

1 **Perceptual difficulty modulates the direction of**

2 **information flow in familiar face recognition**

3 Hamid Karimi-Rouzbahani^{1,2*}, Farzad Ramezani³, Alexandra Woolgar^{1,2}, Anina Rich²,
4 Masoud Ghodrati^{4*}

5

6 ¹Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, UK

7 ²Perception in Action Research Centre and Department of Cognitive Science Macquarie University,
8 Australia

9 ³Department of Computer Science, School of Mathematics, Statistics, and Computer Science, University
10 of Tehran, Iran

11 ⁴Neuroscience Program, Biomedicine Discovery Institute, Monash University, Australia

12 * to whom correspondence should be addressed.

13 **Abstract**

14 Humans are fast and accurate when they recognize familiar faces. Previous
15 neurophysiological studies have shown enhanced representations for the dichotomy of
16 familiar vs. unfamiliar faces. As familiarity is a spectrum, however, any neural correlate
17 should reflect graded representations for more vs. less familiar faces along the
18 spectrum. By systematically varying familiarity across stimuli, we show a neural
19 familiarity spectrum using electroencephalography. We then evaluated the
20 spatiotemporal dynamics of familiar face recognition across the brain. Specifically, we
21 developed a novel informational connectivity method to test whether peri-frontal brain
22 areas contribute to familiar face recognition. Results showed that feed-forward flow
23 dominates for the most familiar faces and top-down flow was only dominant when
24 sensory evidence was insufficient to support face recognition. These results
25 demonstrate that perceptual difficulty and the level of familiarity influence the neural
26 representation of familiar faces and the degree to which peri-frontal neural networks
27 contribute to familiar face recognition.

28 **Keywords:** Face Recognition, Familiar Faces, Multivariate Pattern Analysis (MVPA),
29 Representative Similarity Analysis (RSA), Informational Brain Connectivity

30

Introduction

31 Faces are crucial for our social interactions, allowing us to extract information
32 about identity, gender, age, familiarity, intent and emotion. Humans categorize familiar
33 faces more quickly and accurately than unfamiliar ones, and this advantage is more
34 pronounced under difficult viewing conditions, where categorizing unfamiliar faces often
35 fails (Ramon and Gobbini, 2018; Young and Burton, 2018). The neural correlates of this
36 behavioral advantage suggest an enhanced representation of familiar over unfamiliar
37 faces in the brain (Dobs et al., 2019; Landi and Freiwald, 2017). Here, we focus on
38 addressing two major questions about familiar face recognition. First, whether there is a
39 “familiarity spectrum” for faces in the brain with enhanced representations for more vs.
40 less familiar faces along the spectrum. Second, whether higher-order frontal brain areas
41 contribute to familiar face recognition, as they do to object recognition (Bar et al., 2006;
42 Summerfield et al., 2006; Goddard et al., 2016; Karimi-Rouzbhani et al., 2019), and
43 whether levels of face familiarity and perceptual difficulty (as has been suggested
44 previously (Woolgar et al., 2011; Woolgar et al., 2015)) impact the involvement of frontal
45 cognitive areas in familiar face recognition.

46 One of the main limitations of previous studies, which hinders our progress in
47 answering our first question, is that they mostly used celebrity faces as the familiar
48 category (Ambrus et al., 2019; Collins et al., 2018; Dobs et al., 2019). As familiar faces
49 can range widely from celebrity faces to highly familiar ones such as family members,
50 relatives, friends, and even one's own face (Ramon and Gobbini, 2018), these results
51 might not reflect the full familiarity spectrum. A better understanding of familiar face
52 recognition requires characterizing the computational steps and representations for sub-
53 categories of familiar faces, including personally familiar, visually familiar, famous, and
54 experimentally learned faces. Such face categories do not only differ in terms of how
55 much exposure the individual has had to them, but also the availability of personal
56 knowledge, relationships, and emotions associated with the identities in question

57 (Leppänen and Nelson, 2009; Ramon and Gobbini, 2018; Kovács, 2020). However, we
58 still expect that potentially enhanced representations for more vs. less familiar faces, as
59 they modulate the behavior, can also be detected using neuroimaging analysis.
60 Moreover, these categories may vary in terms of the potential for top-down influences in
61 the process. Importantly, while a few functional magnetic resonance imaging (fMRI)
62 studies have investigated the differences between different levels of familiar faces
63 (Gobbini et al., 2004; Landi and Freiwald, 2017; Leibenluft et al., 2004; Ramon et al.,
64 2015; Sugiura et al., 2015; Taylor et al., 2009), there are no studies that systematically
65 compare the temporal dynamics of *information processing* across this familiarity
66 spectrum. Specifically, while event-related potential (ERP) analyses have shown
67 amplitude modulation by levels of face familiarity (Henson et al., 2008; Kaufmann et al.,
68 2009; Schweinberger et al., 2002; Huang et al., 2017), they remain silent about whether
69 more familiar faces are represented more distinctly than less familiar faces - amplitude
70 modulation does not necessarily mean that information is being represented. To
71 address this issue, we can use multivariate pattern analysis (MVPA or decoding;
72 Ambrus et al., 2019; Karimi-Rouzbahani et al., 2017a), which provides higher sensitivity
73 (Norman et al., 2006) than univariate (e.g., ERP) analysis, to compare the amount of
74 information in each of the familiarity levels.

75 In line with our second question, recent human studies have compared the
76 neural dynamics for familiar versus unfamiliar face processing using the high temporal
77 resolution of electroencephalography (EEG; Ambrus et al., 2019; Collins et al., 2018)
78 and magnetoencephalography (MEG; Dobs et al., 2019). These studies have found that
79 familiarity affects the initial time windows of face processing, which are generally
80 attributed to the feed-forward mechanisms of the brain. In particular, they have explored
81 the possibility that the face familiarity effect occurs because these faces have been
82 seen repeatedly, leading to the development of low-level representations for familiar
83 faces in the occipito-temporal visual system. This in turn facilitates the flow of familiar
84 face information in a bottom-up feed-forward manner from the occipito-temporal to the
85 frontal areas for recognition (di Oleggio Castello and Gobbini, 2015; Ramon et al., 2015;
86 Ellis et al., 1979; Young and Burton, 2018). On the other hand, studies have also shown
87 the role of frontal brain areas in facilitating the processing of visual inputs (Bar et al.,

88 2006; Kveraga et al., 2007; Goddard et al., 2016; Karimi-Rouzbahani et al., 2019), such
89 as faces (Kramer et al., 2018; Summerfield et al., 2006), by feeding back signals to the
90 face-selective areas in the occipito-temporal visual areas, particularly when the visual
91 input is ambiguous (Summerfield et al., 2006) or during face imagery (Mechelli et al.,
92 2004; Johnson et al., 2007). These top-down mechanisms, which were localized in
93 medial prefrontal cortex (MPFC), have been suggested (but not quantitatively
94 supported) to reflect feedback of (pre-existing) face templates, against which the input
95 faces are compared for correct recognition (Summerfield et al., 2006) in a recollection
96 procedure (Brown and Banks, 2015). A more recent fMRI study showed that there is
97 significant face selectivity in the inferior frontal gyrus (IFG) over the frontal cortex and
98 that the same area is strongly connected to the well-established face-selective superior
99 temporal sulcus (STS) over the temporal cortex (Davies-Thompson and Andrews,
100 2012), which was consistent with a previous diffusion tensor imaging study (Ethofer et
101 al., 2011). Despite the large literature of face recognition supporting the roles of both the
102 peri-occipital (e.g. Fusiform face area, STS) and peri-frontal¹ (e.g. IFG, MPFC and
103 posterior cingulate cortex (Ramon et al., 2015)) brain areas (i.e. feed-forward and
104 feedback mechanisms), their potential interactions in familiar face recognition have
105 remained ambiguous (see for reviews Ramon and Gobbini, 2018; Duchaine and Yovel,
106 2015). We develop novel connectivity methods to track the flow of information along the
107 feed-forward and feedback mechanisms and assess the role of these mechanisms in
108 familiar face recognition.

109 One critical aspect of the studies that successfully detected top-down peri-frontal
110 to peri-occipital feedback signals (Bar et al., 2006; Summerfield et al., 2006; Goddard et
111 al., 2016) has been the *active* involvement of the participant in a task. In recent E/MEG
112 studies reporting support for a feed-forward explanation of the face familiarity effect,
113 participants were asked to detect target faces (Ambrus et al., 2019) or find a match
114 between faces in series of consecutively presented faces (Dobs et al., 2019). This
115 makes familiarity irrelevant to the task of the participant. Such indirect tasks may reduce
116 the involvement of top-down familiarity-related feedback mechanisms, as was

¹ Here we use the terms “peri-occipital” and “peri-frontal” to refer broadly to groups of electrodes selected from posterior and anterior parts of the EEG cap, respectively (as indicated in Figure 5).

117 demonstrated by a recent study (Kay et al., 2017), which found reduced feedback
118 signals (from intraparietal to ventro-temporal cortex) when comparing fixation versus an
119 active task in an fMRI study. Therefore, to answer our research questions and fully test
120 the contribution of feedback to the familiarity effect, we need active tasks that are
121 affected by familiarity.

122 Timing information is also crucial in evaluating the flows of feed-forward and
123 feedback information as these processes often differ in the temporal dynamics. With the
124 advent of informational connectivity analyses, we now have the potential to examine the
125 interaction of information between feed-forward and feedback mechanisms to
126 characterize their potential spatiotemporal contribution to familiar face recognition
127 (Goddard et al., 2016; Goddard et al., 2019; Anzellotti and Coutanche, 2018; Basti et
128 al., 2020; Karimi-Rouzbahani et al., 2020a). However, this requires novel methods to
129 track the flow of familiarity information from a given brain area to a destination area and
130 link this flow to the behavioral task goals to confirm its biological relevance. Such
131 analyses can provide valuable insights for understanding the neural mechanisms
132 underlying familiar face recognition in humans.

133 In our study, participants performed a familiar vs. unfamiliar face categorization
134 task on sequences of images selected from four face categories (i.e., unfamiliar,
135 famous, self, and personally familiar faces), with dynamically updating noise patterns,
136 while their EEG data were recorded. It was crucial to use dynamic noise in this study. If
137 stimuli were presented statically for more than ~200ms, this would result in a dominant
138 feed-forward flow of information simply due to the incoming information (Goddard et al.,
139 2016; Karimi-Rouzbahani, 2019; Lamme et al., 2000). On the other hand, if we present
140 stimuli for very brief durations (e.g. < 50 ms), there may be insufficient time to evoke
141 familiarity processing. By varying the signal-to-noise ratio of each image sequence
142 using perceptual coherence, we were able to investigate how information for the
143 different familiar categories gradually builds up in the electrical activity recordable by
144 scalp electrodes, and how this relates to the amount of sensory evidence available in
145 the stimulus (perceptual difficulty). The manipulation of sensory evidence also allowed
146 us to investigate when, and how, feedback information flow affects familiar face

147 recognition. Using univariate and multivariate pattern analyses, representational
148 similarity analysis (RSA) and a novel informational connectivity analysis method, we
149 reveal the temporal dynamics of neural representations for different levels of face
150 familiarity.

151 Our results show that self and personally familiar faces lead to higher perceptual
152 categorization accuracy and enhanced representation in the brain even when sensory
153 information is limited while famous (visually familiar) and unfamiliar face categorization
154 is only possible in high-coherence conditions. Importantly, our novel information flow
155 analysis suggests that in high-coherence conditions the feed-forward sweep of face
156 category information processing is dominant, while at lower coherence levels the
157 exchange of face category information is consistent with feedback flow of information.
158 The change in dominance of feedback versus feed-forward effects as a function of
159 coherence level is consistent with a dynamic exchange of information between higher-
160 order (frontal) cognitive and visual areas depending on the amount of sensory evidence.

161

162 Results

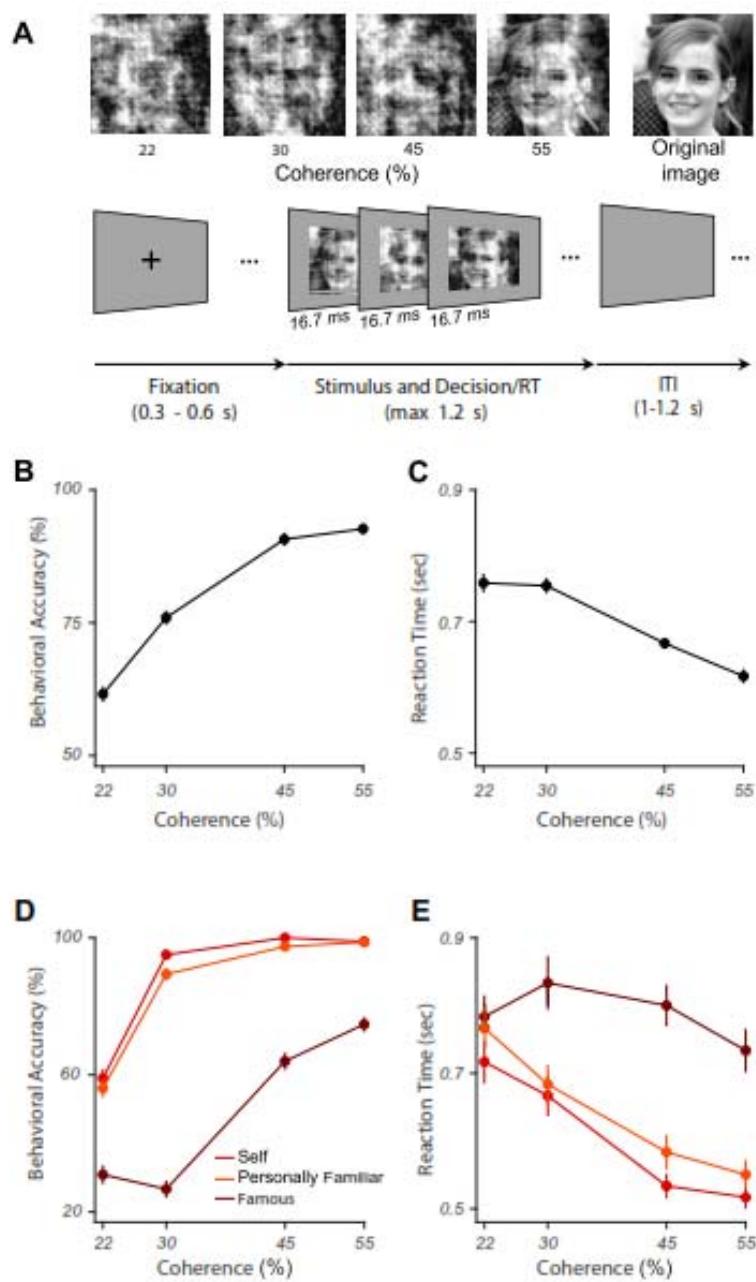
163 We designed a paradigm to study how the stimulus- and decision-related
164 activations for different levels of face familiarity build up during stimulus presentation
165 and how these built-up activations relate to the amount of sensory evidence about each
166 category. We recorded EEG data from human participants (n=18) while they
167 categorized face images as familiar or unfamiliar. We varied the amount of sensory
168 evidence by manipulating the phase coherence of images on different trials (Figure 1A).
169 In each 1.2 s (max) sequence of image presentation (trial), the pattern of noise changed
170 in each frame (16.7 ms) while the face image and the overall coherence level remained
171 the same. Familiar face images (n=120) were selected equally from celebrity faces,
172 photos of the participants' own face, and personally familiar faces (e.g., friends, family
173 members, relatives of the participant) while unfamiliar face images (n=120) were
174 completely unknown to participants before the experiment. Within each block of trials,

175 familiar and unfamiliar face images with different coherence levels were presented in
176 random order.

177

178 Levels of face familiarity are reflected in behavioral performance

179 We quantified our behavioral results using accuracy and reaction times on
180 correct trials. Specifically, accuracy was the percentage of images correctly categorized
181 as either familiar or unfamiliar. All participants performed with high accuracy (>92%) at
182 the highest phase coherence (55%), and their accuracy was significantly lower (~62%)
183 at the lowest coherence (22%; $F(3,272)=75.839$, $p<0.001$; Figure 1B). The correct
184 reaction times show that participants were significantly faster to categorize the face at
185 high phase coherence levels than lower ones ($F(3,272)=65.797$, $p<0.001$, main effect;
186 Figure 1C). We also calculated the accuracy and reaction times for the sub-categories
187 of the familiar category separately (i.e. famous, personally familiar and self). The
188 calculated accuracy here is the percentage of correct responses within each of these
189 familiar sub-categories. The results show a gradual increase in accuracy as a function
190 of phase coherence and familiarity (Figure 1D, two-way ANOVA. factors: coherence
191 level and face category. Face category main effect: $F(2,408)=188.708$, $p<0.001$,
192 coherence main effect: $F(3,408)= 115.977$, $p<0.001$, and interaction: $F(6,408)=12.979$,
193 $p<0.001$), with the highest accuracy in categorizing their own (self), then personally
194 familiar, and finally famous (or visually familiar) faces. The reaction time analysis also
195 showed a similar pattern where participants were fastest to categorize self faces, then
196 personally familiar and famous faces (Figure 1E, two-way ANOVA, factors: coherence
197 level and face category. Face category main effect: $F(2,404)=174.063$, $p<0.001$,
198 coherence main effect: $F(3,404)= 104.861$, $p<0.001$). We did not evaluate any potential
199 interaction between coherence levels and familiarity levels as it does not address any
200 hypothesis in this study. All reported p-values were corrected for multiple comparisons
201 at $p<0.05$ using Bonferroni correction.



202

203 **Figure 1. Experimental design and behavioral results for familiar vs. unfamiliar face**
204 **categorization.** (A) Upper row shows a sample face image (from the famous category) at the four
205 different phase coherence levels (22, 30, 45, and 55%) used in this experiment, in addition to the original
206 image (not used). Lower row shows schematic representation of the experimental paradigm. In each trial,
207 a black fixation cross was presented for 300-600 ms (randomly selected). Then, a noisy and rapidly
208 updating (every 16.7 ms) stimulus of a face image (unfamiliar, famous, personally familiar, or self), at one
209 of the four possible phase coherence levels, was presented until response, for a maximum of 1.2 s.
210 Participants had to categorize the stimulus as familiar or unfamiliar by pressing one of two buttons (button
211 mappings swapped across the two sessions, counterbalanced across participants). There was then a
212 variable inter-trial interval (ITI) lasting between 1-1.2 s (chosen from a uniform random distribution; see a
213 demo of the task here <https://osf.io/n7b8f/>). (B) Mean behavioral accuracy for face categorization across
214 all stimuli, as a function of coherence levels; (C) Median reaction times for correctly categorized face trials

215 across all conditions, as a function of coherence levels. (D) and (E) show the results for different familiar
216 face sub-categories. Error bars in all panels are the standard error of the mean across participants
217 (smaller for panels B and C).

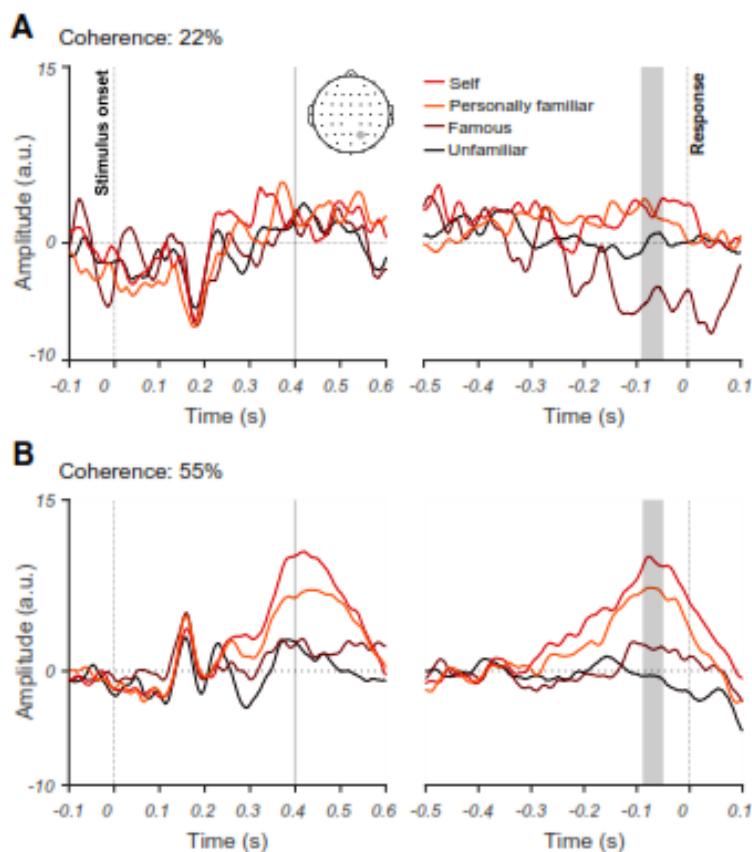
218 Is there a “familiarity spectrum” for faces in the brain?

219 Our behavioral results showed that there is a graded increase in participants'
220 performance as a function of familiarity level - i.e., participants achieve higher
221 performance if the faces are more familiar to them. In this section we address the first
222 question of this study about whether we can find a familiarity spectrum in neural
223 activations, using both the traditional univariate and novel multi-variate analyses of
224 EEG.

225

226 Event-related potentials reflect behavioral familiarity effects

227 As an initial, more traditional, pass at the data, we explored how the neural
228 responses were modulated by different levels of familiarity and coherence by averaging
229 event-related potentials (ERP) across participants for different familiarity levels and
230 phase coherences (Figure 2B). This is important as recent work failed to capture
231 familiar face identity information from single electrodes (Ambrus et al., 2019). At high
232 coherence, the averaged ERPs, obtained from a representative centroparietal electrode
233 (CP2), where previous studies have found differential activity for different familiarity
234 levels (Henson et al., 2008; Kaufmann et al., 2009; Huang et al., 2017), demonstrated
235 an early, evoked response, followed by an increase in the amplitude proportional to
236 familiarity levels. This showed that self faces elicited the highest ERP amplitude,
237 followed by personally familiar, famous, and unfamiliar faces (Figure 2B for 55% phase
238 coherence). This observation of differentiation between familiarity levels at later time
239 points seems to support evidence accumulation over time, which is more pronounced at
240 higher coherence levels where the brain had access to reliable information. This repeats
241 previous findings showing differential activity for different levels of face familiarity after
242 200 ms in the post-stimulus onset window (Caharel et al., 2002; Wiese et al., 2019).



243

244 **Figure 2. The effect of familiarity and sensory evidence on event-related potentials (ERPs).**
245 Averaged ERPs for 22% (A) and 55% (B) phase coherence levels and four face categories across all
246 participants for an electrode at a centroparietal site (CP2). Note that the left panels show stimulus-aligned
247 ERPs while the right panel shows response-aligned ERPs. Shaded areas show the time windows when
248 the difference in ERPs between unfamiliar and the average of the three familiar face categories (i.e.
249 unfamiliar-average of unfamiliar categories) were significantly ($p<0.05$) higher in the 55% vs. 22%
250 coherence levels. The significance was evaluated using one-tailed independent t -test with correction for
251 multiple comparisons across time at $p<0.05$. The differences were significant at later stages of stimulus
252 processing around 400 ms post-stimulus onset and <100 ms before the response was given by the
253 participant in the stimulus- and response-aligned analyses, respectively.

254

255 We also observed a similar pattern between the ERPs of different familiarity
256 levels at the time of decision (just before the response was made). Such systematic
257 differentiation across familiarity levels was lacking at the lowest coherence level, where
258 the amount of sensory evidence, and behavioral performance, were low (c.f. Figure 2A
259 for 22% phase coherence; shading areas). We observed a gradual increase in
260 separability between the four face categories when moving from low to high coherence
261 levels (Supplementary Figure 1). The topographic ERP maps (Supplementary Figure 2)

262 show that the effects are not localized on the CP2 electrode, but rather distributed
263 across the head. There are electrodes which seem to show even more familiarity
264 information than the CP2 electrode. These results reveal the neural correlates of
265 perceptual differences in categorizing different familiar face categories under
266 perceptually difficult conditions.

267 Dynamics of neural representation and evidence accumulation for different
268 face familiarity levels

269 Our results so far are consistent with previous event-related studies showing that
270 the amplitude of ERPs is modulated by the familiarity of the face (Henson et al., 2008;
271 Kaufmann et al., 2009; Schweinberger et al., 2002; Huang et al., 2017). However, more
272 modulation of ERP amplitude does not necessarily mean enhanced representation.
273 Moreover, we observed that the familiarity effects were distributed across the head
274 rather than localized only on the individual CP2 electrode (Supplementary Figure 2).
275 Therefore, looking at individual electrodes might overlook the true temporal dynamics of
276 familiarity information, which may involve widespread brain networks (Ramon and
277 Gobbini, 2018; Duchaine and Yovel, 2015). Here we used multivariate pattern and
278 representational similarity analyses on these EEG data to quantify the time course of
279 familiar vs. unfamiliar face processing. Compared to traditional single-channel
280 (univariate) ERP analysis, MVPA allows us to capture the whole-brain widespread and
281 potentially subtle differences between the activation dynamics of different familiarity
282 levels (Ambrus et al., 2019; Dobs et al., 2019). Specifically, we asked: (1) how the
283 representational dynamics of stimulus- and response-related activations change
284 depending on the level of face familiarity; and (2) how manipulation of sensory evidence
285 (phase coherence) affects neural representation and coding of different familiarity
286 levels.

287 To obtain the temporal evolution of familiarity information across time, at each
288 time point we trained the classifier to discriminate between familiar and unfamiliar faces.
289 Note that the mapping between response and fingers were swapped from the first
290 session to the next (counterbalanced across participants) and the data were collapsed

291 across the two sessions for all analyses, which ensures the motor response cannot
292 drive the classifier. We trained the classifier using 90% of the trials and tested it on the
293 left-out 10% of trials using a standard 10-fold cross-validation procedure (see *Methods*).
294 This analysis used only correct trials. Our decoding analysis shows that, up until ~200
295 ms after stimulus onset, decoding accuracy is near chance for all coherence levels
296 (Figure 3A). The decoding accuracy then gradually increases over time and peaks
297 around 500 ms post-stimulus for the highest coherence level (55%) but remains around
298 chance for the lower coherence level (22%, Figure 3A). The accuracy for intermediate
299 coherence levels (30% and 45%) falls between these two bounds but only reaches
300 significance above chance for the 45% coherence level. This ramping up temporal
301 profile suggests an accumulation of sensory evidence in the brain across the time
302 course of stimulus presentation, which has a processing time that depends on the
303 strength of the sensory evidence (Hanks and Summerfield, 2017; Philiastides et al.,
304 2006).

305 After verifying that we could decode the main effect of familiarity, we turned to
306 the first main question of the study. To examine if neural decoding could reveal the
307 spectrum of familiarity which we observed in behavior and ERPs, we separately
308 calculated the decoding accuracy for each of the sub-categories of familiar faces
309 (Figure 3B): unfamiliar, famous, self and personally familiar faces (on the 55%
310 coherence level, which showed the highest decoding in Figure 3A). The decoding
311 accuracy was highest for self faces, both for stimulus- and response-aligned analyses,
312 followed by personally familiar, famous and unfamiliar faces. Accuracy for the response-
313 aligned analysis shows that the decoding gradually increased to peak decoding ~100
314 ms before the response was given by participants. This temporal evolution of decoding
315 accuracy begins after early visual perception and rises in proportion to the amount of
316 the face familiarity.

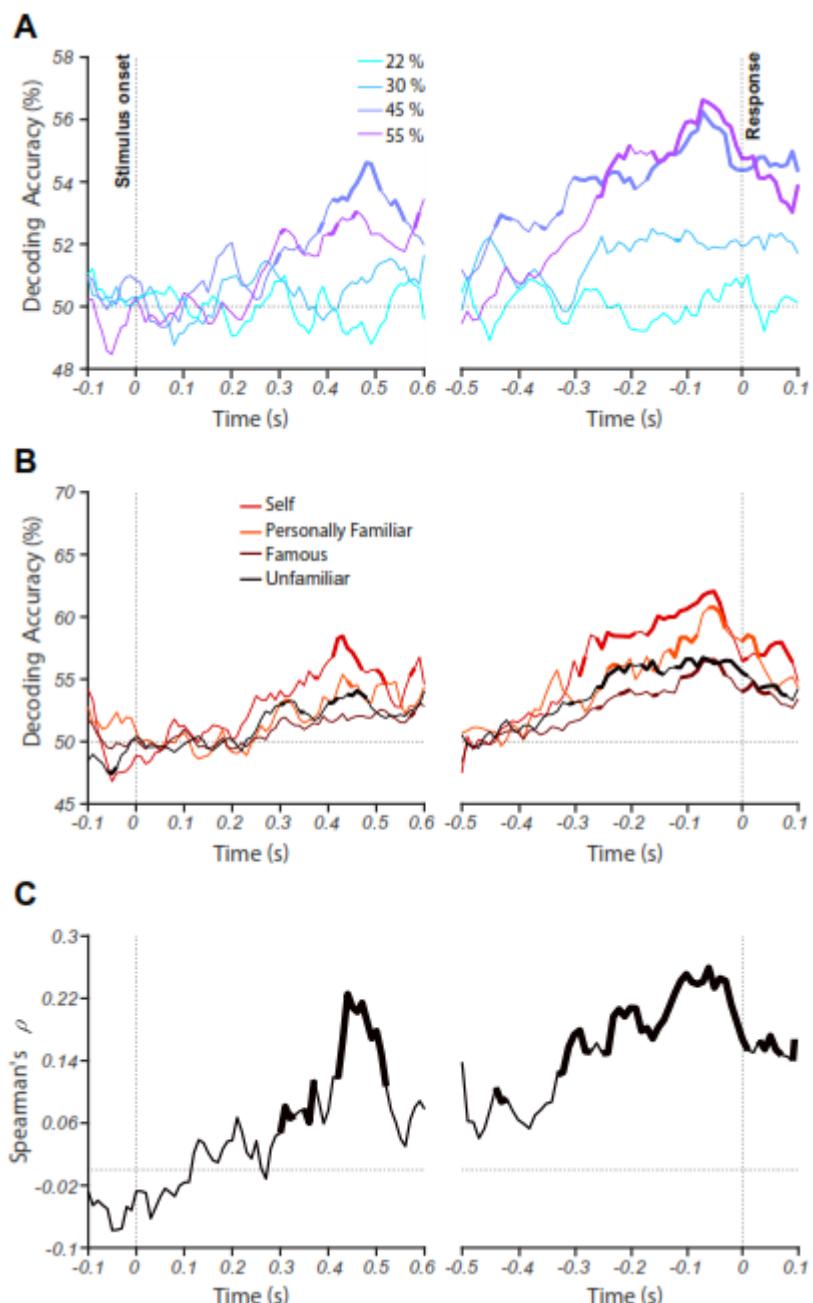
317 To rule out the possibility that an unbalanced number of trials in the sub-
318 categories of familiar faces could lead to the difference in decoding accuracies between
319 the sub-categories, we also repeated the decoding analysis by classifying each familiar
320 subcategory from the unfamiliar category (after equalizing the number of trials across

321 the familiar and unfamiliar categories and also across the three familiar categories),
322 which provided similar results. We also repeated the same analysis for lower coherence
323 levels: only the two high-coherence conditions (i.e. 45% and 55%), but not the low-
324 coherence conditions (i.e. 22% and 30%), showed significantly above-chance decoding
325 for all familiarity conditions (Supplementary Figure 3).

326 Low-level stimulus differences between conditions could potentially drive the
327 differences between categories observed in both ERP and decoding analyses (e.g.,
328 familiar faces being more frontal than unfamiliar faces, leading to images with brighter
329 centers and, therefore, separability of familiar from unfamiliar faces using central
330 luminance of images; Dobs et al., 2019; Ambrus et al., 2019). To address such potential
331 differences, we carried out a supplementary analysis using RSA (Supplementary Text
332 and Supplementary Figures 4 and 5), which showed that such differences between
333 images do not play a major role in the differentiation between categories.

334 To determine whether the dynamics of decoding during stimulus presentation are
335 associated with the perceptual task, as captured by our participants' behavioral
336 performance, we calculated the correlation between decoding accuracy and perceptual
337 performance. For this, we calculated the correlation between 16 data points from
338 decoding accuracy (4 face categories \times 4 phase coherence levels) and their
339 corresponding behavioral accuracy rates, averaged over participants. The correlation
340 peaked ~500 ms post-stimulus (Figure 3C), which was just before the response was
341 given. This is consistent with an evidence accumulation mechanism determining
342 whether to press the button for 'familiar' or 'unfamiliar', which took another ~100 ms to
343 turn into action (finger movement).

344



345

346 **Figure 3. Decoding of face familiarity from EEG signals.** (A) Time course of decoding accuracy for
347 familiar versus unfamiliar faces from EEG signals for four different phase coherence levels (22%, 30%,
348 45%, and 55%). (B) Time course of decoding accuracy for four face categories (i.e., unfamiliar, famous,
349 self and personally familiar faces) from EEG signals at the 55% coherence level. The chance accuracy is
350 50%. Thickened lines indicate the time points when the accuracy was significantly above chance level
351 (sign rank test, FDR corrected across time, $p < 0.05$). (C) Correlation between behavioral performance and
352 decoding accuracy (across all conditions) over time. Thickened lines indicate the time points when the
353 correlation was significant. The left panels show the results for stimulus-aligned analysis while the right
354 panels show the results for response-aligned analysis (averaged over 18 participants).

355

356 Do higher-order peri-frontal brain areas contribute to familiar face
357 recognition?

358 In this section we address the second question of this study about whether peri-
359 frontal brain areas contribute to the recognition of familiar faces in the human brain
360 using a novel informational connectivity analyses on EEG.

361 Task difficulty and familiarity level affect information flow across the brain

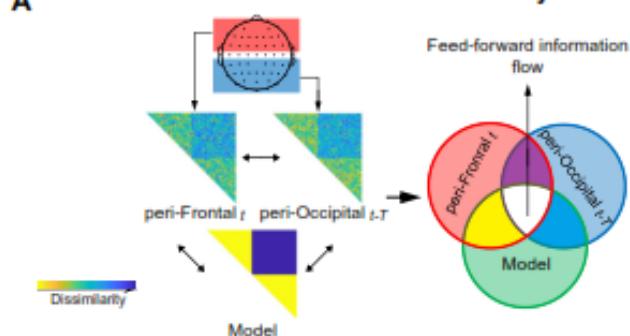
362 We investigated how the dynamics of feed-forward and feedback information flow
363 changes during the accumulation of sensory evidence and the evolution over a trial of
364 neural representations of face images. We developed a novel connectivity method
365 based on RSA to quantify the relationships between the evolution of information based
366 on peri-occipital EEG electrodes and those of the peri-frontal electrodes. As an
367 advantage to previous Granger causality methods (Goddard et al., 2016; Goddard et al.,
368 2019; Karimi-Rouzbahani et al., 2019; Kietzman et al., 2019), the connectivity method
369 developed here allowed us to check whether the transferred signals contained *specific*
370 aspects of *stimulus information*. Alternatively, it could be the case that the transferred
371 signals might carry highly abstract but irrelevant information between the source and
372 destination areas, which can be incorrectly interpreted as connectivity (Anzellotti and
373 Coutanche, 2018; Basti et al., 2020). Briefly, feed-forward information flow is quantified
374 as the degree to which the information from peri-occipital electrodes at present time
375 contributes to the information recorded at peri-frontal electrodes at a later time point,
376 which reflects moving the frontal representation closer to that required for task goals.
377 Feedback flow is defined as the opposite: the contribution to information at peri-frontal
378 electrodes at the present time to that recorded later at peri-occipital electrodes at a later
379 time point (Figure 4A).

380 The results show that at the highest coherence level (55%), information flow is
381 dominantly in the feed-forward direction. This is illustrated by the shaded area in Figure
382 4B where partialling out the peri-frontal from peri-occipital correlations only marginally
383 reduces the total peri-occipital correlation (Figure 4B, black curves and shaded area),

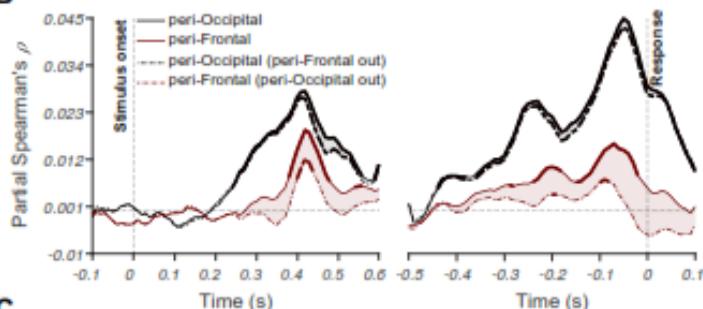
384 meaning that there is limited information transfer from peri-frontal to peri-occipital. In
385 contrast, partialling out the peri-occipital from peri-frontal correlations leads to a
386 significant reduction in peri-frontal correlation, reflecting a feed-forward transfer of
387 information (Figure 4B, brown curves and shaded area). This trend is also seen for
388 response-aligned analysis.

389 These differences are shown more clearly in Figure 4C where the peaks of feed-forward
390 and feedback curves show that the feed-forward information is dominant earlier,
391 followed by feedback information flow, as shown by the later peak of feedback
392 dynamics. These results suggest that when the sensory evidence is high, feed-forward
393 information flow may be sufficient for categorical representation and decision making
394 while feedback only slightly enhances the representation. However, in lower coherence
395 levels (i.e., low sensory evidence), the strength of information flow is either equivalent
396 between feed-forward and feedback directions (30%, 45%) or dominantly feedback
397 (22%, Figure 4D).

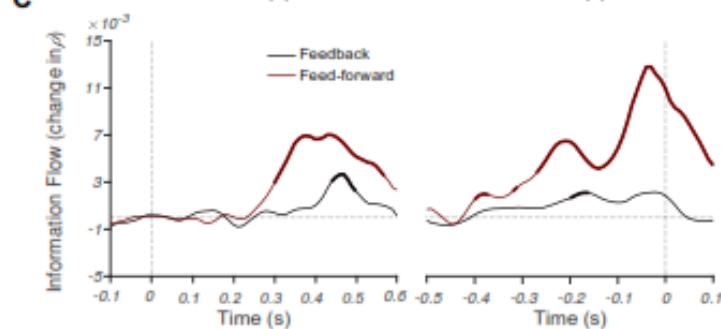
A Feed-forward and feedback flow of familiarity information



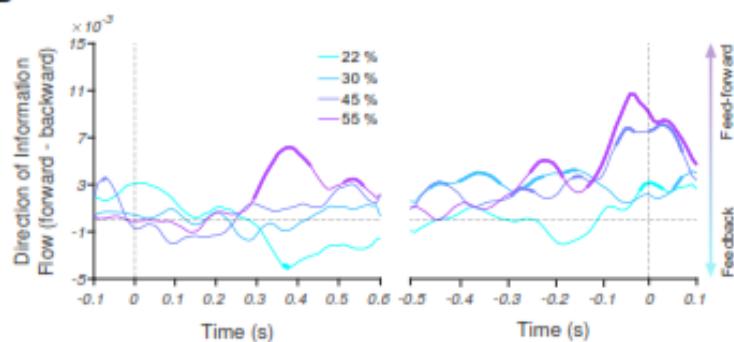
B



C



D



398

399 **Figure 4. Feed-forward and feedback information flow revealed by RSA.** (A) A schematic
400 presentation of the method for calculating informational connectivity between the peri-frontal and peri-
401 occipital electrodes, termed feed-forward and feedback information flow. Feed-forward information flow is
402 calculated as difference of the correlation between the present time peri-frontal neural RDM and the
403 predicted model RDM and the same correlation when the earlier peri-occipital neural RDM is partialled
404 out from it. This is shown in the Venn diagram on the right. The summation of white and yellow areas
405 reflect the correlation between the peri-frontal and the model RDMs while the yellow area reflects the
406 same correlation after partialling out the peri-occipital area at the earlier time point. The difference
407 between the two (i.e. white = (white + yellow) - yellow) is considered to be feed-forward flow of

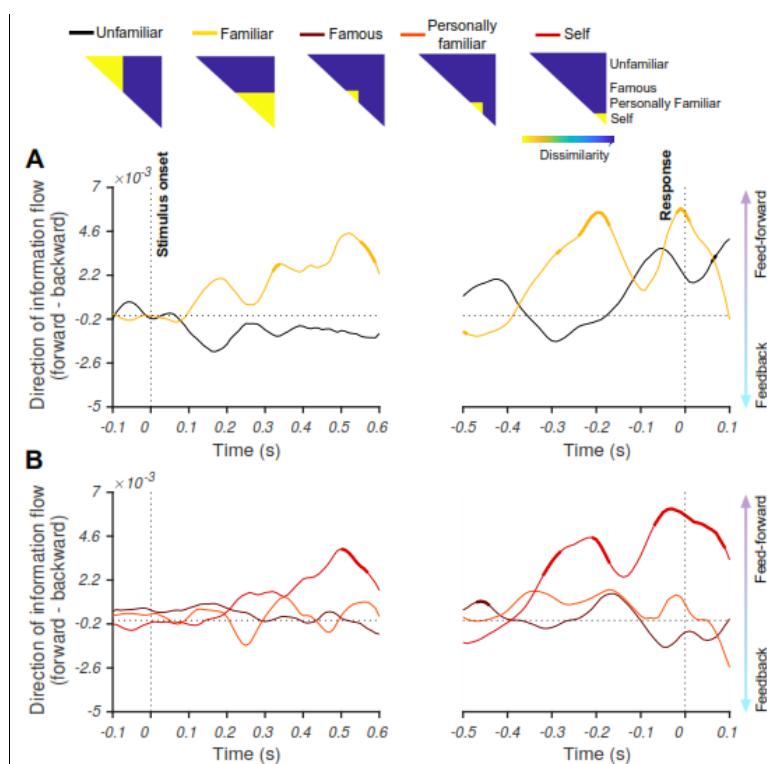
408 information captured by the model. Delay time (T) is 30ms. **(B)** Time course of partial Spearman's
409 correlations representing the partial correlations between the peri-occipital (black) and peri-frontal (brown)
410 EEG electrodes and the model (familiar-unfamiliar model, see the inset in A) while including (solid) and
411 excluding (dashed) the effect of the other area at phase coherence of 55%. The shaded area shows the
412 decline in partial correlation of the current area with the model after excluding (partialling out) the RDM of
413 the other area. Note that in both the dashed and solid lines, the low-level image statistics are partialled
414 out of the correlations, so we call them partial correlations in both cases. **(C)** Feedforward (brown) and
415 feedback (black) information flows obtained by calculating the value of the shaded areas in the
416 corresponding curves in B. **(D)** Direction of information flow for different coherence levels, determined as
417 the difference between feed-forward and feedback information flow showed in C. Thickened lines indicate
418 time points at which the difference is significantly different from zero (sign permutation test and corrected
419 significance level at $p < 0.05$), and black dotted lines indicate 0 correlation. The left panels show the
420 results for stimulus-aligned analysis while the right panels represent the results for response-aligned
421 analysis.

422

423 Here, we can see that the lower sensory evidence correlates with greater
424 engagement of feedback mechanisms, suggesting that feedback is recruited to boost
425 task-relevant information in sensory areas under circumstances where the input is
426 weak. This is consistent with the dynamics and relative contribution of feedback and
427 feed-forward mechanisms in the brain varying with the sensory evidence / perceptual
428 difficulty of the task.

429 Importantly, we also were interested in whether the degree of familiarity changes
430 the direction of information flow between the peri-frontal and peri-occipital brain areas.
431 For this analysis, we collapsed the data across all coherence levels to look specifically
432 at the impact of face familiarity. We generated specific RDM models to evaluate how
433 much information about unfamiliar faces vs. all unfamiliar faces as a group (Figure 5A)
434 and each subcategory of familiar faces (i.e., famous, personally familiar and self; Figure
435 5B) individually were transferred between the two brain areas. To avoid any bias from a
436 different number of elements in the RDM matrices, we only compared equal-sized
437 conditions and present the results in separate panels (i.e. familiar vs. unfamiliar (Figure
438 5A) and subcategories of familiar faces (Figure 5B)). While the unfamiliar category
439 showed a non-significant flow in either direction, the familiar category showed significant
440 feed-forward flow of information in the stimulus-aligned data starting from 300 ms post-
441 stimulus onset (Figure 5A). Among the familiar sub-categories, only the self category
442 showed significant feed-forward information flow starting to accumulate after the
443 stimulus onset, reaching sustained significance ~500 ms. The less familiar categories of

444 famous and personally familiar did not reach significance. In the response-aligned
445 analysis, again, the significant time points show the domination of feed-forward flow for
446 the familiar category (Figure A) but not the unfamiliar category, and the self category but
447 not the other sub-categories of familiar faces (Figure 5B). Together, these results
448 suggest that while the information about the unfamiliar category did not evoke a
449 particular dominance of information flow in either direction, the representations of
450 familiar and self faces showed dominant feed-forward information flow from the peri-
451 occipital to the peri-frontal brain areas. Note that, in this analysis, we also tried to
452 minimize the effect of the participants' decision and motor response in the models by
453 excluding the opposing category (i.e. unfamiliar category when evaluating the familiar
454 models and *vice versa*), which could have potentially contributed to the information
455 flows in the previous analysis (c.f. Figure 4).



456

457 **Figure 5. Directions of information flow for familiar, unfamiliar and different levels of familiarity.**
458 The models, as depicted on the top, are constructed to measure the extent and timing by which
459 information about unfamiliar and familiar (A), and each familiar sub-category (B) moves between the peri-
460 occipital and peri-frontal brain areas. Feed-forward information flow is calculated as difference of the
461 correlation between the present time peri-frontal neural RDM and the predicted model RDM and the same
462 correlation when the earlier peri-occipital neural RDM is partialled out from it. This is shown in the Venn
463 diagram in Figure 4A. The summation of white and yellow areas reflect the correlation between the peri-

464 frontal and the model RDMs while the yellow area reflects the same correlation after partialling out the
465 peri-occipital area at the earlier time point. The difference between the two (i.e. white = (white + yellow) –
466 yellow) is considered to be feed-forward flow of information captured by the model. Delay time (T) is
467 30ms. The yellow areas in the models refer to the target category (including unfamiliar, familiar, famous,
468 personally familiar and self faces). Thickened lines indicate time points at which the difference is
469 significantly different from zero (sign permutation test and corrected for multiple comparisons at
470 significance level of $p < 0.05$), and black horizontal dotted lines indicate 0 correlation. The left panel
471 shows the result for stimulus-aligned analysis while the right panels represent the result for response-
472 aligned analysis.

473

474 Together, the results of the information connectivity analysis suggest that, in
475 familiar face recognition, both top-down and bottom-up mechanisms play a role, with the
476 amount of sensory evidence determining their relative contribution. It also suggests that
477 the degree to which sensory information is processed feed-forward can be modulated
478 by the familiarity level of the stimulus.

479 Discussion

480 This study investigated the neural mechanisms of familiar face recognition. We
481 asked how familiarity affected the contribution of feed-forward and feedback processes
482 in face processing. We first showed that manipulating the familiarity affected the
483 informational content of neural responses about face category, in line with a large body
484 of behavioral literature showing an advantage of familiar over unfamiliar face processing
485 in the brain. Then, we developed a novel method of informational connectivity analysis
486 to track the exchange of familiarity information between peri-occipital and peri-frontal
487 brain areas to see if frontal brain areas contribute to familiar face recognition. Our
488 results showed that when the perceptual difficulty was low (high sensory evidence), the
489 flow of face familiarity information was consistent with a feed-forward account. On the
490 other hand, when the perceptual difficulty was high (low sensory evidence), the
491 dominant flow of face familiarity information reversed, which we interpret as reliance on
492 feedback mechanisms. Moreover, when teasing apart the effect of task and response
493 from neural representations, only the familiar faces, but not the unfamiliar faces,
494 showed a dominance of feed-forward flow of information, with maximum flow for the
495 most familiar category, the self faces.

496 Our results are consistent with the literature suggesting that visual perception
497 comprises both feed-forward and feedback neural mechanisms transferring information
498 between the peri-occipital visual areas and the peri-frontal higher-order cognitive areas
499 (Bar et al., 2006; Summerfield et al., 2006; Goddard et al., 2016; Karimi-Rouzbahani et
500 al., 2017b; Karimi-Rouzbahani et al., 2017c; Karimi-Rouzbahani et al., 2019). However,
501 previous experimental paradigms and analyses did not dissociate feedback and feed-
502 forward information flow in familiar face recognition, and argued for a dominance of
503 feed-forward processing (Dobs et al., 2019; di Oleggio Castello and Gobbini, 2015; Ellis
504 et al., 1979; Young and Burton, 2018). The more nuanced view we present is important
505 because stimulus familiarity, similar to other factors including levels of categorization
506 (superordinate vs. basic level; Besson et al., 2017; Praß et al., 2013), task difficulty
507 (Chen et al., 2008; Woolgar et al., 2015; Kay et al., 2017) and perceptual difficulty (Fan
508 et al., 2020; Hupe et al., 1998; Gilbert and Li, 2013; Gilbert and Sigman, 2007; Lamme
509 and Roelfsema, 2000; Woolgar et al., 2011), may affect the complex interplay of feed-
510 forward and feedback mechanisms in the brain.

511 Our results showed that the contribution of peri-frontal to peri-occipital feedback
512 information was inversely proportional to the amount of sensory evidence about the
513 stimulus. Specifically, we only observed feedback when the sensory evidence was
514 lowest (high perceptual difficulty) in our face familiarity categorization task. Although a
515 large literature has provided evidence for the role of top-down feedback in visual
516 perception, especially when sensory visual information is low, they generally evaluated
517 the feedback mechanisms within the visual system (Ress et al., 2000; Lamme and
518 Roelfsema, 2000; Super et al., 2001; Lamme et al., 2002; Pratte et al., 2013; Fenske et
519 al., 2006; Lee and Mumford, 2003; Felleman et al., 1991; Delorme et al., 2004;
520 Mohsenzadeh et al., 2018; Kietzmann et al., 2019) rather than across the fronto-occipital
521 brain networks (Bar et al., 2006; Summerfield et al., 2006; Goddard et al., 2016; Karimi-
522 Rouzbahani et al., 2018; Karimi-Rouzbahani et al., 2019). Our findings support theories
523 suggesting that fronto-occipital information transfer may feedback (pre-existing) face
524 templates, against which the input faces are compared for correct recognition (Bar et
525 al., 2006; Summerfield et al., 2006). Previous results could not determine the content of
526 the transferred signals (Bar et al., 2006; Summerfield et al., 2006; Goddard et al., 2016;

527 Karimi-Rouzbahani et al., 2018; Karimi-Rouzbahani et al., 2019). Here, using our novel
528 connectivity analyses, we showed that the transferred signal contained information
529 which contributed to the categorization of familiar and unfamiliar faces.

530 Despite methodological differences, our findings support previous human studies
531 showing increased activity in lower visual areas when the cognitive and perceptual
532 tasks were difficult relative to easy, which the authors attributed to top-down
533 contributions (Ress et al., 2000; Kay et al., 2017). However, due to the low temporal
534 resolution of fMRI, these studies cannot show the temporal evolution of these top-down
535 contributions or the validity of the hypothesized direction. Importantly, the observed
536 increase in activity in lower visual areas does not necessarily correspond to the
537 enhancement of neural representations in those areas - increased univariate signal
538 does not show whether there is more information that will support performance.
539 Electrophysiological studies in animals have also shown that cortical feedback
540 projections robustly modulate responses of early visual areas when sensory evidence is
541 low, or the stimulus is difficult to segregate from the background figure (Hupe et al.,
542 1998). A recent study has also found cortical feedback modulated the activity of
543 neurons in the dorsolateral geniculate nucleus (dLGN), which was less consistent when
544 presenting simple vs. complex grating stimuli (Spacek et al., 2019). Therefore, varying
545 perceptual difficulty seems to engage different networks and processing mechanisms,
546 and we show here that this also pertains to faces: less difficult stimuli such as our high-
547 coherence faces seem to be predominantly processed by the feed-forward
548 mechanisms, while more difficult stimuli such as our low-coherence faces recruit both
549 feed-forward and feedback mechanisms. However, the exact location of the feedback in
550 all these studies, including ours, remains to be determined with the development of
551 more accurate modalities for neural activity recording.

552 We observed that the direction of information flow is influenced by the familiarity
553 of the stimulus. The models of familiar faces and self faces, evoked a dominant flow of
554 feed-forward information. The unfamiliar category, however, did not evoke information
555 flow in any direction, as evaluated by our connectivity method. This is consistent with
556 enhanced representations of familiar face categories in the feed-forward pathways

557 (Dobs et al., 2019; di Oleggio Castello and Gobbini, 2015; Ellis et al., 1979; Young and
558 Burton, 2018), which, in turn, requires less top-down contributions to facilitate the
559 perception of relevant information (Bar et al., 2006; Gilbert and Sigman, 2007). Our
560 results might initially seem inconsistent with Fan et al.'s (2020) study, which did not
561 report significant differences between the temporal dynamics of familiar and unfamiliar
562 face representations; however, they only used famous faces within the familiar face
563 spectrum. In our sub-category analysis, we also did not observe differences between
564 famous faces and unfamiliar faces; our main findings were from highly familiar self
565 faces. Overall, then, our results suggest that processing of familiar faces, especially the
566 most familiar (self) faces, is dominated by feed-forward information flow.

567 One assumption in the connectivity analysis of the current work, as in many
568 previous ones (Goddard et al., 2016; Clarke et al., 2018; Kietzman et al., 2019), is that
569 all categories of faces used here involve neural mechanisms from both the peri-frontal
570 and peri-occipital areas. However, this is not necessarily the case; we know from the
571 face recognition literature that peri-frontal brain areas (as defined in this study) play role
572 in the processing of face-relevant information such as social, dynamic and eye-
573 movement-related aspects in cooperation with superior temporal brain areas (Duchaine
574 and Yovel, 2015; superior temporal areas are grouped here in the peri-occipital
575 category). On the other hand, the peri-occipital brain areas have been suggested to
576 dominantly process lower order sensory-level face features with relatively more
577 independence from peri-frontal brain areas (Collins and Olsen, 2014). This suggests
578 that our connectivity analysis might provide a stronger flow for one aspect of information
579 than the other depending on the potentially distinct neural network involved for each.
580 However, to the best of our knowledge, no studies have suggested distinct networks for
581 the processing of the conditions which we compared (familiar vs. unfamiliar faces or
582 familiarity levels). Thus, we cannot rule out the possibility that there might be factors
583 attributable to a subset of categories, but not others, that involve distinct networks. For
584 example, it could be the case that familiar faces, but not unfamiliar ones, involve
585 emotion networks which span from the posterior to the anterior brain areas (Leppänen
586 and Nelson, 2009). To avoid this potential influence, we selected images for both
587 familiar and unfamiliar categories with variable emotional content, but any emotional

588 content associated with basic familiarity could not be avoided. At this point, we interpret
589 our results as an interaction between feed-forward and feedback sweeps of information
590 within a network, but acknowledge the potential contribution of additional frontal areas
591 for one category over another.

592 Our results suggest that processing differs considerably for highly familiar faces.
593 This may be because expectation and prediction play a role in (Ramon and Gobbini,
594 2018; Summerfield and Egner, 2009), and can potentially affect the contribution of
595 feedback neural mechanisms in face detection (Summerfield et al., 2006). Specifically,
596 familiar faces are generally more limited in number compared to unfamiliar faces, which
597 can potentially make the former more predictable. However, according to the earlier
598 visual recognition literature (Bar et al., 2006; Summerfield et al., 2006), if anything, this
599 would have evoked more pronounced feedback signals for the familiar faces vs.
600 unfamiliar faces in this study. In contrast to this prediction, our results showed dominant
601 feed-forward flow of information for familiar faces, and no significant flow in either
602 direction for unfamiliar faces. Therefore, it seems unlikely that the potential difference in
603 expectation between familiar and unfamiliar categories could explain our information
604 flow results.

605 Results also show that, in lower coherence levels, the information about the
606 familiarity levels was generally stronger than the information about familiarity itself (as
607 captured by familiar-unfamiliar model RDM; Supplementary Figure 4). This suggests a
608 lower threshold for the appearance of familiarity level compared to familiar-unfamiliar
609 representations, which are differentially developed through life-time experience and task
610 instructions, respectively. Specifically, development of neural representations reflecting
611 familiarity levels could be a result of exposure to repetitive faces, which can lead to
612 developing face-specific representations in the visual system (Dobs et al., 2019), while
613 task instructions could temporarily enhance the processing of relevant information in the
614 brain through top-down mechanisms (Hebart et al., 2018; Karimi-Rouzbahani et al.,
615 2019). This is probably the reason for the dominance of feedback information flow in the
616 processing of familiarity information (Figure 5A).

617 The RSA-based connectivity method used in this study follows a recent shift
618 towards informational brain connectivity methods (Anzellotti and Coutanche, 2018; Basti
619 et al., 2020; Keitzmann et al., 2019; Goddard et al., 2016; Clarke et al., 2018; Karimi-
620 Rouzbahani, 2018; Karimi-Rouzbahani et al., 2019; Karimi-Rouzbahani et al., 2020a),
621 and introduces a few distinct features compared to previous methods of connectivity
622 analyses. Specifically, traditional connectivity methods examine inter-area interactions
623 through indirect measures such as gamma-band synchronization (Gregoriou et al.,
624 2009), shifting power (Bar et al., 2006) or causality in the activity patterns (Summerfield
625 et al., 2006; Fan et al., 2020). Such connectivity methods consider simultaneous (or
626 time-shifted) correlated activations of different brain areas as connectivity, but they are
627 unable to examine how (if at all) relevant information is transferred across those areas.
628 Goddard et al. (2016) developed an RSA-based Granger connectivity method to solve
629 this issue, which allowed us and others to track the millisecond transfer of stimulus
630 information across peri-frontal and peri-occipital brain areas (Karimi-Rouzbahani, 2018;
631 Karimi-Rouzbahani et al., 2019; Goddard et al., 2019). This was followed by another
632 informational connectivity method, which was similar but used regression instead of
633 correlation in implementation (Keitzmann et al., 2019). While informative, these
634 methods, are silent about what aspects of the representation are transferred and
635 modulated. In other words, we need new methods to tell how (if at all) the transferred
636 information is contributing to the representations in the destination area. Not having
637 access to the transferred contents could lead to incorrect interpretations of connectivity
638 for one main reason: we would not be able to tease apart transactions of distinct types
639 of information across areas (e.g. familiar-unfamiliar discrimination, or different levels of
640 familiarity). To address this issue, one could simply calculate the correlation between
641 the neural and model RDMs from the source and destination areas at every time point
642 and then calculate the Granger causality between the two time-courses of correlations.
643 This is exactly how Clarke et al., (2018) incorporated RDM models into their connectivity
644 to track specific aspects of the transferred information. However, this last method loses
645 the temporal dynamics of information flow in the calculation of Granger causality, and
646 only provides the direction of information flow. Our method circumvents this limitation
647 (i.e. lack of temporal information) by making use of the high-dimensional

648 representational space of the RDMs in the source and destination areas for the
649 calculation of inter-area and area-model relationship leaving the time samples available
650 for the evaluation of the temporal dynamics. Our method allows us to explicitly
651 determine the content (using model RDMs), the direction (using delayed time samples
652 across areas) and the temporal evolution (using temporally-resolved analysis) of the
653 information transferred from the peri-frontal to peri-occipital areas and vice versa. The
654 relevance of the transferred information is determined by the amount that the
655 representations in the destination area are shifted towards our predefined predicted
656 RDM models. In this way, we could determine the temporal dynamics of the contributory
657 element of the transferred information. Despite the specificity that the model-based
658 methods (including our proposed one) provide about the content of the transferred
659 information, such model-based methods have the characteristic to ignore other model-
660 irrelevant aspects of information which might be similarly or distinctly represented in the
661 source and destination areas. In other words, while the source and destination areas
662 might show high levels of connectivity through the “lens” of the model used, their
663 representational geometry (as evaluated here using RDMs) might be very distinct from
664 one another when directly compared or vice versa. Therefore, the results of model-
665 based connectivity methods do not make any prediction about the direction and the
666 amount of potential connectivity when using model-free connectivity methods.

667 Despite the advantage that informational connectivity methods provide over
668 conventional univariate connectivity methods, further investigations (using simulated
669 datasets with known ground-truth of information flow) are needed to fully uncover their
670 characteristics and potential limitations. As an initial step in that direction, we simulated
671 a simplified well-controlled dataset and applied our connectivity analysis to it to check if
672 it could detect the imposed information flow between our simulated source and
673 destination areas (Supplementary Figure 6). Results showed that our connectivity
674 analysis detected correct direction and temporal dynamics of the simulated information
675 flow. Despite this successful simulation, a full mathematical and analytical investigation
676 will need to be performed to compare the available and the proposed informational
677 connectivity analyses in the future.

678 Our results specify the neural correlates for the behavioral advantage in
679 recognizing more vs. less familiar faces in a “familiarity spectrum”. As in previous
680 studies, our participants were better able to categorize highly familiar than famous or
681 unfamiliar faces, especially in low-coherence conditions (Kramer et al., 2018; Young
682 and Burton, 2018). This behavioral advantage could result from long-term exposure to
683 variations of personally familiar faces under different lighting conditions and
684 perspectives, which is usually not the case for famous faces.

685 Our neural decoding results quantified a neural representational advantage for
686 more familiar faces compared to less familiar ones (i.e. higher decoding for the former
687 than the latter) to suggest that more familiar faces also lead to more distinguishable
688 neural representations. Decoding accuracy was also proportional to the amount of
689 sensory evidence: the higher the coherence levels, the higher the decoding accuracy.
690 We observed that the decoding accuracy “ramped-up” and reached its maximum ~100
691 ms before participants expressed their decisions using a key press. These results are
692 suggestive of sensory evidence accumulation and decision making processes during
693 face processing in humans, consistent with previously reported data in monkey and
694 recent single-trial ERP studies (Kelly et al., 2013; Hanks and Summerfield, 2017;
695 Philiastides et al., 2006; Philiastides and Sajda, 2006; Shadlen and Newsome, 2001).

696 There was a significant correlation between MVPA accuracy and our behavioral
697 results, showing a relationship between neural representation and behavioral outcomes.
698 While it would be ideal to see perfect correlation between neural data and behavior, it is
699 not usually the case (Dobs et al., 2019), which may reflect several reasons including the
700 noise in the neural data and sub-optimal decoding of the neural codes (Karimi-
701 Rouzbahani et al., 2020b) and/or possible non-linear relationships between neural data
702 and behavior. In our study, while there was a difference between the neural data from
703 personally familiar and self faces (c.f. Figures 2 and 3), there was no detectable
704 difference in behavior, potentially reflecting a ceiling effect for both categories in above-
705 chance conditions (i.e. coherence levels > 30%). Despite this, the correlation was
706 significant across the four familiarity × four coherence level conditions overall during
707 time windows later in the trial and immediately before the response. This suggests that

708 the behavioral advantages of self and familiar faces and/or having higher sensory
709 evidence (highest coherence) could have been driven by the enhanced neural
710 representations.

711 Previous studies generally show face familiarity modulation during early ERP
712 components such as N170, N250, and P300 (Dobs et al., 2019; Ambrus, 2019; Fan et
713 al., 2020; Henson et al., 2008; Kaufmann et al., 2009; Schweinberger et al., 2002;
714 Huang et al., 2017). In contrast, the time course of our EEG results showed their
715 maximum effects after 400 ms post-stimulus onset. However, these studies typically use
716 event-related paradigms, which evoke initial brain activations peaking at around 200
717 ms, whereas our dynamic masking paradigm releases the information gradually along
718 the time course of the trial. Moreover, the extended (>200ms) static stimulation used in
719 previous studies has been suggested to bias towards domination of feed-forward
720 processing (Goddard et al., 2016; Karimi-Rouzbahani, 2018), because of the co-
721 processing of the incoming sensory information and the recurrence of earlier windows of
722 the same input (Kietzmann et al., 2019; Mohsenzadeh et al., 2018), making it hard to
723 measure feedback. Our paradigm, while providing a delayed processing profile
724 compared to previous studies, avoids this and also slows down the process of evidence
725 accumulation so that it becomes more trackable in time. This does mean, however, that
726 our time courses are not really comparable with previous ERP results.

727 In conclusion, our study demonstrates that the processing of face information
728 involves both feed-forward and feedback flow of information in the brain, and which
729 predominates depends on the strength of incoming perceptual evidence and the
730 familiarity of the face stimulus. Our novel extension of multivariate connectivity analysis
731 methods allowed us to disentangle feed-forward and feedback contributions to
732 familiarity representation. This connectivity method can be applied to study a wide
733 range of cognitive processes, wherever information is represented in the brain and
734 transferred across areas. We also showed that the behavioral advantage for familiar
735 face processing is robustly reflected in neural representations of familiar faces in the
736 brain and can be quantified using multivariate pattern analyses. These new findings and

737 methods emphasize the importance of, and open new avenues for, exploring the impact
738 of different behavioral tasks on the dynamic exchange of information in the brain.

739 Materials and Methods

740 Participants

741 We recorded from 18 participants (15 male, aged between 20-26 years, all with
742 normal or corrected-to-normal vision). Participants were students from the Faculty of
743 Mathematics and Computer Science at the University of Tehran, Iran. All participants
744 voluntarily participated in the experiments and gave their written consent prior to
745 participation. All experimental protocols were approved by the ethical committee of the
746 University of Tehran. All experiments were carried out in accordance with the guidelines
747 of the Declaration of Helsinki.

748

749 Stimuli

750 We presented face images of four categories, including unfamiliar, famous, self
751 and personally familiar faces. The unfamiliar faces (n=120) were unknown to
752 participants. The famous faces (n=40) were pictures of celebrities, politicians, and other
753 well-known people. These faces were selected from different, publicly available face
754 databases². In both categories, half of the images were female, and half were male. To
755 ensure that all participants knew the famous face identities, participants completed a
756 screening task prior to the study. In this screening, we presented them with the names
757 of famous people in our data set and asked if they were familiar with the person.

758 The personally familiar faces were selected from participants' family, close
759 relatives, and friends (n=40); self-images were photographs of participants (n=40). The
760 images of self and personally familiar faces were selected to have varied backgrounds

² <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>
<https://megapixels.cc/datasets/msceleb/>

761 and appearances. On average, we collected n=45 for personally familiar and n=45 for
762 self faces for every individual participant. All images were cropped to have 400x400
763 pixels and were converted to greyscale (Figure 1A). We ensured that spatial frequency,
764 luminance, and contrast were equalized across all images. The magnitude spectrum of
765 each image was adjusted to the average magnitude spectrum of all images in our
766 database³.

767 The phase spectrum was manipulated to generate noisy images characterized by
768 their percentage phase coherence (Dakin et al., 2002). We used a total of four different
769 phase coherence values (22%, 30%, 45%, and 55%), chosen based on behavioral pilot
770 experiments, so overall behavioral performance spanned the psychophysical dynamic
771 range. Specifically, the participants scored 52.1%, 64.7%, 85.2% and 98.7% in the
772 mentioned coherence levels in the piloting. At each of the four phase coherence levels,
773 we generated multiple frames for every image: the number of frames generated
774 depended on the reaction time of the participants, as explained below. Different sets of
775 participants were used for the actual and pilot experiments.

776

777 EEG acquisition and Apparatus

778 We recorded EEG data from participants while they were performing the face
779 categorization task. EEG data were acquired in an electrostatically shielded room using
780 an ANT Neuro Amplifier (eego 64 EE-225) from 64 Ag/AgCl scalp electrodes and from
781 three periocular electrodes placed below the left eye and at the left and right outer
782 canthi. All channels were referenced to the left mastoid with input impedance <15k and
783 chin ground. Data were sampled at 1000 Hz and a software-based 0.1-200 Hz
784 bandpass filter was used to remove DC drifts, and high-frequency noise and 50 and 100
785 Hz (harmonic) notch filters were applied to minimize line noise. These filters were
786 applied non-causally (using MATLAB `filtfilt`) to avoid phase-related distortions. We used
787 Independent Component Analysis (ICA) to remove artefactual components in the signal.

³ https://github.com/Masoud-Ghodrati/face_familiarity

788 The components which were reflecting artefactual signals (eye movements, head
789 movements) were removed based on ADJUST's criteria (Mognon et al., 2011). Next,
790 trials with strong eye movement or other movement artifacts were removed using visual
791 inspection. On average, we kept $98.74\% \pm 1.5\%$ artifact-free trials for any given
792 condition.

793 We presented images on LCD monitor (BenQ XL2430, 24", 144 Hz refresh rate,
794 resolution of 1920 \times 1080 pixels) and the stimulus presentation was controlled using
795 custom-designed MATLAB codes and Psychtoolbox 3.0 (Brainard, 1997; Pelli, 1997).
796 We presented stimuli at a distance of 60 cm to the participant, and each image
797 subtended $8^\circ \times 8^\circ$ of visual angle.

798

799 Procedure

800 Participants performed a familiar vs. unfamiliar face categorization task by
801 categorizing dynamically updating sequences of either familiar or unfamiliar face images
802 in two recording sessions (Figure 1A). Image sequences were presented in rapid serial
803 visual presentation (RSVP) fashion at a frame rate of 60 Hz frames per second (i.e.
804 16.67 ms per frame without gaps). Each trial consisted of a single sequence of up to 1.2
805 seconds (until response) with a series of images from the same stimulus (i.e., selection
806 from either familiar or unfamiliar face categories) at one of the four possible phase
807 coherence levels. Importantly, within each phase coherence level, the overall amount of
808 noise remained unchanged, whereas the spatial distribution of the noise varied across
809 individual frames such that different parts of the underlying image was revealed
810 sequentially.

811 We instructed participants to fixate at the center of the monitor and respond as
812 accurately and quickly as possible by pressing one of two keyboard keys (left and right
813 arrow keys) to identify the image as familiar or unfamiliar using the right index and
814 middle fingers, respectively. The mapping between familiar-unfamiliar categories and
815 the two fingers were swapped from the first session to the next (counterbalanced across

816 participants) and the data were collapsed across the two sessions before analyses. As
817 soon as a response was given, the RSVP sequence stopped, followed by an inter-trial
818 interval of 1–1.2 s (random with uniform distribution). The maximum time for the RSVP
819 sequence was 1.2 secs. If participants failed to respond within the 1.2 secs period, the
820 trial was marked as a no-choice trial and was excluded from further analysis. We had a
821 total of 240 trials (i.e., 30 trials per perceptual category, familiar and unfamiliar, each at
822 four phase coherence levels) during the experiment. Participants were naïve about the
823 number and proportion of the face stimuli in categories. We presented six blocks of 36
824 trials each, and one block of 24 trials and participants had some resting time between
825 the blocks. Each image from the image set was presented to the participants once in
826 each session.

827

828 Analysis

829 Decoding (MVPA) analysis

830 We decoded the information content of our conditions using Multivariate Pattern
831 Analysis (MVPA) methods with Support Vector Machine (SVM) classifiers (Cortes et al.,
832 1995). MVPA utilizes within-condition similarity of trials and their cross-condition
833 dissimilarity to determine the information content of individual conditions. We trained an
834 SVM classifier on the patterns of brain activity (from 64 EEG electrodes) from 90% of
835 familiar (including personally familiar, famous, and self sub-categories) and 90% of
836 unfamiliar trials, and then tested the trained classifier on the left-out 10% of trials from
837 each category. The classification accuracy from categorization of the testing data shows
838 whether there is information about familiarity in the neural signal. We only used the trials
839 in which the participant *correctly* categorized the stimulus as familiar or unfamiliar. We
840 repeated this procedure iteratively 10 times until all trials from the two categories were
841 used in the testing of the classifier once (no trial was included both in the training and
842 testing sets in a single run), hence 10-fold cross-validation, and averaged the
843 classification accuracy across the 10 validation runs for each participant. To obtain the

844 decoding accuracy through time, we down-sampled the EEG signals to 100 Hz and
845 repeated the same classification procedure for every 10 ms time point from -100 to 600
846 ms relative to the onset of the stimulus, and from -500 to 100 ms relative to the
847 response. This allowed us to assess the evolution of face familiarity information relative
848 to the stimulus onset and response separately.

849 To investigate the potential differences in the temporal evolution of the sub-
850 categories contained in the familiar category (i.e., famous, personally familiar and self),
851 we additionally calculated the decoding accuracy for each sub-category separately.
852 Note that the same decoding results obtained from decoding of familiar vs. unfamiliar
853 categories were used here, only calculated separately for each sub-category of familiar
854 faces. Finally, we averaged the decoding accuracies across participants and reported
855 the group-level results.

856 We used random bootstrapping testing to evaluate the significance of the
857 decoding accuracies at every time point for the group of participants. For every time
858 point, this involved randomizing the labels of the familiar and unfamiliar trials 10,000
859 times and obtaining 10,000 decoding accuracies using the above procedure for each
860 participant. Then we averaged the 10,000 decoding accuracies across (18) participants
861 obtaining a single decoding accuracy for each of the 10,000 randomization for group-
862 level analysis. For every time point, the p-value of the true group-averaged decoding
863 accuracy was obtained as $[1 - p(10,000 \text{ randomly generated decoding accuracies which}$
864 $\text{were surpassed by the corresponding true group-averaged decoding value})]$. Since
865 there is a different number of trials in each familiar sub-category, in the random
866 bootstrapping, we maintained the same proportion of trials in each sub-category to
867 preserve the original structure and generate an appropriate null distribution. We then
868 corrected the p values for multiple comparisons across time (using MATLAB's mafdr
869 function at $p < 0.05$). After the correction, the true decoding values with $p < 0.05$ were
870 considered significantly above chance (e.g., 50%).

871

872 Brain-behavior correlation

873 To investigate if the decoding results could explain the observed behavioral face
874 categorization results, we calculated the correlation between the decoding and the
875 behavioral results using Spearman's rank correlation. We calculated the correlation
876 between a 16-element vector containing each participant's behavioral accuracy for the
877 four coherence levels of the four familiarity levels (i.e. Familiar, Famous, Self and
878 Unfamiliar), and another vector with the same structure containing the decoding values
879 from the same conditions (Karimi-Rouzbahani et al., 2020b). We repeated this
880 procedure for every time point and each individual participant separately. Finally, we
881 averaged the correlations across participants and reported the group-level results.

882 To determine the significance of the group-averaged correlations, the same
883 bootstrapping procedure as described above was repeated at every time point by
884 generating 10,000 random correlations after shuffling the elements of the 16-element
885 behavioral vector. We repeated this procedure for every time point and each individual
886 participant separately. Then we averaged the 10,000 random correlations across (18)
887 participants obtaining a single correlation value for each of the 10,000 randomization for
888 group-level analysis. For every time point, the *p*-value of the true group-averaged
889 correlation was obtained as [1- *p*(10,000 randomly generated correlations which were
890 surpassed by the corresponding true group-averaged correlation)]. We then corrected
891 the *p* values for multiple comparisons across time (using MATLAB's mafdr function at
892 *p*<0.05). After the correction, the true correlation values with *p* < 0.05 were considered
893 significantly above chance (i.e., 0).

894

895 Representational similarity analysis

896 Representational similarity analysis is used here for three purposes. First, to
897 partial out the possible contributions of low-level image statistics to our decoding
898 results, which is not directly possible in the decoding analysis (Supplementary Text).
899 Second, to investigate possible coding strategies that the brain might have adopted

900 which could explain our decoding, specifically, whether the brain was coding familiar
901 versus unfamiliar faces, the different levels of familiarity or a combination of the
902 superordinate and subordinate categories. Third, to measure the contribution of
903 information from other brain areas to the representations of each given area (see
904 Information flow analysis).

905 We constructed neural representational dissimilarity matrices (RDMs) by
906 calculating the (Spearman's rank) correlation between every possible representation
907 obtained from every single presented image which resulted in a *correct* response
908 (leading to a 240 by 240 RDM matrix if all images were categorized correctly, which was
909 never the case for any participant). The matrices were constructed using signals from
910 the electrodes over the whole brain as well as from peri-occipital and peri-frontal
911 electrodes separately as explained later (Figures 4-6). We also constructed *image*
912 RDMs for which we calculated the correlations between every possible pair of images
913 which had generated the corresponding neural representations used in the neural
914 RDMs (i.e. only from *correct* trials). Finally, to evaluate how much the neural RDMs
915 coded the familiar vs. unfamiliar faces, familiar and unfamiliar faces separately,
916 familiarity levels and each level of familiarity, we constructed different model RDMs. For
917 examples, in the *Familiar-Unfamiliar* model RDM, the elements which corresponded to
918 the correlations of familiar with familiar, or unfamiliar with unfamiliar, representations
919 (and not their cross-correlations) were valued as 1, and the elements which
920 corresponded to the cross-correlations between familiar and unfamiliar faces were
921 valued as 0. The *Familiarity level* model, on the other hand, was filled with 0s (instead of
922 1s) for the representations which corresponded to the cross-correlations between
923 different sub-categories of familiar faces (e.g. personally familiar vs. famous) with
924 everything else being the same as the *Familiar-Unfamiliar* model RDM. Please note that
925 the number of trials within all conditions of the RDM were down-sampled to the
926 minimum number available for all conditions. This avoided potential difference across
927 conditions as a result of unbalanced number of trials across conditions. To correlate the
928 RDMs, we selected and reshaped the upper triangular elements of the RDMs (excluding
929 the diagonal elements) into vector RDMs (or RDVs). To evaluate the correlation
930 between the neural RDVs and the model RDVs, we used Spearman's partial correlation

931 in which we calculated the correlation between the neural and the model RDV while
932 partialling out the image RDV as in equation (1):

933
$$\rho_{NM.I}(t) = \frac{\rho_{NM}(t) - \rho_{NI}(t)\rho_{MI}(t)}{\sqrt{1-\rho_{NI}^2(t)}\sqrt{1-\rho_{MI}^2(t)}} \quad (1)$$

934 where ρ refers to Spearman correlation and $\rho_{NM.I}$ refers to the Spearman correlation
935 between the neural and model RDVs after partialling out the image RDV. N , M and I
936 respectively refer to Neural, Model and Image RDVs. As indicated in the equation, the
937 partial correlation was calculated for every time point of the neural data (10 ms time
938 steps), relative to the stimulus onset and response separately using the time-invariant
939 model and image RDVs. To evaluate the significance of the partial correlations, we
940 used a similar bootstrapping procedure as was used in decoding. However, here we
941 randomized the elements of the model RDV 10,000 times (while keeping the number of
942 ones and zeros equal to the original RDV) and calculated 10,000 random partial
943 correlations. Finally, we compared the true partial correlation at every time point with the
944 randomly generated partial correlations for the same time point and deemed it
945 significant if it exceeded 95% of the random correlations ($p < 0.05$) after correcting for
946 multiple comparisons.

947

948 Informational connectivity analysis

949 We developed a novel model-based method of information flow analysis to
950 investigate how earlier information content of other brain areas contributes to the
951 present-time information content of a given area. While several recent approaches have
952 suggested for information flow analysis in the brain (Goddard et al., 2016; Karimi-
953 Rouzbahani, 2018; Karimi-Rouzbahani et al., 2019), following the recent needs for
954 these approaches in answering neuroscience questions (Anzellotti and Coutanche,
955 2018), none of the previously developed methods could answer the question of whether
956 the transferred information was improving the representation of the target area in line
957 with the behavioral task demands. Our proposed model, however, explicitly incorporates

958 the specific aspects of behavioral goals or stimuli in its formulation and allows us to
959 measure if the representations of target areas are shifted towards the behavioral/neural
960 goals by the received information. An alternative would be that the incoming information
961 from other areas are just epiphenomenal and are task-irrelevant. This new method can
962 distinguish these alternatives.

963 Accordingly, we split the EEG electrodes in two groups, each with 16 electrodes:
964 peri-frontal and peri-occipital (Figure 4A) to see how familiarity information is (if at all)
965 transferred between these areas that can be broadly categorized as “cognitive” and
966 “sensory” brain areas, respectively. We calculated the neural RDMs for each area
967 separately and calculated the correlation between the neural RDV and the model RDV,
968 partialling out the image RDM from the correlation (as explained in equation (1)). This
969 resulted in a curve when calculating the partial correlation at every time point in 10 ms
970 intervals (see the solid lines in Figure 4B). Note that the partial correlation curve for the
971 peri-frontal area could have received contributions from the present and earlier
972 representations of the same area (i.e., the latter being imposed by our sequential
973 stimulus presentation). It could also have received contributions from earlier peri-
974 occipital representations through information flow from peri-occipital to the peri-frontal
975 area. To measure this potential contribution, we partialled out the earlier peri-occipital
976 representations in calculation of the partial correlation between peri-frontal and model
977 RDVs and calculated the difference between the former and the latter partial
978 correlations as feed-forward information flow according to equation (2):

$$979 \text{Feed-forward information flow } (t) = \rho_{FM,I}(t) - \frac{\rho_{FM,I}(t) - \rho_{FO,I}(t-T)\rho_{MO,I}(t-T)}{\sqrt{1-\rho_{FO,I}^2(t-T)}\sqrt{1-\rho_{MO,I}^2(t-T)}} \quad (2)$$

980 where $\rho_{FM,I}$ refers to the partial correlation between the peri-frontal and the model RDV,
981 $\rho_{FO,I}$ the partial correlation between peri-frontal and peri-occipital RDVs and $\rho_{MO,I}$ the
982 partial correlation between the peri-occipital and model RDVs. Please note that the
983 image RDV is partialled out from all pairwise correlations to remove its effect in the
984 analysis, so the subscript I and the term “partial”. This determines the contribution of
985 earlier peri-occipital representations to the present peri-frontal areas which we called

986 “feed-forward information flow” (as indicated by the brown shades in Figure 4). To
987 determine the contribution of the peri-frontal representations in modulating the peri-
988 occipital representations, we used equation (3):

989
$$\text{Feedback information flow } (t) = \rho_{OM,I}(t) - \frac{\rho_{OM,I}(t) - \rho_{OF,I}(t-T)\rho_{MF,I}(t-T)}{\sqrt{1-\rho_{OF,I}^2(t-T)}\sqrt{1-\rho_{MF,I}^2(t-T)}} \quad (3)$$

990 with the same notations as in equation (2). Accordingly, equation (3) determines the
991 contribution of earlier peri-frontal representations in directing the peri-occipital
992 representations towards the model RDV, namely 'feedback information flow'. In
993 equations (2) and (3), the delay time (T) was 30ms, which was selected based on
994 previously reported delay times between the peri-occipital and peri-frontal areas in
995 visual processing (Foxe and Simpson, 2002). To that end, five earlier RDVs were
996 averaged (5 time points centered on -30ms) leading to an average delay time of 30ms.

997 Finally, to characterize the information flow dynamics between the peri-occipital
998 and peri-frontal areas, we calculated the difference between the feed-forward and
999 feedback contribution of information flows. This allowed us to investigate the transaction
1000 of targeted information between the brain areas aligned to the stimulus onset and
1001 response. We repeated the same procedure using the Familiar-Unfamiliar as well as
1002 Familiarity level models to see if they differed. We validated the proposed informational
1003 connectivity method using simulated well-controlled dataset (Supplementary Figure 6).
1004 We determined the significance of the partial correlations using the above-explained
1005 random bootstrapping procedure. We determined the significance of the differences
1006 between partial correlations (the shaded areas in Figure 4 and the lines in panel C) and
1007 the differences in the feed-forward and feedback contribution of information using
1008 Wilcoxon's signed-rank test using $p < 0.05$ threshold for significance after correction for
1009 multiple comparisons (using Matlab mafdr).

1010 Acknowledgements

1011 We would like to thank Ali Yoonessi for supporting our electroencephalography data
1012 collection in his lab. We would also like to appreciate Chris I Baker, Mark Williams, Erin

1013 Goddard and Hamed Nili for providing their valuable feedback on the manuscript. This
1014 research was funded by UK Royal Society's Newton International Fellowship
1015 NIF\R1\192608 to H.K.R. and MRC intramural funding SUAG/052/G101400 to A.W.

1016 References

1017 Ambrus, Géza Gergely, Daniel Kaiser, Radoslaw Martin Cichy, and Gyula Kovács.
1018 2019. "The Neural Dynamics of Familiar Face Recognition." *Cerebral Cortex* 29 (11):
1019 4775–4784.

1020 Anzellotti, Stefano, and Marc N. Coutanche. 2018. "Beyond Functional Connectivity:
1021 Investigating Networks of Multivariate Representations." *Trends in Cognitive Sciences*
1022 22 (3): 258–269.

1023 Bar, Moshe, Karim S. Kassam, Avniel Singh Ghuman, Jasmine Boshyan, Annette M.
1024 Schmid, Anders M. Dale, Matti S. Hämäläinen, Ksenija Marinkovic, Daniel L. Schacter,
1025 and Bruce R. Rosen. 2006. "Top-down Facilitation of Visual Recognition." *Proceedings
1026 of the National Academy of Sciences* 103 (2): 449–454.

1027 Basti, Alessio, Hamed Nili, Olaf Hauk, Laura Marzetti, and Richard Henson. 2020.
1028 "Multi-dimensional Connectivity: A Conceptual and Mathematical Review." *Neuroimage*
1029 117179.

1030 Besson, Gabriel, Gladys Barragan-Jason, Simon J. Thorpe, Michèle Fabre-Thorpe,
1031 Sébastien Puma, Mathieu Ceccaldi, and Emmanuel J. Barbeau. 2017. "From Face
1032 Processing to Face Recognition: Comparing Three Different Processing Levels."
1033 *Cognition* 158: 33–43.

1034 Brainard, David H. 1997. "The Psychophysics Toolbox." *Spatial Vision* 10 (4): 433–436.

1035 Brown, M. W., and P. J. Banks. 2015. "In search of a recognition memory engram."
1036 *Neuroscience & Biobehavioral Reviews* 50: 12-28.

1037 Caharel, Stephanie, Stephane Poiroux, Christian Bernard, Florence Thibaut, Robert
1038 Lalonde, and Mohamed Rebai. 2002. "ERPs associated with familiarity and degree of
1039 familiarity during face recognition." *International Journal of Neuroscience* 112(12): 1499-
1040 1512.

1041 Chen, Yao, Susana Martinez-Conde, Stephen L. Macknik, Yulia Bereshpolova, Harvey
1042 A. Swadlow, and Jose-Manuel Alonso. 2008. "Task Difficulty Modulates the Activity of
1043 Specific Neuronal Populations in Primary Visual Cortex." *Nature Neuroscience* 11 (8):
1044 974.

1045 Clarke, Alex, Barry J. Devereux, and Lorraine K. Tyler. 2018. "Oscillatory Dynamics of
1046 Perceptual to Conceptual Transformations in the Ventral Visual Pathway." *Journal of*
1047 *Cognitive Neuroscience* 30 (11): 1590–1605.

1048 Collins, Elliot, Amanda K. Robinson, and Marlene Behrmann. 2018. "Distinct Neural
1049 Processes for the Perception of Familiar versus Unfamiliar Faces along the Visual
1050 Hierarchy Revealed by EEG." *NeuroImage* 181: 120–131.

1051 Collins, Jessica A., and Ingrid R. Olson. 2014. "Beyond the FFA: the role of the ventral
1052 anterior temporal lobes in face processing." *Neuropsychologia* 61: 65-79.

1053 Dakin, S. C., R. F. Hess, T. Ledgeway, and R. L. Achtman. 2002. "What Causes Non-
1054 Monotonic Tuning of fMRI Response to Noisy Images?" *Current Biology* 12 (14):
1055 R476–R477.

1056 Davies-Thompson, Jodie, and Timothy J. Andrews. 2012. "Intra-and interhemispheric
1057 connectivity between face-selective regions in the human brain." *Journal of*
1058 *Neurophysiology* 108 (11): 3087-3095.

1059 Delorme, Arnaud, Guillaume A. Rousselet, Marc J.-M. Macé, and Michele Fabre-
1060 Thorpe. 2004. "Interaction of Top-down and Bottom-up Processing in the Fast Visual
1061 Analysis of Natural Scenes." *Cognitive Brain Research* 19 (2): 103–113.

1062 Dobs, Katharina, Leyla Isik, Dimitrios Pantazis, and Nancy Kanwisher. 2019. "How Face
1063 Perception Unfolds over Time." *Nature Communications* 10 (1): 1–10.

1064 Duchaine, Brad, and Galit Yovel. 2015. "A revised neural framework for face
1065 processing." *Annual review of vision science* 1: 393-416.

1066 Ellis, Hadyn D., John W. Shepherd, and Graham M. Davies. 1979. "Identification of
1067 Familiar and Unfamiliar Faces from Internal and External Features: Some Implications
1068 for Theories of Face Recognition." *Perception* 8 (4): 431–439.

1069 Ethofer, Thomas, Markus Gschwind, and Patrik Vuilleumier. 2011. "Processing social
1070 aspects of human gaze: a combined fMRI-DTI study." *Neuroimage* 55 (1): 411-419.

1071 Fan, Xiaoxu, Fan Wang, Hanyu Shao, Peng Zhang, and Sheng He. 2020. "The Bottom-
1072 up and Top-down Processing of Faces in the Human Occipitotemporal Cortex." *ELife* 9:
1073 e48764.

1074 Felleman, Daniel J., and DC Essen Van. 1991. "Distributed Hierarchical Processing in
1075 the Primate Cerebral Cortex." *Cerebral Cortex (New York, NY: 1991)* 1 (1): 1–47.

1076 Fenske, Mark J., Elissa Aminoff, Nurit Gronau, and Moshe Bar. 2006. "Top-down
1077 Facilitation of Visual Object Recognition: Object-Based and Context-Based
1078 Contributions." *Progress in Brain Research* 155: 3–21.

1079 Foxe, John J., and Gregory V. Simpson. 2002. "Flow of Activation from V1 to Frontal
1080 Cortex in Humans." *Experimental Brain Research* 142 (1): 139–150.

1081 Gilbert, Charles D., and Mariano Sigman. 2007. "Brain States: Top-down Influences in
1082 Sensory Processing." *Neuron* 54 (5): 677–696.

1083 Gilbert, Charles D., and Wu Li. 2013. "Top-down Influences on Visual Processing."
1084 *Nature Reviews Neuroscience* 14 (5): 350–363.

1085 Gobbini, M. Ida, Ellen Leibenluft, Neil Santiago, and James V. Haxby. 2004. "Social and
1086 Emotional Attachment in the Neural Representation of Faces." *Neuroimage* 22 (4):
1087 1628–1635.

1088 Goddard, Erin, Thomas A. Carlson, and Alexandra Woolgar. 2019. "Spatial and
1089 Feature-Selective Attention Have Distinct Effects on Population-Level Tuning." *BioRxiv*,
1090 530352.

1091 Goddard, Erin, Thomas A. Carlson, Nadene Dermody, and Alexandra Woolgar. 2016.
1092 "Representational Dynamics of Object Recognition: Feedforward and Feedback
1093 Information Flows." *Neuroimage* 128: 385–397.

1094 Gregoriou, Georgia G., Stephen J. Gotts, Huihui Zhou, and Robert Desimone. 2009.
1095 "High-Frequency, Long-Range Coupling between Prefrontal and Visual Cortex during
1096 Attention." *Science* 324 (5931): 1207–1210.

1097 Hanks, Timothy D., and Christopher Summerfield. 2017. "Perceptual Decision Making in
1098 Rodents, Monkeys, and Humans." *Neuron* 93 (1): 15–31.

1099 Hebart, Martin N., Brett B. Bankson, Assaf Harel, Chris I. Baker, and Radoslaw M.
1100 Cichy. 2018. "The representational dynamics of task and object processing in humans."
1101 *Elife* 7: e32816.

1102 Henson, Richard N., Elias Mouchlianitis, William J. Matthews, and Sid Kouider. 2008.
1103 "Electrophysiological Correlates of Masked Face Priming." *Neuroimage* 40 (2): 884–
1104 895.

1105 Huang, Wanyi, Xia Wu, Liping Hu, Lei Wang, Yulong Ding, and Zhe Qu. 2017.
1106 "Revisiting the Earliest Electrophysiological Correlate of Familiar Face Recognition."
1107 *International Journal of Psychophysiology* 120: 42–53.

1108 Hupé, J. M., A. C. James, B. R. Payne, S. G. Lomber, P. Girard, and J. Bullier. 1998.
1109 "Cortical Feedback Improves Discrimination between Figure and Background by V1, V2
1110 and V3 Neurons." *Nature* 394 (6695): 784–787.

1111 Johnson, Matthew R., Karen J. Mitchell, Carol L. Raye, Mark D'Esposito, and Marcia K.
1112 Johnson. 2007. "A Brief Thought Can Modulate Activity in Extrastriate Visual Areas:
1113 Top-down Effects of Refreshing Just-Seen Visual Stimuli." *Neuroimage* 37 (1): 290–
1114 299.

1115 Karimi-Rouzbahani, Hamid, Alexandra Woolgar, and Anina N. Rich. 2020a. "Neural
1116 signatures of vigilance decrements predict behavioral errors before they occur." *bioRxiv*.

1117 Karimi-Rouzbahani, Hamid, Ehsan Vahab, Reza Ebrahimpour, and Mohammad Bagher
1118 Menhaj. 2019. "Spatiotemporal Analysis of Category and Target-Related Information
1119 Processing in the Brain during Object Detection." *Behavioral Brain Research* 362: 224–
1120 239.

1121 Karimi-Rouzbahani, Hamid, Mozghan Shahmohammadi, Ehsan Vahab, Saeed
1122 Setayeshi, and Thomas Carlson. 2020b. "Temporal codes provide additional category-
1123 related information in object category decoding: a systematic comparison of informative
1124 EEG features." *bioRxiv*.

1125 Karimi-Rouzbahani, Hamid, Nasour Bagheri, and Reza Ebrahimpour. 2017a. "Average
1126 activity, but not variability, is the dominant factor in the representation of object
1127 categories in the brain." *Neuroscience* 346: 14–28.

1128 Karimi-Rouzbahani, Hamid, Nasour Bagheri, and Reza Ebrahimpour. 2017b. "Hard-
1129 Wired Feed-Forward Visual Mechanisms of the Brain Compensate for Affine Variations
1130 in Object Recognition." *Neuroscience* 349: 48–63.

1131 Karimi-Rouzbahani, Hamid, Nasour Bagheri, and Reza Ebrahimpour. 2017c. "Invariant
1132 Object Recognition Is a Personalized Selection of Invariant Features in Humans, Not
1133 Simply Explained by Hierarchical Feed-Forward Vision Models." *Scientific Reports* 7 (1):
1134 1–24.

1135 Karimi-Rouzbahani, Hamid. 2018. "Three-Stage Processing of Category and Variation
1136 Information by Entangled Interactive Mechanisms of Peri-Occipital and Peri-Frontal
1137 Cortices." *Scientific Reports* 8 (1): 1–22.

1138 Kaufmann, Jürgen M., Stefan R. Schweinberger, and A. Mike Burton. 2009. "N250 ERP
1139 Correlates of the Acquisition of Face Representations across Different Images." *Journal
1140 of Cognitive Neuroscience* 21 (4): 625–641.

1141 Kay, Kendrick N., and Jason D. Yeatman. 2017. "Bottom-up and Top-down
1142 Computations in Word-and Face-Selective Cortex." *Elife* 6: e22341.

1143 Kelly, Simon P., and Redmond G. O'Connell. 2013. "Internal and External Influences on
1144 the Rate of Sensory Evidence Accumulation in the Human Brain." *Journal of*
1145 *Neuroscience* 33 (50): 19434–19441.

1146 Kietzmann, Tim C., Courtney J. Spoerer, Lynn Sörensen, Radoslaw M. Cichy, Olaf
1147 Hauk, and Nikolaus Kriegeskorte. 2019. "Recurrence Required to Capture the Dynamic
1148 Computations of the Human Ventral Visual Stream." *ArXiv Preprint ArXiv:1903.05946*.

1149 Kovács, Gyula. 2020 "Getting to Know Someone: Familiarity, Person Recognition, and
1150 Identification in the Human Brain." *Journal of Cognitive Neuroscience* 1-19.

1151 Kramer, Robin SS, Andrew W. Young, and A. Mike Burton. 2018. "Understanding Face
1152 Familiarity." *Cognition* 172: 46–58.

1153 Lamme, Victor AF, and Pieter R. Roelfsema. 2000. "The Distinct Modes of Vision
1154 Offered by Feedforward and Recurrent Processing." *Trends in Neurosciences* 23 (11):
1155 571–579.

1156 Lamme, Victor AF, Karl Zipser, and Henk Spekreijse. 2002. "Masking Interrupts Figure-
1157 Ground Signals in V1." *Journal of Cognitive Neuroscience* 14 (7): 1044–1053.

1158 Landi, Sofia M., and Winrich A. Freiwald. 2017. "Two Areas for Familiar Face
1159 Recognition in the Primate Brain." *Science* 357 (6351): 591–595.

1160 Lee, Tai Sing, and David Mumford. 2003. "Hierarchical Bayesian Inference in the Visual
1161 Cortex." *JOSA A* 20 (7): 1434–1448.

1162 Leibenluft, Ellen, M. Ida Gobbini, Tara Harrison, and James V. Haxby. 2004. "Mothers'
1163 Neural Activation in Response to Pictures of Their Children and Other Children."
1164 *Biological Psychiatry* 56 (4): 225–232.

1165 Leppänen, Jukka M., and Charles A. Nelson. 2009. "Tuning the developing brain to
1166 social signals of emotions." *Nature Reviews Neuroscience* 10 (1): 37-47.

1167 Mechelli, Andrea, Cathy J. Price, Karl J. Friston, and Alumit Ishai. 2004. "Where
1168 Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery."
1169 *Cerebral Cortex* 14 (11): 1256–1265.

1170 Mognon, Andrea, Jorge Jovicich, Lorenzo Bruzzone, and Marco Buiatti. 2011.
1171 "ADJUST: An Automatic EEG Artifact Detector Based on the Joint Use of Spatial and
1172 Temporal Features." *Psychophysiology* 48 (2): 229–240.

1173 Mohsenzadeh, Yalda, Sheng Qin, Radoslaw M. Cichy, and Dimitrios Pantazis. 2018.
1174 "Ultra-Rapid Serial Visual Presentation Reveals Dynamics of Feedforward and
1175 Feedback Processes in the Ventral Visual Pathway." *Elife* 7: e36329.

1176 Norman, Kenneth A., Sean M. Polyn, Greg J. Detre, and James V. Haxby. 2006.
1177 "Beyond mind-reading: multi-voxel pattern analysis of fMRI data." *Trends in cognitive
1178 sciences* 10 (9): 424-430.

1179 Pelli, Denis G. 1997. "The VideoToolbox Software for Visual Psychophysics:
1180 Transforming Numbers into Movies." *Spatial Vision* 10 (4): 437–442.

1181 Philiastides, Marios G., and Paul Sajda. 2006. "Temporal Characterization of the Neural
1182 Correlates of Perceptual Decision Making in the Human Brain." *Cerebral Cortex* 16 (4):
1183 509–518.

1184 Philiastides, Marios G., Roger Ratcliff, and Paul Sajda. 2006. "Neural Representation of
1185 Task Difficulty and Decision Making during Perceptual Categorization: A Timing
1186 Diagram." *Journal of Neuroscience* 26 (35): 8965–8975.

1187 Praß, Maren, Cathleen Grimsen, Martina König, and Manfred Fahle. 2013. "Ultra Rapid
1188 Object Categorization: Effects of Level, Animacy and Context." *PLoS One* 8 (6).

1189 Pratte, Michael S., Sam Ling, Jascha D. Swisher, and Frank Tong. 2013. "How
1190 Attention Extracts Objects from Noise." *Journal of Neurophysiology* 110 (6): 1346–1356.

1191 Ramon, Meike, and Maria Ida Gobbini. 2018. "Familiarity Matters: A Review on
1192 Prioritized Processing of Personally Familiar Faces." *Visual Cognition* 26 (3): 179–195.

1193 Ramon, Meike, Luca Vizioli, Joan Liu-Shuang, and Bruno Rossion. 2015. "Neural
1194 Microgenesis of Personally Familiar Face Recognition." *Proceedings of the National
1195 Academy of Sciences* 112 (35): E4835–E4844.

1196 Ress, David, Benjamin T. Backus, and David J. Heeger. 2000. "Activity in Primary
1197 Visual Cortex Predicts Performance in a Visual Detection Task." *Nature Neuroscience* 3
1198 (9): 940–945.

1199 Schweinberger, Stefan R., Esther C. Pickering, Ines Jentzsch, A. Mike Burton, and
1200 Jürgen M. Kaufmann. 2002. "Event-Related Brain Potential Evidence for a Response of
1201 Inferior Temporal Cortex to Familiar Face Repetitions." *Cognitive Brain Research* 14 (3):
1202 398–409.

1203 Shadlen, Michael N., and William T. Newsome. 2001. "Neural Basis of a Perceptual
1204 Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey." *Journal of
1205 Neurophysiology* 86 (4): 1916–1936.

1206 Spacek, Martin A., Gregory Born, Davide Crombie, Steffen A. Katzner, and Laura
1207 Busse. 2019. "Robust Effects of Cortical Feedback on Thalamic Firing Mode during
1208 Naturalistic Stimulation." *BioRxiv*, 776237.

1209 Sugiura, Motoaki, Carlos Makoto Miyauchi, Yuka Kotozaki, Yoritaka Akimoto, Takayuki
1210 Nozawa, Yukihito Yomogida, Sugiko Hanawa, Yuki Yamamoto, Atsushi Sakuma, and
1211 Seishu Nakagawa. 2015. "Neural Mechanism for Mirrored Self-Face Recognition."
1212 *Cerebral Cortex* 25 (9): 2806–2814.

1213 Summerfield, Christopher, and Tobias Egner. 2009. "Expectation (and attention) in
1214 visual cognition." *Trends in cognitive sciences* 13 (9): 403-409.

1215 Summerfield, Christopher, Tobias Egner, Matthew Greene, Etienne Koechlin, Jennifer
1216 Mangels, and Joy Hirsch. 2006. "Predictive codes for forthcoming perception in the
1217 frontal cortex." *Science* 314 (5803): 1311-1314.

1218 Supèr, Hans, Henk Spekreijse, and Victor AF Lamme. 2001. "Two Distinct Modes of
1219 Sensory Processing Observed in Monkey Primary Visual Cortex (V1)." *Nature
1220 Neuroscience* 4 (3): 304–310.

1221 Taylor, Margot J., Marie Arsalidou, Sarah J. Bayless, Drew Morris, Jennifer W. Evans,
1222 and Emmanuel J. Barbeau. 2009. "Neural Correlates of Personally Familiar Faces:
1223 Parents, Partner and Own Faces." *Human Brain Mapping* 30 (7): 2008–2020.

1224 Visconti, M. di Oleggio Castello, and M. I. Gobbini. 2015. "Familiar Face Detection in
1225 180 Ms." *PLoS One* 10 (8): e0136548–e0136548.

1226 Wiese, Holger, Simone C. Tüttenberg, Brandon T. Ingram, Chelsea YX Chan, Zehra
1227 Gurbuz, A. Mike Burton, and Andrew W. Young. 2009. "A robust neural index of high
1228 face familiarity." *Psychological science* 30 (2): 261-272.

1229 Woolgar, Alexandra, Adam Hampshire, Russell Thompson, and John Duncan. 2011.
1230 "Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex."
1231 *Journal of Neuroscience* 31 (41): 14592–14599.

1232 Woolgar, Alexandra, Soheil Afshar, Mark A. Williams, and Anina N. Rich. 2015.
1233 "Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for
1234 Flexible Cognitive Control." *Journal of Cognitive Neuroscience* 27 (10): 1895–1911.

1235 Young, Andrew W., and A. Mike Burton. 2018. "Are We Face Experts?" *Trends in
1236 Cognitive Sciences* 22 (2): 100–110.

1237

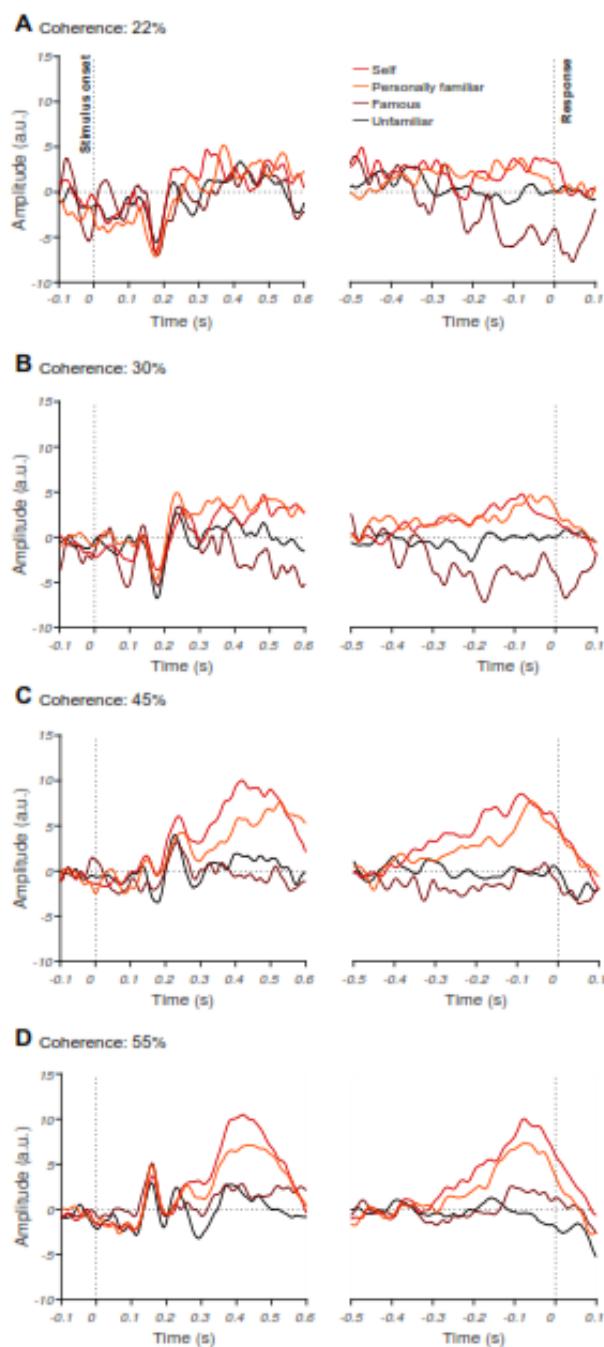
1238

1239

1240

1241

1242 Supplementary Materials

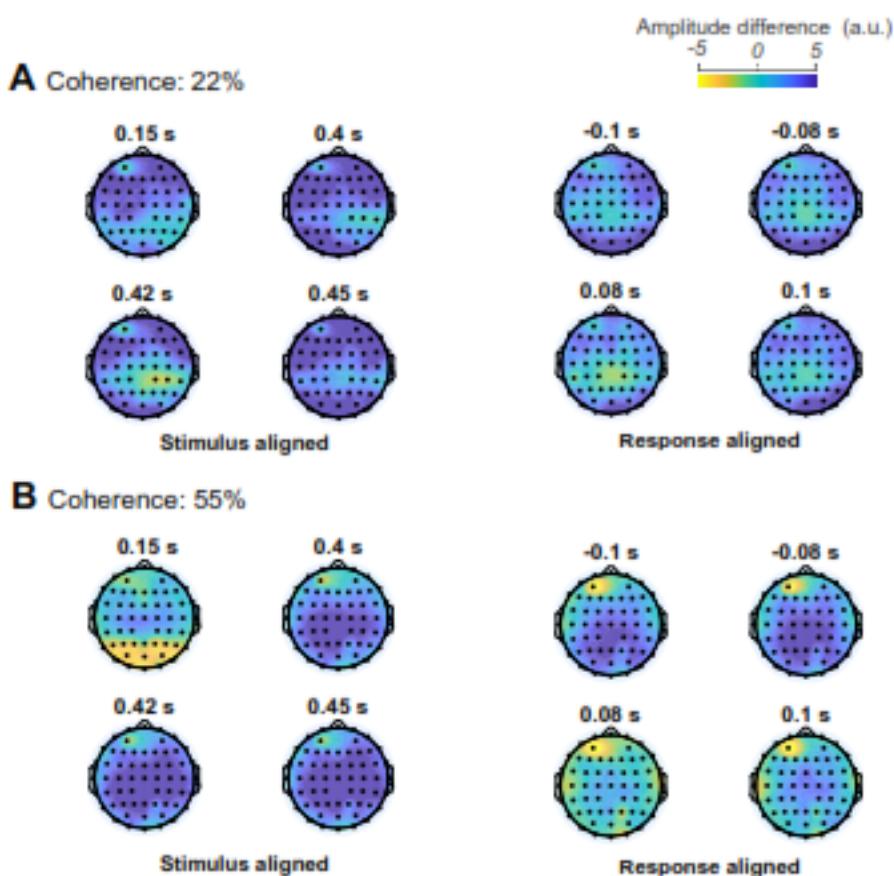


1243

1244 **Supplementary Figure 1. The effect of familiarity and sensory evidence on event-related potentials**
1245 **(ERPs).** Averaged ERPs for 22% (A), 30% (B), 45% (C) and 55% (D) phase coherence levels and four
1246 face categories across all participants for an electrode at a centroparietal site (CP2). Note that the left
1247 panels show stimulus-aligned ERPs while the right panel shows response-aligned ERPs. The differences
1248 between levels of familiarity were more pronounced at later stages of stimulus processing around 400 ms
1249 post-stimulus onset and <100 ms before the response was given by the participant in the stimulus- and
1250 response-aligned analyses, respectively.

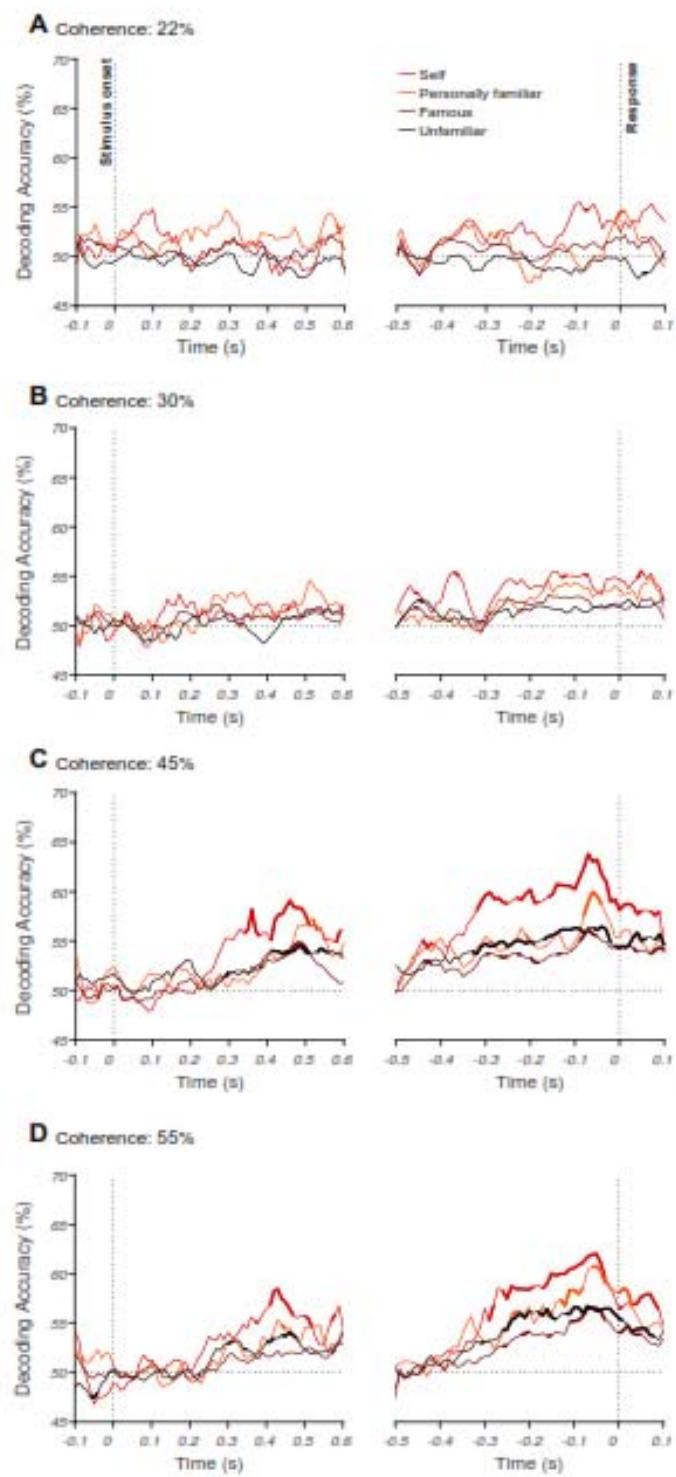
1251

1252



1253

1254 **Supplementary Figure 2. Familiarity information on the whole-brain event-related potentials**
1255 **(ERPs).** The topographic maps show the difference in ERPs between unfamiliar and the average of the
1256 three familiar face categories (i.e. unfamiliar-average of unfamiliar categories) at specific time points
1257 averaged across participants. The time points were chosen based on the results from Figure 2, when the
1258 ERPs were significantly ($p < 0.05$) higher in the 55% vs. 22% coherence levels. Note that the left panels
1259 show stimulus-aligned ERPs while the right panel shows response-aligned ERPs.



1260

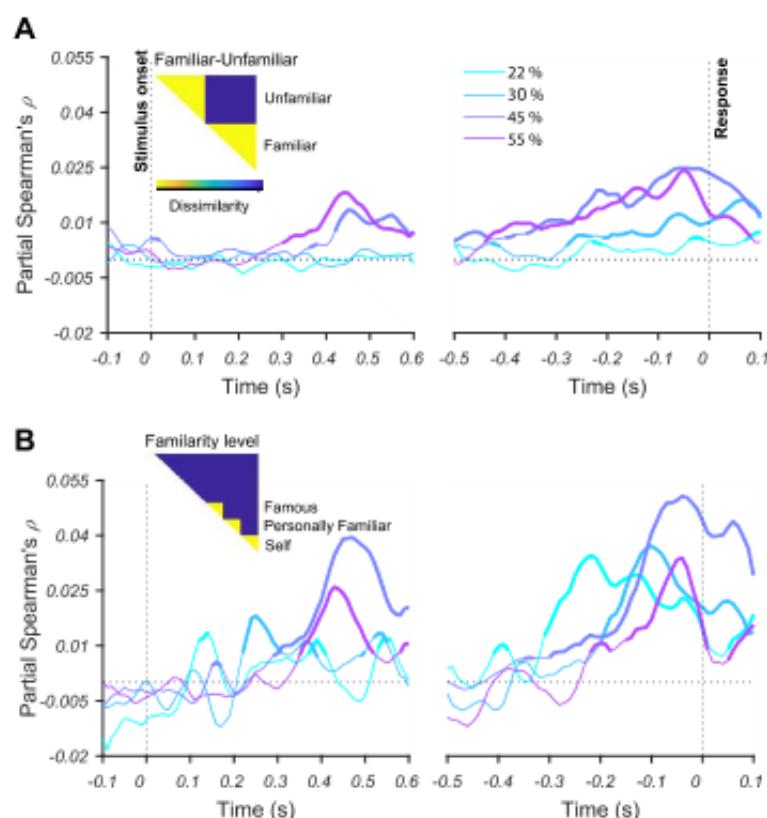
1261 **Supplementary Figure 3. Decoding of face familiarity from EEG signals.** Time course of decoding
1262 accuracy for familiar versus unfamiliar faces from EEG signals for four different phase coherence levels
1263 (22% (A), 30% (B), 45% (C), and 55% (D)). The chance accuracy is 50%. Thickened lines indicate the
1264 time points when the accuracy was significantly above chance level (sign rank test, FDR corrected across
1265 time, $p < 0.05$). The left panels show the results for stimulus-aligned analysis while the right panels show
1266 the results for response-aligned analysis (averaged over 18 participants).

1267 Supplementary Text: Low-level image statistics do not explain the
1268 separation of familiar from unfamiliar faces

1269 Although, we did equalize the frequency content, pixel intensities and contrast of
1270 the images of our dataset (see *methods*), but we checked whether there are other low-
1271 level differences by creating a model representational dissimilarity matrix (RDM) for
1272 each of the categories under different phrase coherences. Briefly, *neural* RDMs are
1273 constructed by calculating the correlations (or dissimilarities) of the *brain* response to
1274 different face stimuli to give an abstract representation of information encoding in the
1275 brain. We also construct a *low-level feature* RDM, for which we calculate the
1276 correlations between images corresponding to each brain response. *Model* RDMs
1277 *predicted* representations in the brain (see *Methods*). The model RDMs were created
1278 for discriminating (1) familiar from unfamiliar (Supplementary Figure 4A) and also (2) the
1279 familiarity levels from one another (Supplementary Figure 4B). We then computed
1280 partial Spearman's correlations between one of the models and neural RDMs for every
1281 time point and participant, while partialling out (Supplementary Figure 4)/not partialling
1282 out (Supplementary Figure 5) low-level feature model RDM.

1283 This analysis revealed the emergence of familiarity representation (familiar vs.
1284 unfamiliar faces) at around 270 ms post-stimulus for the highest coherence level (55%,
1285 Supplementary Figure 4A). The onset of significant representation is slightly later for
1286 lower coherence levels (e.g., 45%, Supplementary Figure 4A), which may suggest the
1287 need for additional processing time required to evaluate the sensory evidence.
1288 Interestingly, while the dynamics of familiarity level representations also showed gradual
1289 accumulation of information (Supplementary Figure 4B), especially for the 45% and
1290 55% coherence, the correlation values are generally higher for the model of familiarity
1291 level compared to familiar-unfamiliar (c.f. Supplementary Figure 4A). This suggests that
1292 there might be well-established neural mechanisms in the brain that discriminate levels
1293 of familiarity so strongly that is not suppressed-dominated by the task (i.e. here familiar-
1294 unfamiliar) or the response of the participants. This could also be supported by the
1295 observation that, as opposed to the familiar-unfamiliar representations, for which the
1296 55% coherence provided the most information (at least in the stimulus-aligned analysis),

1297 the familiarity level representations provided their highest information in lower
1298 coherence levels such as 45% (in both stimulus- and response-aligned analyses) and
1299 30% or even 22% in the response-aligned analysis. Note that participants' task and
1300 response could have also potentially contributed to the analysis of face familiarity model
1301 as those factors matched the familiar-unfamiliar model used in Supplementary Figure
1302 4A.

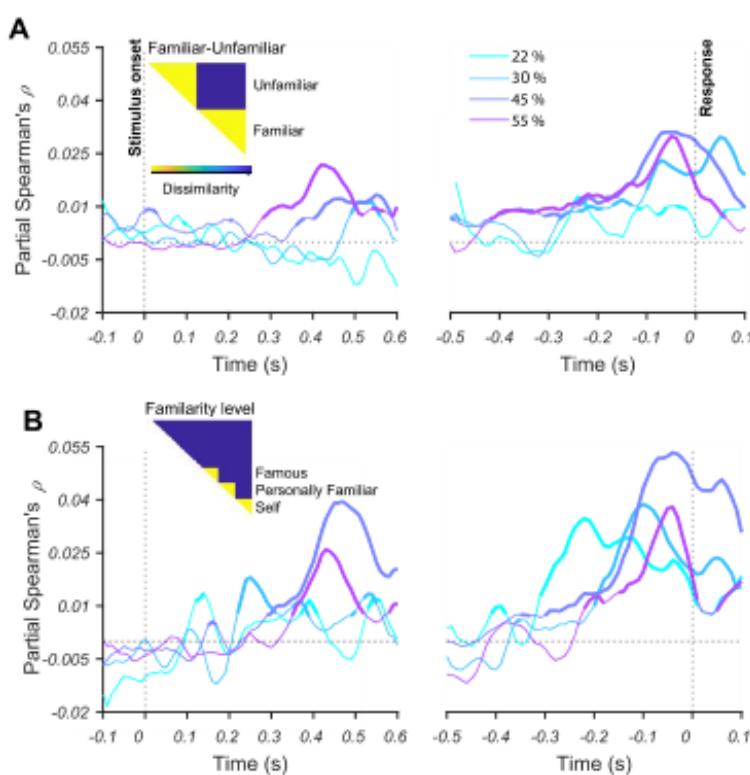


1303
1304 **Supplementary Figure 4. Representations of face familiarity and categories revealed by RSA.** Time
1305 course of Spearman's correlations between neural RDMs and model RDM (shown as insets) for (A) face
1306 familiarity; and (B) face familiarity levels, famous, self and personally familiar faces, after partialling out
1307 contributions from low-level features (see Methods). Each colored trace shows the correlations over time
1308 for one phase coherence level. Thickened lines indicate time points where the correlation is significant
1309 (sign permutation test, FDR-corrected significance level at $p < 0.05$), and black horizontal dotted lines
1310 indicate 0 correlation. The left panels show the results for stimulus-aligned analysis while the right panels
1311 represent the results for response-aligned analysis.

1312
1313 Apart from a small difference in absolute decoding rates, the dynamics of neural
1314 representations were similar when not partialling out the low-level feature model RDM

1315 (Supplementary Figure 5), presenting the ramping up of information, with earlier and
1316 most mounting trends for highest coherence levels (i.e. 45% and 55%). The similar
1317 patterns of neural information decoding between the correlation patterns with and
1318 without the low-level feature model suggest that low-level image statistics may only play
1319 a minor role in driving the observed decoding analyses. Nonetheless, we partialled out
1320 the low-level feature model in all the following RSA-based analyses to avoid their
1321 potential contribution to the results.

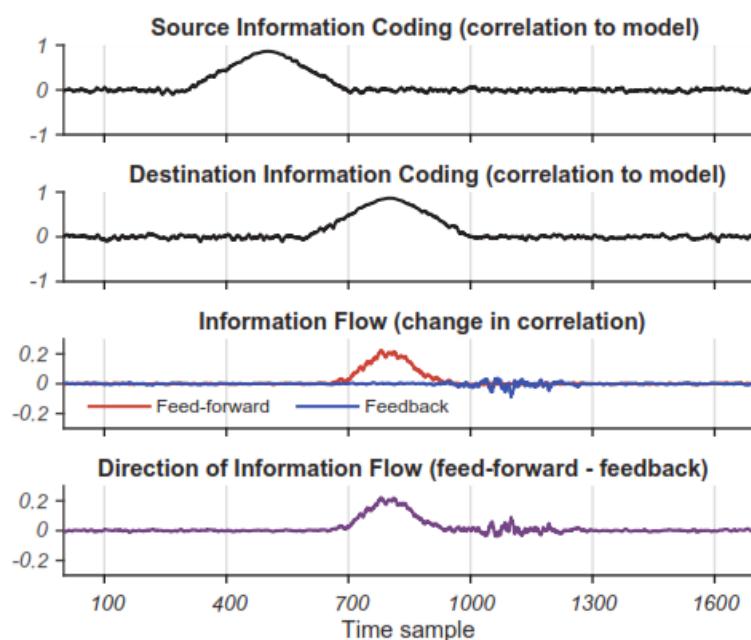
1322



1323

1324 **Supplementary Figure 5. Representations of face familiarity and categories revealed by RSA.** Time
1325 course of Spearman's correlations between neural RDMs and model RDM (shown as insets) for (A) face
1326 familiarity; and (B) face familiarity levels, famous, self and personally familiar faces, before partialling out
1327 contributions from low-level features (see Methods). Each colored trace shows the correlations over time
1328 for one phase coherence level. Thickened lines indicate time points where the correlation is significant
1329 (sign permutation test, FDR-corrected significance level at $p < 0.05$), and black horizontal dotted lines
1330 indicate 0 correlation. The left panels show the results for stimulus-aligned analysis while the right panels
1331 represent the results for response-aligned analysis. Note that the correlation values are higher compared
1332 to the results after partialling out contributions from low-level features (see Supplementary Figure 4).

1333



1334

1335 **Supplementary Figure 6. Simulation of informational connectivity and its measurement using the**
1336 **proposed connectivity analysis.** Feed-forward and feedback refer to information flow from source to
1337 destination and vice versa, respectively. For this stimulation, we initially generated three static 16×16
1338 RDMs, one for source, one for destination and one for model, all similar to the familiar-unfamiliar RDMs
1339 we presented throughout the manuscript (e.g. Figure 4A). Next, by adding varying levels of uniform noise,
1340 in the range between 0 and 1, to the artificially generated source and destination static RDMs, we
1341 generated temporally changing dynamics of information coding (i.e. measured as how much they
1342 correlated with the desired model RDM) in those RDMs across time samples. The peak of information
1343 coding in the destination RDM was designed to appear **300** samples after the coding in the source RDM,
1344 so that it simulates flow of information in the feed-forward direction. Finally, we applied our connectivity
1345 analysis to these data to check if it could detect the information flow from source to destination area. The
1346 two top panels show the correlation between temporally varying simulated source and destination RDMs
1347 and the model RDM. Third panel from top shows the amount of feed-forward and feedback information
1348 and the bottom panel shows their difference as measured by our informational connectivity analysis.
1349 Results show a clear feed-forward (from source to destination) information flow and almost zero
1350 information flow in the feedback (from destination to source) direction, which peaks almost simultaneously
1351 with the information peak in the destination area. This result suggests that our informational connectivity
1352 detects the simulated connectivity in the correct direction and temporal dynamics.

1353

1354

1355

1356