at both frequency pairs (N = 9). **A)** Waveforms of the bilateral OT ROI for 12 and 20 Hz (left), and 24 and 30 Hz (right), for both behaviorally categorized faces only (red) and behaviorally categorized and uncategorized faces combined (blue). The indicated times of the P1-face (144 ms) and N1-face (202 ms) components are taken from Experiment 1 (see Figure 6A). **B)** Scalp topographies of these two components; outlines indicating condition are colored as in the previous panels. **C)** The combined amplitude of the P1-face and N1-face components, at the OT ROI. Responses to behaviorally categorized faces alone did not differ across these conditions (left), but responses to behaviorally categorized and uncategorized faces together were significantly larger at 12 and 20 Hz than at 24 and 30 Hz (right). **D)** The percent accuracy dropped from 12 and 20 Hz to 24 and 30 Hz by about 32%; note the correspondence with the 31% decreased EEG amplitude across these conditions in Panel C for to behaviorally categorized and uncategorized faces.*

When behaviorally categorized faces were considered alone, a significant difference in the combined amplitude of the P1-face and N1-face components was not found across 12

1   and 20 Hz, (M = 3.12 µV; SE = 0.47 µV) vs. 24 and 30 Hz (M = 3.38 µV; SE = 0.72 µV), t_8

2   = -0.06, p = .48, d = 0.02, replicating the finding in the previous section (**Figure 4C**, left

3   panel). However, the combined amplitude was significantly larger at 12 and 20 Hz (M = 2.80

4   µV; SE = 0.53 µV) than at 24 and 30 Hz (M = 1.95 µV; SE = 0.48 µV) when behaviorally

5   categorized and uncategorized faces were considered together, t_8 = 2.29, p = .026, d = 0.56

6   (**Figure 4C**, middle panel).

7   　　　Behaviorally, the accuracy was reduced by 32.2% from 12 and 20 Hz to 24 and 30 Hz

8   (from 85.8%, SE = 3.27%, to 58.1%, SE = 6.44%) for these participants, t_8 = 6.35, p = .0001,

9   d = 1.80 (**Figure 4D**). The EEG amplitude across the P1-face and N1-face components was

10   decreased similarly, by 30.6%, from 12 and 20 Hz to 24 and 30 Hz (from 2.80 µV to 1.95

11   µV), when responses to behaviorally categorized and uncategorized faces were combined.

12   This corresponding decrease of behavioral accuracy and amplitude supports the hypothesis

13   that missed face categorizations contribute proportionally to decreased neural response

14   amplitudes at shorter presentation durations.

15   **4. Qualitative Investigation of Face-Categorization Responses**

16   *4.1 Multi-Harmonic Response Distributions (Experiment 1)*

17   In the temporal frequency-domain amplitude spectrum, face-categorization responses were

18   distributed across 1 Hz and its harmonics below 20 Hz (**Figure 5A**). Harmonic responses

19   were generally higher at lower frequencies, and strongest between 1 and 10 Hz, with a local

20   peak typically around 6 Hz. Stimulus-presentation responses, which occurred at higher

21   frequencies, were spread across fewer harmonics (**Figure 5C**). In contrast to face-

22   categorization harmonic responses, they were typically dominated by the amplitude of the

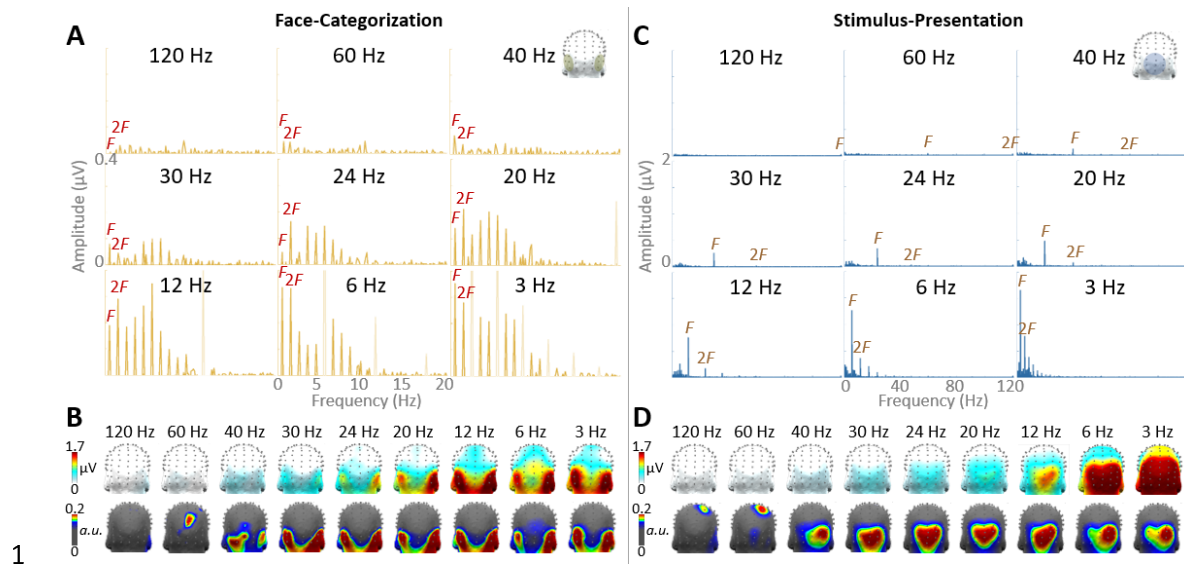23   first harmonic frequency and decayed exponentially as harmonic frequency increased.

25

Figure 5*. Face-categorization responses characterized in the frequency domain, and compared to stimulus-presentation responses. **A)** Face-categorization baseline-subtracted amplitude spectra. Face-selective responses were tagged at F (= 1 Hz) in every condition, with higher harmonics occurring at 2F (2 Hz), 3F (3 Hz), etc. (the first two harmonics are labeled above the spectra). The data are shown over the region of the scalp (bilateral occipito-temporal ROI) and frequency range (up to 20 Hz) that were selected to capture face-selective responses. Higher harmonic frequencies coinciding with stimulus-presentation are drawn in very light yellow, and may surpass the plotted amplitude range. For ease of comparison across conditions, all graphs are plotted with common axes. **B)** Face-categorization response topographies. Top row: grand-averaged baseline-subtracted summed-harmonic topographies. Bottom row: the same responses normalized. **C)** Stimulus-presentation responses are tagged at the frequency of presentation (F) corresponding to the condition name. As in Panel A, the first two harmonic frequencies are labeled above the spectra for each condition. The data is shown over the region of the scalp (medial-occipital ROI) and frequency range (up to 120 Hz) that were used to capture stimulus-presentation responses. All graphs are plotted with common axes, although different from those in Panel*

1   *A. **D**) Stimulus-presentation topographies. Top row: grand-averaged baseline-subtracted*

2   *summed-harmonic topographies. Bottom row: the same responses normalized.*

3   Note that in the previous sections, given the ranges of the distributed harmonic

4   responses, as well as expected ranges from previous studies (e.g., Jacques, Retter & Rossion,

5   2016; Retter & Rossion, 2016a; Retter et al., 2018), baseline-subtracted harmonics were

6   summed from 1 to 20 Hz for face categorization responses (Retter & Rossion, 2016a). For

7   stimulus presentation responses, in order to encapsulate the full range of presentation

8   frequencies used across conditions, baseline-subtracted harmonic responses were summed up

9   to 120 Hz (see Methods). The face-categorization amplitudes across presentation rates were

10   shown in Figure 2A, and can be compared to the corresponding stimulus-presentation

11   amplitudes in Supplemental Figure 1.

12   ***4.2 Response Topographies (Experiment 1)***

13   The a-priori bilateral occipito-temporal (OT) region of interest was verified for its ability

14   represent face-categorization responses. Indeed, the OT ROI represented 7 to 9 out of the 10

15   channels displaying the maximal face-categorization response amplitudes across presentation

16   conditions from 3 to 40 Hz, although only 2 at 60 Hz and 0 at 120 Hz (**Figure 5B**: top row).

17   For further verification, we controlled for differences in response amplitude across

18   presentation duration conditions by normalizing the response topographies: this shows

19   occipito-temporally dominated responses again from about 40 to 3 Hz (**Figure 5B**: bottom

20   row). Note that for response quantification, this region was also broken into the right and left

21   hemispheres to assess potential right lateralization effects generally predicted for face

22   perception.

23   In terms of lateralization, the ROT appeared to consistently have a higher amplitude

24   than the LOT, except at 3 Hz. Indeed, the lateralization index (see Methods) revealed a right

25   lateralization (values > 0) across all significant conditions (30 – 3 Hz), except 3 Hz (M = -

27

1 3.13, SE = 4.33; see Figure 3A for data at each the right and left OT ROI). The right

2 lateralization was typically low, ranging from 2.18 to 16.1 across 30 to 6 Hz; it was maximal

3 at 20 Hz (SE = 11.6). Yet, at the group level, there were no significant differences in

4 lateralization across conditions, $F_{1.4,11} = 1.35$, p = .31, $\eta_p^2 = 0.38$.

5   The responses to stimulus presentation were shown to contrast with the face-

6 categorization responses. These stimulus-presentation responses typically occurred

7 maximally over the medial occipital (MO) scalp region (**Figure 5D**: top row). Their response

8 topographies varied somewhat as the presentation frequency varied, being most different at

9 60 and 120 Hz, where a centro-parietal region dominated the responses (**Figure 5D**: top row).

10 *4.3 Spatio-Temporal Dynamics (Experiment 1)*

11   The face-categorization responses from the previous section were re-examined in the

12 time domain to explore potential qualitative differences across presentation rates. The

13 following analyses use data selectively filtering out the stimulus-presentation frequency (see

14 Methods; for unfiltered data, see **Supplemental Figure 2**). Four prominent time-windows of

15 interest appeared across conditions, corresponding to those reported previously as P1-face,

16 N1-face, P2-face, and P3-face (at a 12.5-Hz stimulus-presentation rate; Retter & Rossion,

17 2016a). These time-windows encompassed nearly all significant deviations over the bilateral

18 OT ROI (**Figure 6A**). Each of these deflections, or "components", was significant across

19 three to six frequency conditions, without any clear trend of component differences as

20 stimulus presentation duration varied. Across significant conditions, these four components

21 peaked over the OT ROI at: 144 ms (SE = 1.95 ms); 202 ms (SE = 5.09 ms); 281 ms (SE =

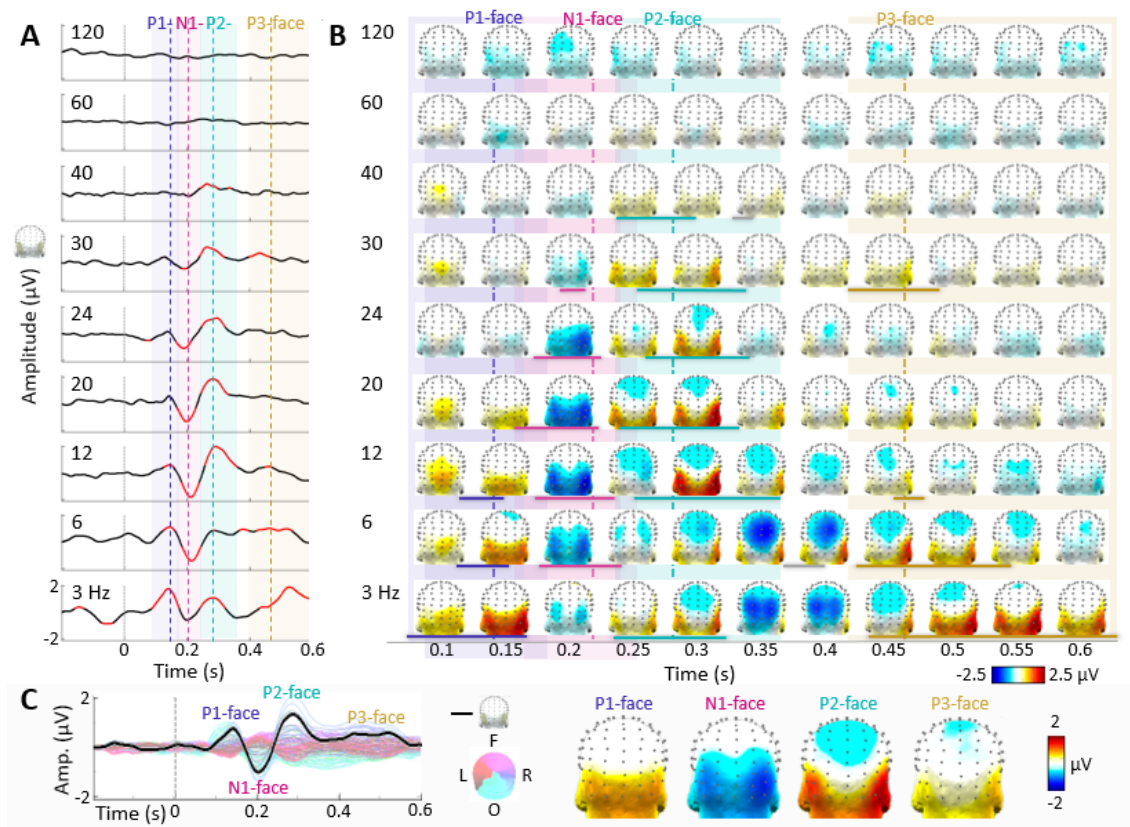22 5.34 ms); and 463 ms (SE = 12.8 ms) post face-stimulus onset.

28

*Figure 6. Time-domain responses to face presentation, with each frequency condition filtered to remove the response at its respective stimulus-presentation frequency. Four components were present repeatedly across conditions: P1-face, N1-face, P2-face, and P3-face, and are indicated by a vertical line showing the average peak time, and shading depicting the range, across significant conditions. **A**) Waveforms of the bilateral OT ROI in response to face stimulus onset (0 s). For each condition, significant time periods (p<.001) are plotted in red, while those insignificant are plotted in black. **B**) Scalp topographies, plotted from 0.1-0.6 s post-stimulus onset, every 0.05 s. Significant response periods are underlined in the color of their respective component for each condition; significant periods not corresponding to one of these four components are indicated in gray. **C**) An average of all conditions with significant face-categorization responses in the frequency domain. The bilateral occipito-temporal ROI is plotted in thick black, superimposed above the data from all 128 EEG channels, colored accordingly to the adjacent 2D scalp topography (F: frontal; R: right; O:*

1   *occipital; L: left). The topographies are plotted to the right at the peaks of the P1-face (142*

2   *ms), N1-face (202 ms), P2-face (286 ms), and P3-face (519 ms) for this waveform.*

3   These deflections, and responses more generally across time, occurred mainly over

4   the bilateral OT ROI, with some coverage of the occipito-inferior, occipito-parietal, and

5   medial-occipital regions (**Figure 6B**). In concurrence with the frequency-domain analysis, no

6   significant deflections were found for the conditions at 120 or 60 Hz; however, one

7   significant time-window was found for 40 Hz at the P2-face time, perhaps contributing to the

8   non-significant amplitude in the frequency domain over the OT region for this condition (see

9   again Figure 4B). To characterize a frequency-independent face-categorization response, an

10   average of all conditions with significant face-categorization responses as determined from

11   the frequency domain (3 – 30 Hz) was created (**Figure 6C**). Over the OT ROI, this response

12   first reached significance at 103 ms post-stimulus onset, and retained significance for 460 ms:

13   it was consistently significant at the times of the P1-face (103 to 159 ms), N1-face (175 to

14   234 ms), P2- and P3-face (247 to 560 ms).

15   In addition to the four main time-windows, a significant time period emerged in the 6

16   Hz condition between 366 – 398 ms, peaking at 380 ms. This 6-Hz OT positive deflection

17   was accompanied by a large medial parieto-occipital negativity, peaking about 15 ms earlier,

18   which was previously described as the "N2-face" (with a 5.88-Hz sinewave presentation;

19   Jacques, Retter & Rossion, 2016). This response signature is also apparent when stimuli are

20   presented at 3 Hz, and perhaps at 12 Hz and 24 Hz, although not reaching significance over

21   the OT scalp ROI in these conditions (see again Figure 6B).

22   ***4.4 Summary***

23   In sum, face-categorization responses were qualitatively similar in terms of their harmonic

24   response distributions in the frequency domain, their summed-harmonic topographies, and

25   their spatio-temporal dynamics.

30

1   **5. Individual Differences in Generic Face Categorization**

2   *5.1 The Diagnosticity of Presentation Rate (Experiments 1 and 2)*

3   Here, we aimed to investigate whether individuals' EEG amplitudes (Experiment 1) were

4   related to their behavioral face-categorization performance (Experiment 2) overall. As a first

5   step, we checked what effect presentation rate had on the range of individual differences,

6   both behaviorally and neurally. Individual participants were ranked for their behavioral face-

7   categorization performance based on their inverse efficiency (IE) score across all presentation

8   rates (range = 0.65 to 1.36; M = 0.97; SE = 0.054; see Table 3C).

9   The best participant at behavioral face detection (S07) was contrasted to the worst

10   (S01), in terms of percent accuracy and EEG amplitude. (Note that the same participants

11   would have been ranked as the best and worst if accuracy alone was used as a metric.) Large

12   differences between these participants in both behavioral and EEG amplitude responses were

13   apparent in the mid-frequency range, from 20 – 30 Hz (50 – 33.3 ms; **Figure 7A**). No

14   differences were apparent across these participants at the lowest three presentation rates (3 –

15   12 Hz, i.e., 333 – 83.3 ms) in either measure, and differences were greatly reduced by 40 and

16   60 Hz (25 and 16.7 ms, respectively) until they were abolished at 120 Hz (8.33 ms). The data

17   of all participants, grouped into quartiles, followed these trends, at least between the best
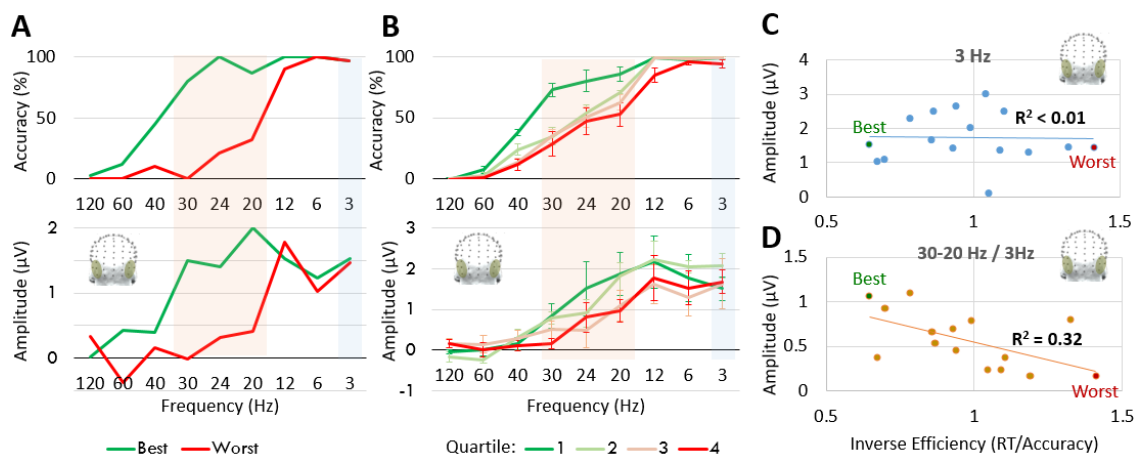
18   25% and worst 25% of participants (**Figure 7B**).

19

1   *Figure 7. Individual differences in behavioral performance related to EEG amplitude for*

2   *face categorization. **A**) Presentation rates diagnostic of individual differences. A comparison*

3   *of the participant with the best vs. worst behavioral (inverse efficiency; IE) responses across*

4   *presentation rates. The highlighted sections indicate data ranges used in panels C and D.*

5   *Above: Accuracy; Below: Summed baseline-subtracted harmonic amplitude over the bilateral*

6   *OT ROI. **B**) As in Part A, except the 16 participants were split into quartiles of four*

7   *participants each, ranked 1st-4th behaviorally (Group 1), 5th-8th (Group 2), 9th-12th (Group 3),*

8   *and 13th-16th (Group 4). **C**) The non-significant correlation of individuals' IE (across all*

9   *presentation rates) with their OT ROI EEG amplitudes at 3 Hz (highlighted in blue in the*

10   *previous two panels). **D**) The significant correlation between individuals' mid-frequency OT*

11   *EEG amplitudes (averaged over 20, 24, and 30 Hz, highlighted in orange in the previous two*

12   *panels), weighted by the amplitude at 3 Hz. In both panels C and D, the data from the*

13   *participants yielding the best and worst behavioral performance are labeled, respectively.*

14   The lack of significant EEG responses in the high presentation rates (40 – 120 Hz)

15   would prevent a relationship between individuals' EEG and behavior at this range. At low

16   frequencies (3 – 12 Hz), behavior near ceiling would prevent a relationship between

17   individuals' EEG and behavior. The best separation in EEG response amplitudes was

18   produced at 20 Hz, and behavioral responses seemed well-separated across the mid-frequency

19   range (from 20 – 30 Hz; 50 – 33.3 ms). In the following section, therefore, we will relate

20   individuals' EEG response amplitudes to their behavioral performance using this mid-

21   frequency range.

22   ***5.2 The Correlation of Individual Participant's EEG Amplitudes (Experiment 1)***

23   ***and Behavioral Accuracy (Experiment 2)***

24   To examine whether there was an inherent relationship between individuals' behavioral face-

25   categorization performance (IE scores, computed across presentation rates) and EEG face-

1   categorization amplitudes, we correlated these measures at the slowest presentation rate, 3 Hz

2   (333 ms). This correlation was not significant, without any apparent trend, R^2 = .0006; p =

3   0.93 (**Figure 7C**), which is likely explained by variable physiological factors influencing the

4   EEG response across participants (see Discussion).

5   When correlating individuals' behavior and amplitudes in the mid-frequency range,

6   from 20 to 30 Hz, as defined in the previous section, the correlation was improved but still

7   not significant, R^2 = 0.22; p = .067. However, when weighting the individuals' mid-

8   frequency range EEG amplitudes by their baseline amplitudes at 3 Hz, a significant

9   correlation was found, R^2 = 0.32, p = .043 (**Figure 9D**). This measurement is reflective of the

10   change in amplitude of the EEG response from its amplitude at a presentation rate with

11   ceiling-level behavioral face-categorization performance, and thus likely reflects the percent

12   change in behavioral performance as presentation rate decreases. Yet, note that the weighted

13   correlation was not significantly stronger than the unweighted correlation, Z = 0.99, p = .16

14   (one-tailed).

15   **<u>Discussion</u>**

16   With a novel sweep frequency-tagging visual categorization paradigm, we (1) determined the

17   minimal and optimal presentation duration necessary to elicit both behavioral and neural

18   face-categorization responses in human observers, (2) related these responses at the group

19   and individual levels, and (3) determined whether visual categorization is graded or all-or-

20   none. In a first experiment, we presented faces periodically within non-face object RSVP

21   sequences, and investigated implicit response in the EEG frequency domain. In a second

22   experiment, we presented faces non-periodically within non-face object RSVP sequences,

23   and investigated explicit behavioral categorization responses, as well as their associated

24   electrophysiological responses.

25   **All-or-None Categorization Responses**

33

1   Our results suggest that stimuli are either categorized as faces and elicit full responses, even

2   at very brief presentation durations (e.g., 17 ms), or they are not categorized, even sometimes

3   at relatively low presentation durations (e.g., 50 ms), and no face-selective responses are

4   elicited. Indeed, no significant neural responses to behaviorally uncategorized faces were

5   present in the categorical, or differential, neural responses of Experiment 2. However, across

6   presentation rates from 12 to 60 Hz (83.3 to 17.6 ms), large deflections were present at the

7   times of the P1-face and N1-face in response only to behaviorally categorized faces (Figure

8   3). Additionally, these face-categorization responses obtained at 12 and 20 Hz were no larger

9   than at 24 and 30 Hz when only behaviorally categorized faces were taken into account

10   (Figure 4).

11   All-or-none categorization responses are not inconsistent with the apparent gradient

12   of response amplitude observed as presentation rate changed in Experiment 1 (Figure 2A).

13   Instead, we propose that the proportion of categorized to uncategorized faces drives graded

14   amplitude differences, similar to how behavioral categorization of "hits" and "misses" drives

15   graded behavioral response accuracies. Suggesting such an explanation, the behavioral and

16   frequency-tagged neural face-categorization responses were strongly correlated across

17   Experiments 2 and 1, respectively (Figure 2C). Moreover, the decrease in the percent of faces

18   that were categorized behaviorally (a drop of 31% when comparing 24 and 30 Hz to 12 and

19   20 Hz) matched the decrease in the amplitude (32%) of the neural face-categorization

20   response as quantified in the same experiment (Figure 4C and D).

21   It should be noted that the correspondence of neural and behavioral measures as

22   image processing time varies (as a function of presentation duration and/or backward-mask

23   latency) is a replication of a number of previous studies (e.g., Rolls & Tovee, 1994; Kovács,

24   Vogels & Orban, 1995; Vanni et al., 1996; Grill-Spector et al., 2000; Keysers et al., 2001;

25   Bacon-Macé et al., 2005). Note that such studies, if interpreted in this light at all, have been

34

1   taken as evidence for graded perceptual effects, attributing the decrease in response as a

2   decrease in the response to each stimulus presented (e.g., see Keysers et al., 2001; Bacon-

3   Mace et al., 2005). Our account and evidence of this relationship as variations in the

4   *proportion* of all-or-none neural responses, is to our knowledge novel and suggests a different

5   underlying process.

6   Further in line with all-or-none categorization, even as neural amplitude varied with

7   presentation rate in Experiment 1, face-categorization responses remained qualitatively

8   similar in terms of scalp topography (consistent maximal responses over the right occipto-

9   temporal region: Figure 5B), time course (no trends by presentation speed affecting the P1-

10   face, N1-face, P2-face, or P3-face deflections: Figure 6), and the distribution of amplitude

11   across face-categorization harmonic frequencies (Figure 5A). Qualitatively similar face-

12   categorization responses in both the frequency and time domain have also been reported with

13   this frequency-tagging categorization paradigm across studies using different stimulus-

14   presentation rates from about 6 to 20 Hz (e.g., Rossion et al., 2015; Retter & Rossion, 2016a;

15   Quek & Rossion, 2017; Retter et al., 2018) and across variations of stimulus visibility (i.e.,

16   low-pass filtering, Quek et al., 2018). Moreover, raw responses to faces or other visual

17   objects appear similar both quantitatively and qualitatively when using variable image

18   presentation durations, for example from about 40 to 500 ms (Cichy, Pantazis & Oliva,

19   2014).

20   While many arguments have been made for a progressively increasing response

21   amplitude enabling stimulus recognition in the visual cortex (e.g., Kleinschmidt et al., 1998;

22   Bar et al., 2001; Moutoussis & Zeki, 2002; Jemel et al., 2003), all-or-none neural responses

23   have been proposed previously for stimulus detection or identification (e.g., Del Cul, Baillet

24   & Dehaene, 2007; Quiroga et al., 2008; see also Navajas et al., 2013; Sekar, Findley &

25   Llinas, 2012; Sekar et al., 2013 for unmasked stimuli). However, the stimuli used in these

35

1   previous studies are limited in range and well-segmented (e.g., a few letters, numbers or

2   words in Del Cul et al., 2007; Sekar, Findley & Llinas, 2012; Sekar et al., 2013; a few

3   repeated exemplar images in Quiroga et al., 2008; full front segmented faces in Navajas et al.,

4   2013). Moreover, the timing of these all-or-none neural responses is unclear, with relatively

5   late all-or-none responses often reported (e.g., more than 270 ms in widespread regions in

6   Del Cul et al., 2007; more than 300 ms in Quiroga et al., 2008 for neurons in the human

7   medial temporal lobe). Compared to these studies, our study provides original evidence for

8   all-or-none neural categorization responses, given that: 1) it directly measures categorization

9   (i.e., a differential and generalizable response across widely variable exemplars); 2) through

10   frequency tagging, the neural response is objectively identified and quantified relative to

11   surrounding noise level in the frequency domain; 3) the neural markers of face categorization

12   in this paradigm are known to originate from face-selective neuronal populations in the

13   ventral occipito-temporal cortex (see Jonas et al., 2016; Gao et al., 2018); and 4) it shows

14   unambiguous face categorization responses emerging at a 17 ms duration/SOA, and with a

15   latency of 100 ms following stimulus onset.

16   Finally, the interpretation of all-or-none categorization responses does not imply that

17   at very short presentation durations the perception of a face stimulus is full, i.e., that all the

18   potential visual information, such as age, gender, and identity, has been extracted from the

19   image. Thus, perception may be graded in the sense that a face may be recognized as a face

20   before its identity is recognized. However, we suggest that the categorization process itself is

21   all-or-none: if neural responses differentiating faces from objects and generalizing across face

22   exemplars are elicited at all, they are elicited fully.

23   **Why Are Some Stimuli Not Categorized?**

24   While high-level (i.e., category-selective) visual processes are all-or-none, they may still be

25   reliant on the gradual accumulation of low-level visual information in early visual areas of

36

1   the brain (e.g., Windey et al., 2014). At an early stage, we theorize that if enough visual

2   information is assembled from the stimulus within about 20 ms, it goes on to develop and

3   trigger a full category-selective response after approximately 100-ms post-stimulus onset

4   (e.g., Crouzet et al., 2010; Crouzet & Thorpe, 2011; Liu et al., 2009; Jacques et al., 2016;

5   Retter & Rossion, 2016a; see Carandini et al., 2005 for a review of this process up to V1).

6   Indeed, in previous studies, the minimum duration to evoke a neural response was between

7   about 12-20 ms when images were backward-masked (e.g., with scrambled images; see

8   Breitmeyer, 1984); this measure was replicated with a wide variety of recording techniques,

9   including single unit in monkeys (Rolls & Tovee, 1994; Kovács, Vogels & Orban, 1995;

10   Keysers et al., 2001), MEG (Vanni et al., 1996), fMRI (Grill-Spector et al., 2000), and EEG

11   (Bacon-Macé et al., 2005) in humans. Here, a highly sensitive 20 ms window is consistent

12   with responses occurring rarely when interrupted with a backward-mask after 17 ms, and

13   large increases in the amount of categorization with the next durations (SOAs) of 25 and 30

14   ms.

15   However, if early visual processing is interrupted after about 20 ms but before the

16   category-selective response onset (at approximately 100 ms here, and also in Retter &

17   Rossion, 2016a), perceptual categorization may still not occur. Beyond its initiation, visual

18   information persists in the visual system, e.g., for about 60-100 ms, in early (e.g., V1) brain

19   areas (e.g., see the Discussions of Rolls & Tovee, 1994; Keysers et al., 2001). At these stages,

20   the accumulation of visual information may still be disrupted from masking producing

21   competition in early visual areas (see the "neural competition" theory of Keysers and Perrett,

22   2002; see also the discussion of Potter, 2012). In these experiments, the earlier a mask is

23   presented following a face, the shorter the sensory processing time, and the more likely that

24   the face will not be categorized. In our experiment, missed face categorizations occurred

25   more often at shorter presentation durations, and even up to SOAs of 50 ms (20 Hz; Figure

1 2). At each face presentation, whether or not the image is categorized or not likely depends

2 on many factors, including: properties of the face image (e.g., how central or high-contrast

3 the face is, or the face's head orientation), the overlap of low-level properties with its

4 flanking images (Crouzet & Thorpe, 2011; see again Potter, 2012; but also Maguire & Howe,

5 2016; Broers et al., 2018), the local context (e.g., whether the preceding face has been

6 categorized), and the neural representation of faces in the stimulated brain.

7 Note that in this case, a strong effect of masking is not predicted in high-level visual

8 areas, because stimuli temporally flanking the faces consist of non-face objects for at least 1 s

9 before and after (in contrast, see Retter & Rossion, 2016a for a case of high-level face

10 competition when face stimuli are presented with less than about 400 ms SOAs). In previous

11 forward- and backward-masking studies, early competition from masking stimuli was shown

12 to contribute to reduce mean response amplitude when target stimuli were shown for as low

13 as 28 ms; in contrast, with a "gap" between stimuli resulting in SOAs of at least 100 ms, a

14 reduction in amplitude was not produced (Keysers et al., 2005; Retter et al., 2018).

15 Nevertheless, we suggest that if a face is not categorized because of low-level interference

16 before about 100 ms, it does not enter perceptual awareness (see also Edelman, 1978, pp. 81-

17 88).

18 **The Minimal Speed for Conscious Face Categorization**

19 A minimum duration of 17 ms (60 Hz presentation rate) was required to elicit occasional face

20 categorization responses in both neural and behavioral measures. This minimal duration is

21 conservative in that it results from highly variable images: the faces varied in size,

22 eccentricity, head orientation, gaze direction, expression, etc. Nevertheless, behavioral face

23 categorization responses were present significantly, but minimally, with 17 ms of image

24 presentation (60 Hz; M = 3.03%, SE = 1.26%, with more frequent responses than false alarms

25 evident at this rate in half (8/16) of the individual participants). In contrast, face

38

1 categorization was clearly not possible at the shortest presentation duration tested, 8 ms (120

2 Hz), at which only one participant had one categorization, albeit with no false alarms.

3 At a neural level, face-categorization responses occurred occasionally at 17 ms over

4 the bilateral occipito-temporal ROI (evident in Experiment 2; data not shown), although these

5 were significant for only one participant in Experiment 1, who had the highest accuracy at

6 this presentation rate. Neural responses were instead not significant at 17 ms for most

7 participants in Experiment 1, likely due to an insufficient number and proportion of detected

8 trials to separate EEG signal from noise: e.g., a participant detecting the average 3% of faces

9 in the second experiment would only have detected about 4 of 126 faces presented in the first

10 experiment at 17 ms. In comparison, the best participant behaviorally detected 12.5% of faces

11 in the second experiment, predicting about 16 faces detected in the first experiment at 17 ms.

12 Note that these 16 responses to faces presumably compete for amplitude with the 110 faces

13 that were not detected, such that recording any significant response is notable, and likely due

14 to the high signal-to-noise ratio afforded from frequency-domain analyses of periodic

15 responses (Norcia et al., 2015).

16 Perhaps due to such factors, group-level significance was not evident here until 33

17 ms, when group mean accuracy was 22% (predicting about 28/126 faces detected in the first

18 experiment; Figure 1; Table 1). Across participants, the minimum presentation duration

19 required for a significant EEG face-categorization response over the bilateral OT ROI ranged

20 from 17 to 83 ms (Figure 1). Note that while a linear relation between behavior and neural

21 responses was significant across all presentation rates here, a threshold for recording neural

22 responses predicts a non-linear relationship between behavior and neural responses at the

23 fastest presentation rates (i.e., below about 17 ms here; see Figure 2C).

24 Although very short, the minimum duration of 17 ms for face categorization is

25 consistent with the results of several previous studies, both at behavioral and neural levels.

39

1   Notably, in a study by Keysers and colleagues (2001; 2005; see also Perrett et al., 2009), the

2   temporal tuning of the responses of single cells and populations in macaque monkeys was

3   investigated in response to monkey heads presented at varying orientations in RSVP

4   sequences. Neural firing rates were shown to decrease at faster presentation rates, but

5   responses to non-periodic target stimuli could still be detected at up to 14 ms (with 300-400

6   trials), the fastest rate in that experiment. Behavioral accuracy was measured separately with

7   human participants, which also decreased with presentation speed and remained significant

8   up to 14 ms. A later behavioral study using RSVP closely replicated this presentation

9   duration limit for high-level visual perception: Potter et al. (2014) showed recognition of

10 various natural images (e.g., people or flowers) by human participants at 13 ms, the fastest

11 rate in their experiment. Further, additional studies have confirmed the presence of face (and

12 object) detection responses at 16-17 ms with behavior, ECoG, and MEG, in masked or short

13 RSVP sequences (Grill-Spector & Kanwisher, 2005; Fisch et al., 2009; Mohsenzadeh et al.,

14 2018). The common limit of approximately 15 ms across these studies perhaps reflects

15 something general in visual perception: the minimum amount of time required for a visual

16 stimulus to be coherently processed in early visual areas, before triggering a high-level

17 classification response.

18   It should be noted that our study was limited in its ability to present stimuli at many

19 durations near potential thresholds, due to the constraints of monitor refresh rates. For

20 example, durations between 17 and 8 ms could not be tested, since these durations consist of

21 two and one screen display frames, respectively, on a 120 Hz monitor. This limitation in

22 available presentation durations has similarly affected many previous studies, and for

23 example limits pinpointing whether a limit is at 13 (Potter et al., 2014), 14 (Keysers et al.,

24 2001), 16 (Fisch et al., 2009), or 17 ms (Grill-Spector & Kanwisher, 2005; Mohsenzadeh et

25 al., 2018). Accordingly, it would also be difficult to assess subtler questions, such as whether

40

1   faces may be identified faster than other objects around such upper limits (after 17 ms, three

2   display frames at 120 Hz determined the next available duration at 25 ms; but see Fisch et al.,

3   2009; and for a different approach, see Rousselet et al., 2003). Such questions require faster

4   framerates than most current displays afford.

5   **The Optimal Speed of Face-Categorization**

6   The maximal behavioral accuracy and neural amplitude occurred at the same time, around 80

7   ms (12 Hz; Figure 2). This optimum time aligns well with ranges from previous studies: for

8   example, maximum accuracy and EEG responses were reported at 81 ms for categorizing

9   animal images followed by dynamic visual masks (adjacent durations of 44 and 106 ms;

10   Bacon-Mace et al., 2005). Additionally, a rough plateau of fMRI and behavioral responses

11   between two duration steps of 120 to 500 ms was reported for stimulus naming (reduced

12   responses at the adjacent lower step of 40 ms; see Fig. 3 of Grill-Spector et al., 2000). While

13   we expect variation by the type of visual process targeted, our results suggest that 80 ms is

14   sufficient for neurotypical adult observers to categorize any natural view of a face in the

15   visual scene (exempting the cases that faces are occluded, artificially degraded, etc.). A

16   practical implication is that studies measuring neural as well as behavioral generic face (vs.

17   object) categorization responses need not present stimuli longer than 80 ms.

18   Above 80 ms, behavioral accuracy remains at ceiling (above 97% at 6 and 3 Hz;

19   Table 3). Neural amplitude does not increase at these rates: if anything, it decreases (Figure

20   2A). In the section "Why Are Some Stimuli Not Categorized?", we discussed how early

21   visual responses may produce masking effects, or competition at a neural level, across

22   neighboring images in RSVP streams. These responses were not specifically represented in

23   the targeted face-categorization responses of the present experiments, which reflect the

24   contrast of faces with non-face objects. However, we also measured general visual responses

25   to stimulus presentation, maximal over the medial occipital cortex, and increasing in

41

1   amplitude exponentially as stimulus presentation duration increased (Supplementary Figure

2   1). It is possible these visual responses were more developed as presentation duration

3   increased, being more likely to reach higher-level general visual processing that nevertheless

4   might have interfered more with face-selective processing. Indeed, these stimulus-

5   presentation responses showed some variance in topography across presentation duration,

6   with a particularly distinct response localization at 60 and 120 Hz (Figure 5D), although these

7   latter responses may have been contaminated by noise from the power line frequency at 60

8   Hz. Thus, one possible explanation for lower amplitudes at longer presentation durations

9   could be an interaction between face and non-face object processing at rates below 12 Hz

10   (compare Supplemental Figure 1 with Figure 2A; and Figure 5B with 5D).

11   **Individual Differences in Generic Face Categorization**

12   Finding reliable correlations between neural amplitudes and behavioral measures of visual

13   categorization, even when using frequency-tagging with well quantified measures, remains a

14   challenge (e.g., Retter & Rossion, 2016b; Retter & Rossion, 2017; Xu et al., 2017). This is

15   likely because neural responses are heavily influenced by physiological factors: in particular,

16   for EEG the orientation of dipole generators, influenced by individual differences in cortical

17   folding, is thought to play a large role; additional factors include skull thickness and the

18   conductivity of tissue in between the generators and captors (Luck, 2005; Nunez &

19   Srinivasan, 2006; Woodman, 2010).

20   Here, we developed a method to identify the relationship between individual

21   participants' behavioral and neural measures. We were able to show a significant correlation

22   between participants' behavior (measured in Experiment 2 across all presentation rates in

23   terms of inverse efficiency) and face-categorization EEG amplitude (measured in Experiment

24   1 by weighting the mid-frequency responses from 20 – 30 Hz (30 – 50 ms) by that of 3 Hz

25   (333 ms), over the OT ROI; Figure 7D).

42

1   This correlation relies on two elements: first, we used a range of presentation rates

2   that accentuated individual differences in generic face categorization. While a great amount

3   of research is being done in terms of individual differences in individual face recognition,

4   differences in generic face vs. object categorization have not been reported to our knowledge,

5   likely due to the high level of expertise that human adults have at performing this function

6   (Crouzet et al., 2010; Herschler & Hochstein, 2005; Scheirer et al., 2014). Here, at

7   presentation rates from 20 to 30 Hz (30 to 50 ms; see Figure 7A and B), we were able to

8   make this function challenging enough to keep all participants below a behavioral ceiling (at

9   80 ms, i.e., 12 Hz, and later; Figure 2B) and yet above the threshold for measuring significant

10   neural responses in most individual participants (Figure 1).

11   Second, we used a relative measure of neural activity, weighting individuals' neural

12   amplitudes in this mid-frequency range by their amplitudes at the lowest presentation rate. At

13   this rate, 3 Hz, individual differences had no relationship to behavioral performance (Figure

14   7C), again likely due to physiological influences on the neural amplitudes. Using a relative

15   neural measure likely assisted in cancelling out such influences, enabling a more diagnostic

16   signature of individuals' EEG response amplitudes relative to behavior (again, Figure 7D).

17   Note again that this relationship is found when comparing explicit behavioral face

18   categorization (Experiment 2, face detection task) to implicit neural face-categorization

19   responses (Experiment 1, naïve, fixation-cross task). Although most (14/16) participants

20   spontaneously reported perceiving faces when describing the sequences after Experiment 1,

21   none reported observing their periodic presentation. Note also that the responses to stimulus-

22   presentation duration showed very different trends, being significant at all durations but

23   decreasing exponentially as duration increased (Supplemental Figure 1), suggesting that the

24   targeted responses are truly specific to face categorization.

25   **Summary**

43

1 We have shown that widely variable natural images of faces can be categorized in contrast to

2 non-face objects at presentation rates as short as 17 ms, and optimally at 80 ms. Individual

3 differences were evident within this range, such that individuals' relative neural amplitudes

4 (across 30-50 ms) were significantly correlated with their behavioral performance (across all

5 presentation rates). While overall the amplitude of face-categorization responses appeared to

6 increase gradually as presentation duration increased, this was completely accounted for by

7 the proportion of discrete "hits" and "misses" in categorization, with categorized faces

8 producing full neural face-categorization responses and uncategorized faces producing no

9 neural face-categorization responses. This provides convergent neural and behavioral

10 evidence that human perceptual (face) categorization is all-or-none.

11 ## **Acknowledgments**

23 ## **References**

24 Adrian E. D., Matthews B. H. C. (1934). The Berger rhythm: Potential changes from the

25   occipital lobes in man. *Brain*, 4 (57), 355–385.

44

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A. Luby, M. & Spencer, D. D. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology,* 71(2), 821-825.

Bachman, T. (2013). On the all-or-none rule of conscious perception. *Frontiers in Human Neuroscience,* 7, 387.

Bacon-Macé, N., Macé, M. J., Fabre-Thorpe, M. & Thorpe, S. J. (2005). The time course of visual processing: backward masking and natural scene categorisation. *Vision Research,* 45(11), 1459-1469.

Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D. Rosen, B. R. & Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron,* 29(2), 529-535.

Breitmeyer, B. G. (1984). *Visual masking: an integrative approach.* Oxford: Oxford University Press.

Broers, N., Potter, M. C. & Nieuwenstein, M. R. (2018). *Psychon Bull Rev*, 25:1080.

Caharel, S., Leleu, A., Bernard, C., Viggiano, M.P., Lalonde, R., & Rebaï, M. (2013). Early holistic face-like processing of Arcimboldo paintings in the right occipito-temporal cortex: evidence from the N170 ERP component. *Int J Psychophysiol.*, 90, 157-64.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y, & Olshausen, B. A. et al. (2005). Do we know what the early visual system does? *Journal of Neuroscience*, 25, 10577–10597.

Christensen, M. S., Ramsøy, T. Z., Lund, T. E., Madsen, K. H., & Rowe, J. B. (2006). An fMRI study of the neural correlates of graded visual perception. *Neuroimage*, 31, 1711–1725.

Cichy, R. M., Pantazis, D. & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience,* 17(3), 455-462.

bioRxiv preprint doi: https://doi.org/10.1101/658625; this version posted June 6, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Crouzet, S. M., Kirchner, H. & Thorpe, S. J. (2010). Fast saccades toward faces: face
2   detection in just 100 ms. Journal of Vision, 10 :3.

3 Crouzet, S. M. & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers*
4   *in Psychology,* 2:342.

5 de Gardelle, V., Charles, L., & Kouider, S. (2011). Perceptual awareness and categorical
6   representation of faces: Evidence from masked priming. *Consciousness and*
7   *Cognition: An International Journal*, 20(4), 1272-1281.

8 Del Cul, A. Baillet, S. & Dehaene, S. (2007). Brain Dynamics Underlying the Nonlinear
9   Threshold for Access to Consciousness. *PLoS Biology,* 5(10), e260.

10 Dzhelyova, M. & Rossion, B. (2014). The effect of parametric stimulus size variation on
11   individual face discrimination indexed by fast periodic visual stimuliation. *BMC*
12   *Neuroscience*, 15:87, 1-12.

13 Edelman, G. M. (1978). Group selecting and phasic reenterant signalling : A theory a higher
14   brain function. In *The Mindful Brain* (pp. 51-100). Cambridge, MA: MIT Press.

15 Edelman, G. M. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection.*
16   Cambridge, MA: MIT Press.

17 Enns, J.T., & Di Lollo V. (2000). What's new in visual masking? *Trends in Cognitive*
18   *Sciences*, 4, 345-352.

19 Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S. Andelman, F. et al.
20   (2009). Neural "ignition": enhanced activation linked to perceptual awareness in
21   human ventral stream visual cortex. *Neuron,* 64, 562-574.

22 Foldiak, P., Xiao, D., Keysers, C., Edwards, R. & Perrett, D. I. (2004). Rapid serial visual
23   presentation for the determination of neural selectivity in area STSa. *Progress in*
24   *Brain Research*, 144, 107-116,

46

Gao, X., Gentile, F., Rossion, B. (2018). Fast periodic stimulation: a highly effective approach in fMRI brain mapping. *Brain Structure and Function,* 223, 2433-2454.

Goldstone, R. L., Kersten, A. & Carvalho, P. F. (2018). *Categorization and Concepts*. In Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, J. T. Wixted (Ed.). doi:10.1002/9781119170174.epcn308

Grill-Spector, K., Kushnir, T., Hendler, T. & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience,* 3(8), 837-843.

Grill-Spector, K. & Kanwisher, N. (2005). Visual recognition: as soon as you know it's there, you know what it is. *Psychological Science,* 16(2), 152-160.

Harris, J. A., Wu, C. T. & Woldorff, M. G. (2011). Sandwich masking eliminates both visual awareness of faces and face-specific brain activity through a feedforward mechanism. *Journal of Vision,* 11(7): 10.

Helmholtz, H. (1867). "*Handbuch der physiologischen Optik*", in Allgemeine Encyklopddie der Physik volume 9, Ed. G Karsten (Leipzig: Voss)

Herrmann, C.S. (2001). Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Experimental Brain Research, 137, 346-353.

Herschler, O. & Hochstein, S. (2005). At first sight: a high-level pop out effect for faces. *Vision Research,* 45(13), 1707-1724.

Jacques, C.*, Retter, T.L.*, Rossion, B. (2016). A single glance at a face generates larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. *NeuroImage*, 137, 21-33

47

Jacques, C., Witthoft, N., Weiner, K. S., Foster, B. L., Rangarajan, V., Hermes, D. et al. (2016). Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia,* 83, 14-28.

Jonas, J.*, Jacques, C.*, Liu-Shuang, J., Brissart, H., Colnat-Coulbois, S., Maillard, L., Rossion, B. (2016). A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proc. Natl. Acad. Sci. USA*, 113, E4088-E4097.

Kanwisher, N., McDermott, J. & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.

Keysers, C., Xiao, D.-K., Foldiak, P. & Perrett, D. I. (2005). Out of sight but not out of mind: the neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology,* 22, 3-4, 316-332.

Kovács, G. Vogels, R. & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *PNAS USA,* 92(12), 5587-5591.

Keysers, C., Xiao, D.- K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, 13, 90–101.

Keysers, C. & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Science*, 6, 120–125.

Kleinschmidt, A., Buchel, C., Zeki, S. & Frackowiak, R. S. J. (1998). Human brain activity during spontaneously reversing perception of ambiguous figures. *Proceedings of the Royal Society of London B,* 265:1413.

Lee, I. A. & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

48

1 Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: fast
2   decoding of object information from intracranial field potentials in human visual
3   cortex. *Neuron*, 62(2), 281-90.

4 Luck, S. J. (2005). *An introduction to the event-related potential technique.* Cambridge, MA:
5   MIT Press.

6 Maguire, J. F. & Howe, P. D. L. (2016). Failure to detect meaning in RSVP at 27 ms per
7   picture. *Attention, Perception, and Psychophysics*. 78:1405.

8 McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An
9   ambiguity associated with analysis of variance models. Electroencephalography and
10   Clinical Neurophysiology, Evoked Potentials Section, 62, 203–208.

11 Mohsenzadeh, Y., Qin, S., Cichy, R. M. & Pantazis, D. (2018). Ultra-Rapid serial visual
12   presentation reveals dynamics of feedforward and feedback processes in the ventral
13   visual pathway. *eLife,* 7:e36329.

14 Moutoussis, K. & Zeki, S. (2002). The relationship between cortical activation and
15   perception investigated with invisible stimuli. *PNAS USA,* 99(14): 9527–9532.

16 Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

17 Navajas, J., Ahmadi, M., & Quian Quiroga, R. (2013). Uncovering the mechanisms of
18   conscious face perception: a single-trial study of the n170 responses. *J Neurosci.* 33,
19   1337-1343.

20

21 Navajas, J., Rey, H. G. & Quiroga, R. Q. (2014). Perceptual and contextual awareness:
22   methodological considerations in the search for the neural correlates of consciousness.
23   *Frontiers in Psychology,* 5, 959.

1   Norcia, A.M., Appelbaum, L.G., Ales, J.M., Cottereau, B., Rossion, B. (2015). The steady-
2   state visual evoked potential in vision research: a review. *Journal of Vision,* 15(6):4,
3   1-46.

4   Nunez, P. L. & Srinivasan, R. (2006). *Electric Fields of the Brain.* Oxford: Oxford University
5   Press.

6   Omer, Y., Sapir, R., Hatuka, Y., & Yovel G. (2019). What Is a Face? Critical Features for
7   Face Detection. *Perception*, 48, 437-446.

8   Oostenveld, R. & Praamstra, P. (2001). The five percent electrode system for high-resolution
9   EEG and ERP measurements. 112(4), 713-719.

10 Perrett, D. I. (2009). Seeing the future: Natural image sequences produce "anticipatory"
11   neuronal activity and bias perceptual report. *The Quarterly Journal of Experimental*
12   *Psychology,* 62(11), 2081-2104.

13 Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in*
14   *Psychology*, 3:32.

15 Potter, M. C. & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *J.*
16   *Exp. Psychol*, 81, 10–15.

17 Potter, M.C., Wyble, B., Hagmann, C.E., & McCourt, E.S. (2014). Detecting meaning in
18   RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, 76(2), 270-279.

19 Puce, A., Allison, T., Gore, J. C. & McCarthy, G. (1995). Face-sensitive regions in human
20   extrastriate cortex studied by functional MRI. *Journal of Neurophysiology,* 74(3),
21   1192-1199.

22 Quek, G. & Rossion, B. (2017). Category-selective human brain processes elicited in fast
23   periodic visual stimulation streams are immune to temporal
24   predictability. *Neuropsychologia*, 104, 182-200.

Quiroga, R. Q., Mukamel, R., Isham, E. A., Malach, R. & Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. *PNAS,* 105(9), 3599-3604.

Regan D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine*. Amsterdam, the Netherlands: Elsevier.

Retter, T. L., Jiang, F., Webster, M. A. & Rossion, B. (2018). Dissociable effects of inter-stimulus interval and presentation duration on rapid face categorization. *Vision Research*, 145, 11-20.

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.

Retter, T. L. & Rossion, B. (2016b). Visual adaptation provides objective electrophysiological evidence of facial identity discrimination. *Cortex*, 80, 35-50.

Rolls, E. T. & Tovée, M. J. (1994). Processing Speed in the Cerebral Cortex and the Neurophysiology of Visual Masking. *Proceedings of the Royal Society B,* 257, 9-15.

Rossion, B., Torfs, K., Jacques, C., Liu-Shuang, J. (2015). Fast periodic presentation of natural face images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*, 15(1), 1-18.

Rousselet, G. A., Macé M. J.-M, Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, 3(6):5.

Scheirer, W. J., Anthony, S. E., Nakayama, K. & Cox, D. D. (2014). Perceptual annotation: Measuring human vision to improve computer vision. *IEE Transactions on Pattern Analysis and Machine Intelligence,* 36(8), 1679-1686.

Sekar, K., Findley, W. M. & Linas, R. R. (2012). Evidence for an all-or-none perceptual response: single-trial analyses of magnetoencephalography signals indicate an abrupt transition between visual perception and its absence. *Neuroscience,* 206, 167-182.

51

Sekar, K., Findley, W. M., Poeppel, D., & Llinás, R. R. (2013). Cortical response tracking the conscious experience of threshold duration visual stimuli indicates visual perception is all or none. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(14), 5642-7.

Sergent, J., Ohta, S. & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain,* 115(1), 15-36.

Shafto J. P., Pitts M. A. (2015). Neural signatures of conscious face perception in an inattentional blindness paradigm. *Journal of Neuroscience,* 35, 10940–10948.

Smith, E. E. & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Thorpe, S. J., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature,* 381(6582), 520-522.

Townsend, J. T. & Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.

Vanni, S., Revonsuo, A., Saarinen, J. & Hari, R. (1996). Visual awareness of objects correlates with activity of right occipital cortex. *Neuroreport,* 8(1), 183-186.

Weiner, K. S. & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage,* 52(4): 1559–1573.

Windey, B. & Cleeremans, A. (2015). Consciousness as a graded and an all-or-none phenomenon: A conceptual analysis. *Consciousness and cognition,* 35, 185-191.

Windey, B., Vermeiren, A., Atas, A. & Cleeremans, A. (2014). The graded and dichotomous nature of visual awareness. *Phil. Trans. R. Soc. B,* 369:20130282.

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, and Psychophysics.* 72(8), 2031-2046.

1   Xu, B., Liu-Shuang, J., Rossion, B. & Tanaka, J. (2017). Individual differences in face

2   identity processing with fast periodic visual stimulation. *Journal of Cognitive*

3   *Neuroscience,* 29(8), 1368-1377.

4   Zhen, Z., Yang, Z., Huang, L., Kong, X.-z., Wang, X., Dang, X., et al. (2015). Quantifying

5   interindividual variability and asymmetry of face-selective regions: a probabilistic

6   functional atlas. *Neuroimage*, 113, 13-25.