1 **Ensemble encoding of facial expressions is**

2 **greater for static faces than dynamic faces**

3

4

5 Mia Hanson & David Pitcher

6

7 Department of Psychology, University of York, Heslington, York, YO10 5DD, U.K.
8

9

10 <u>**Abstract**</u>

11 Understanding the behaviour and emotions of other people is predicated on accurately
12 decoding nonverbal facial cues. In the current study, we investigated whether the
13 ensemble encoding of these facial cues is processed differently between moving and
14 static faces. Across two experiments, participants performed a delayed-match-to-
15 sample expression recognition task using dynamic and static stimuli. Ensemble size
16 was manipulated by increasing the number of faces in the target array. In Experiment 1,
17 we presented one, two, and four target faces. In Experiment 2, we presented one, four,
18 and eight target faces. Results demonstrated that while recognition accuracy for
19 dynamic and static faces was comparable when one target face was presented, task
20 accuracy diverged as the ensemble size increased. Specifically, increasing the number
21 of faces in the target array failed to improve task performance for the dynamic
22 expression stimuli. In addition, the results of Experiment 2 demonstrated that when
23 eight faces were presented in the dynamic condition, performance was impaired for all
24 expressions (disgust, fear, and happy) as well as in the non-emotional air-puff condition.
25 This suggests motion in general disrupted the participants ability to extract an accurate
26 summary statistic as the ensemble size was increased.

27

28

29 <u>**Keywords**</u>

31

32

33

34

Introduction

Faces convey a broad range of socially relevant visual cues that are used in everyday interaction and communication. These cues can inform decisions about a person's identity, gender, age, gaze direction, intent, and emotional state. The ability to efficiently and accurately interpret these cues is a cornerstone of successful social navigation, empathy, and interpersonal interaction. Cognitive and neurobiological models of face processing (Allison et al., 2000; Bruce & Young, 1986; Haxby et al., 2000; Pitcher & Ungerleider, 2021) propose that these cues can be broadly segregated into cues that rely on invariant aspects (e.g. facial structure and shape) and changeable aspects (e.g. eye, mouth and head movements). The invariant (or static) cues contribute to identity discrimination (Grill-Spector et al., 2004; Megreya & Burton, 2006; Sliwinska et al., 2022) while the changeable aspects contribute to facial expression discrimination (Krumhuber et al., 2023; Pitcher et al., 2020). This functional dissociation between static and dynamic facial cues is supported by evidence from different methods. These include behavioural experiments (Calvo et al., 2016; Dobs et al., 2018; Lander et al., 1999), neuropsychology (Pitcher, 2025; Prabhakar et al., 2025; Sliwinska, Bearpark, et al., 2020), neuroimaging (Fox et al., 2009; Kucuk et al., 2024; Puce et al., 1998; Schultz & Pilz, 2009) and developmental studies (Kosakowski et al., 2022).

The majority of behavioural research into facial expression discrimination has used static images as experimental stimuli (Dawel et al., 2022). Static stimuli, typically photographs of peak emotional expressions, offer a controlled and precise method for examining the recognition of affective information from discrete, snapshot-like cues

(Ekman & Friesen, 1976; Pitcher, 2014; Pitcher et al., 2008; Pitcher et al., 2023;

Sliwinska, Elson, et al., 2020). However, in the real world, faces are rarely static. They

are constantly in motion as facial expressions unfold over time. This temporal dimension

is a rich source of socially relevant information (Fiorentini et al., 2012; Krumhuber &

Scherer, 2011; Krumhuber et al., 2023). Dynamic stimuli provide this critical temporal

information and reveal the unique trajectory and biomechanics of how an expression is

formed and how it dissipates. Consequently, some empirical evidence has reported an

advantage for expression recognition using dynamic stimuli (Arsalidou et al., 2011;

Fiorentini & Viviani, 2011; Jiang et al., 2014; Kamachi et al., 2001; Richoz et al., 2024).

In the current study we investigated whether dynamic stimuli would also produce a

behavioural advantage when recognising multiple facial expressions simultaneously, a

process known as ensemble encoding.

Ensemble encoding is the cognitive process in which observers are able to extract an

average summary statistic from multiple exemplars of a stimulus category rather than

processing each exemplar individually (Whitney & Yamanashi Leib, 2018). This

produces a redundancy gain, where the coactivation of multiple signals increases the

overall signal strength and results in faster, more accurate task performance. Prior

research has demonstrated that expression recognition benefits from ensemble

encoding when using both static (Haberman & Whitney, 2007, 2009; Hao et al., 2025;

Sama et al., 2025) and dynamic (Elias et al., 2017) stimuli. However, it is unclear

whether the redundancy gain is the same for both dynamic and static expressions. The

additional visual information conveyed in dynamic stimuli may interfere with the ability to

form an efficient ensemble summary. Instead of a redundancy gain, processing multiple dynamic faces might result in a redundancy cost, impairing performance as the number of faces increases.

To test this hypothesis, we designed two experiments utilizing a delayed match-to-sample task facial expression discrimination task. Participants were asked to match facial expressions presented as either dynamic videos or as static images of the peak expression (van der Gaag et al., 2007). We systematically manipulated ensemble size by varying the number of faces presented in the target array (Figure 1a). In Experiment 1, we compared accuracy for one, two, and four target faces. In Experiment 2, we compared accuracy for one, four, and eight target faces. Facial expression stimuli included a range of basic emotions (happy, fear, disgust) as well as neutral faces (the actor staring expressionless into the camera) and a non-emotional motion control ("air-puff") to determine if these any differential effects were limited to facial expressions or generalised to any facial movement. We hypothesised that ensemble encoding may be more challenging when using dynamic compared to static stimuli as the visual information needs to be summated over a longer time scale.

Materials and Methods

Participants

50 participants (35 females, 13 males and 2 non-binary) took part in Experiment 1. 40 participants (27 females, 9 males and 4 non-binary) took part in Experiment 2. All

1   subjects had normal or corrected-to-normal vision and gave informed consent as

2   directed by the ethics committee in the Psychology department at the University of York.

3

4   **Stimuli**

5   Dynamic stimuli consisted of videos of actors depicting different emotional facial

6   expressions (van der Gaag et al., 2007). These stimuli have been used in previous

7   studies of emotion discrimination (Nikel et al., 2022; Pitcher et al., 2020). The videos

8   show one of twelve actors (6 males, 6 females, all Caucasian) depicting different facial

9   expressions (disgust, fear, happy) together with the actors blowing out cheeks (motion

10   with no expression) and neutral (in which the actors stare into the camera without

11   expression). This resulted in a total of 60 video stimuli. Static stimuli were created using

12   a single frame of peak emotional expression from each video. All stimuli were then

13   organised by emotion and numbered from 1-12 based on the actor (e.g. happy 1,

14   disgust 5, etc.). Stimuli were presented in full colour on a grey background. Dynamic

15   stimuli were presented without audio. Dynamic and static stimuli were shown for 2

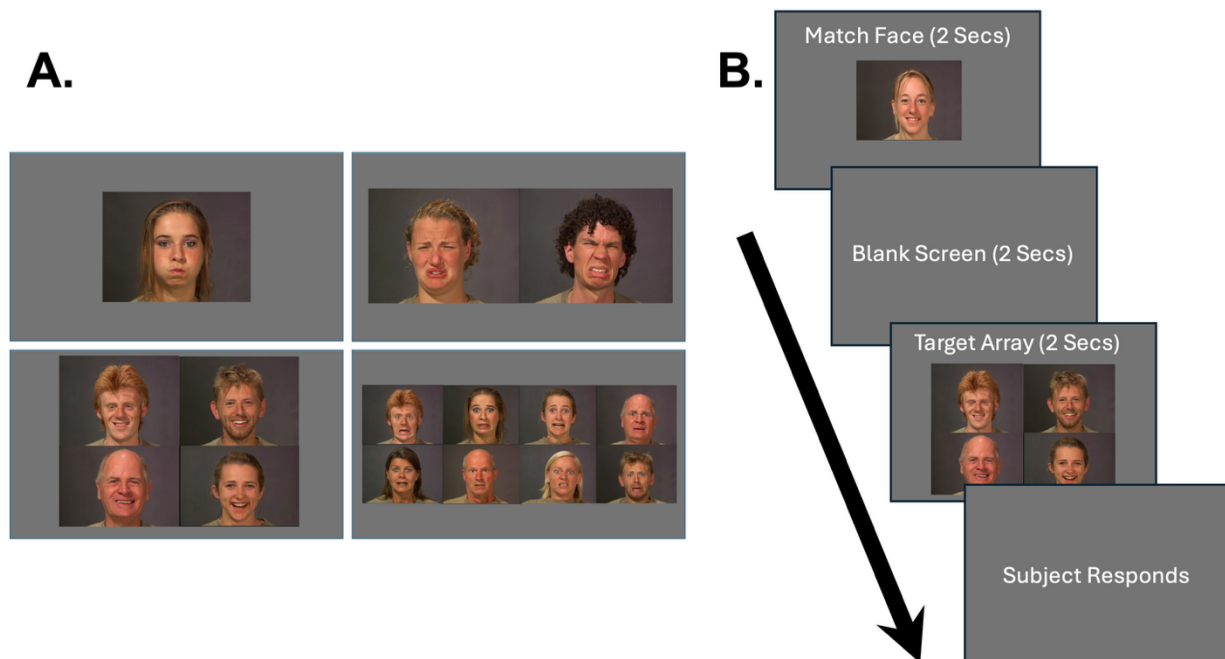16   seconds at a frame rate of 143 frames per second.

17

Figure 1a. Shows the format of the Target Arrays for the static condition. The dynamic arrays presented videos in the same configuration. In Experiment 1 participants saw arrays with 1, 2 or 4 faces. In Experiment 2 participants saw arrays with 1, 4 or 8 faces. The stimuli in the arrays for the 2, 4 and 8 faces all expressed the same emotion. Figure 1b. The trial procedure for Experiment 1 and 2. Participants performed a delayed match to sample expression discrimination task. Participants had to indicate whether the expression of the match face was the same as the faces (or face) in the target array.

Procedure

Participants performed a delayed-match-to-sample facial expression discrimination task using dynamic and static stimuli (Figure 1a). The experiment was conducted using Psychopy Version 3 (Peirce et al., 2019), which controlled randomisation and presentation of the stimuli, as well as measuring response time and accuracy. Each trial began with a single match face (dynamic or static) presented centrally for 2 seconds. This was followed a blank grey screen for two seconds. The target face array (dynamic or static) was then presented. In Experiment 1 the target array consisted of 1 face

1   (Figure 1b), 2 faces (Figure 1c) or 4 faces (Figure 1d). In Experiment 2 the target array

2   consisted of 1 face (Figure 1b), 4 faces (Figure 1d) or 8 faces (Figure 1e). All faces in

3   the target arrays presented the same expressions in all conditions (e.g. all 8 videos in

4   the dynamic condition would all depict the same expression). The participants were

5   asked to match if the expression of the match face was the same as the face, or faces,

6   depicted in the target arrays. This was done using the arrow keys to indicate whether

7   faces in the target array showed the same emotion (right arrow key) or a different

8   emotion (left arrow key) to the match face. In Experiment 1 participants completed six

9   blocks: dynamic stimuli with 1, 2 and 4 face arrays and static stimuli with 1, 2 and 4 face

10 arrays. Each block consisted of fifty trials. In Experiment 2 participants completed six

11 blocks: dynamic stimuli with 1, 4 and 8 face arrays and static stimuli with 1, 4 and 8 face

12 arrays. Each block consisted of fifty trials. Block order was randomised across

13 participants.

14

15 **Results**

16 Experiment 1

17 The results of Experiment 1 (Figure 2a) demonstrated that static facial expressions

18 were more accurately matched than dynamic facial expressions but only when two or

19 four target faces were presented. Accuracy data were entered into a two (motion:

20 dynamic and static) by three (number of faces: one face, two faces and four faces) by

21 five (emotion: happy, neutral, air-puff, fear and disgust) repeated measures analysis of

22 variance (ANOVA). We found significant main effects of motion ($F_{(1,49)} = 15.44$, $p <$

1   0.001; partial $\eta^2 = 0.24$), number of faces ($F_{(2, 98)} = 5.48$, $p = 0.006$; partial $\eta^2 = 0.10$)

2   and emotion ($F_{(4,196)} = 37.65$, $p = 0.001$; partial $\eta^2 = 0.44$). Motion and number of

3   faces ($F_{(2,98)} = 6.01$, $p = 0.003$; partial $\eta^2 = 0.11$) and motion and emotion ($F_{(4,196)} =$

4   20.50, $p < 0.001$; partial $\eta^2 = 0.30$) both combined in significant two-way interactions.

5   Number of faces and emotion failed to show a significant two-way interaction ($F_{(8,392)}$

6      $= 0.94$, $p = 0.48$; partial $\eta^2 = 0.02$). There was no significant three-way interaction

7   between motion, number of faces and emotion ($F_{(8,392)} = 0.85$, $p = 0.56$; partial η^2

8      $= 0.02$).

9

10   To understand what factors were driving the significant two-way interaction between

11   motion and number of faces we analysed the data in a separate ANOVA (Figure 2a).

12   Results showed significant main effects of motion ($F_{(1,49)} = 19.25$, $p < 0.001$; partial

13      $\eta^2 = 0.28$), number of faces ($F_{(2,98)} = 5.00$, $p = 0.009$; partial $\eta^2 = 0.09$) and a

14   significant two-way interaction ($F_{(2,98)} = 4.57$, $p = 0.013$; partial $\eta^2 = 0.85$). Pairwise

15   comparisons showed that static faces were more accurately recognized than dynamic

16   faces when 2 faces ($p < 0.001$) and 4 faces ($p < 0.001$) were presented. There was no

17   significant difference between dynamic and static faces when 1 face was presented ($p =$

18   0.82). Further comparisons showed significant differences in the static condition

19   between the 1 and 2 target face condition ($p = 0.027$) and between the 1 and 4 target

20   face condition ($p = 0.008$) but not between the 2 and 4 target face condition ($p = 1$).

21   There were no significant differences in the dynamic face condition between any of the

22   target face conditions ($p > 0.26$).

23

1   Next, to understand what factors were driving the significant two-way interaction

2   between motion and emotion we analysed the data in a separate two-way ANOVA

3   (Figure 2b). Results showed significant main effects of motion ($F_{(1,49)}$ = 15.44, p =

4   0.001; partial η^2 = 0.24), emotion ($F_{(4,196)}$ = 37.65, $p < 0.001$; partial η^2 = 0.44) and a

5   significant two-way interaction ($F_{(4,196)}$ = 20.50, $p < 0.001$; partial η^2 = 0.30). Pairwise

6   comparisons showed significant differences between dynamic and static faces when the

7   neutral expressions were presented ($p < 0.001$) and for air-puff expressions (p = 0.044),

8   but not when happy (p = 0.5), fear (p = 0.09) or disgust (p = 0.8) expressions were
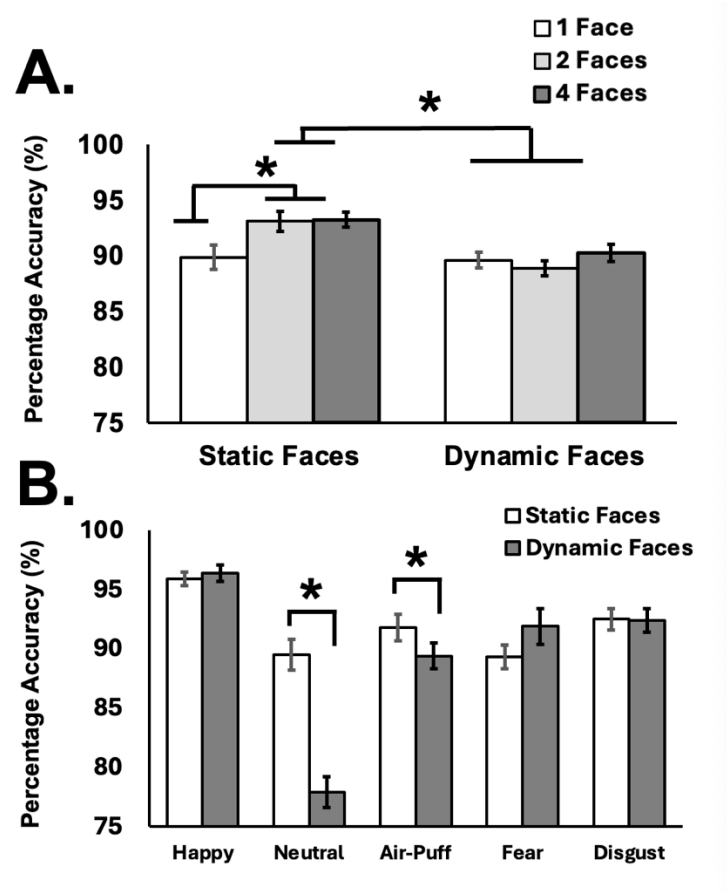
9   presented.

10

11   **Figure 2a. The results of the significant two-way ANOVA between motion and**
12   **number of faces in the target array. Pairwise comparisons showed that static**

1 **faces were more accurately recognized than dynamic faces when 2 faces ($p <$**
2 **0.001) and 4 faces ($p < 0.001$) were presented. Figure 2b. The results of the**
3 **significant two-way ANOVA between motion and emotion. Pairwise comparisons**
4 **showed significant differences between dynamic and static faces when the**
5 **neutral ($p < 0.001$) and air-puff faces ($p = 0.044$). were presented**

6

7 Reaction time (RT) data were entered into a two (motion: dynamic and static) by three

8 (number of faces: one face, two faces and four faces) by five (emotion: happy, neutral,

9 air-puff, fear and disgust) repeated measures analysis of variance (ANOVA). We found

10 significant main effects of motion ($F_{(1,49)} = 280.00$, $p < 0.001$; partial $\eta^2 = 0.85$),

11 number of faces ($F_{(2,98)} = 8.04$, $p < 0.001$; partial $\eta^2 = 0.14$) and emotion ($F_{(4,196)} =$

12 18.30, $p < 0.001$; partial $\eta^2 = 0.27$). Motion and emotion combined in significant two-way

13 interaction ($F_{(4,196)} = 3.96$, $p = 0.004$; partial $\eta^2 = 0.08$). However, there were no two-

14 way interactions between motion and number of faces ($F_{(2,98)} = 0.15$, $p = 0.86$; partial

15 $\eta^2 = 0.003$) or between emotion and number of faces ($F_{(8,392)} = 0.6$, $p = 0.8$; partial

16 $\eta^2 = 0.012$). There was no significant three-way interaction between motion, number of

17 faces and emotion ($F_{(8,392)} = 0.58$, $p = 0.79$.; partial $\eta^2 = 0.012$). Pairwise

18 comparisons of the significant main effects showed that static faces were more quickly

19 recognised than dynamic faces ($p < 0.001$), that 4 faces in the target array were more

20 quickly recognised than 1 face ($p = 0.001$) and happy expressions were more quickly

21 recognised than all other expressions ($p < 0.001$). No other tests reached significance

22 ($p > 0.075$).

23

24 <u>Experiment 2</u>

1   Accuracy data (Figure 3) were entered into a two (motion: dynamic and static) by three

2   (number of faces: one face, four faces and eight faces) by five (emotion: happy, neutral,

3   air-puff, fear and disgust) repeated measures analysis of variance (ANOVA). We found

4   significant main effects of motion ($F_{(1,39)} = 43.37$, $p < 0.001$; partial $\eta^2 = 0.53$), number

5   of faces ($F_{(2,78)} = 7.89$, $p < 0.001$; partial $\eta^2 = 0.17$) and emotion ($F_{(4,156)} = 18.41$, p

6   < 0.001; partial $\eta^2 = 0.32$). Motion and number of faces ($F_{(2,78)} = 6.14$, $p = 0.003$;

7   partial $\eta^2 = 0.14$), motion and emotion ($F_{(4,156)} = 6.02$, $p < 0.001$; partial $\eta^2 = 0.13$)

8   and number of faces and emotion ($F_{(8,312)} = 2.06$, $p = 0.039$; partial $\eta^2 = 0.05$).

9   Finally, motion, number of faces and emotion combined in a significant three-way

10   interaction ($F_{(8,312)} = 4.51$, $p < 0.001$; partial $\eta^2 = 0.10$).

11

12   To understand what factors were driving the significant three-way interaction we

13   analysed the data in separate two-way ANOVAs. Firstly, we ran an ANOVA between

14   motion and number of faces to compare the results of experiment 2 (in which 1, 4 and 8

15   target faces were presented) with experiment 1 (in which 1, 2 and 4 target faces were
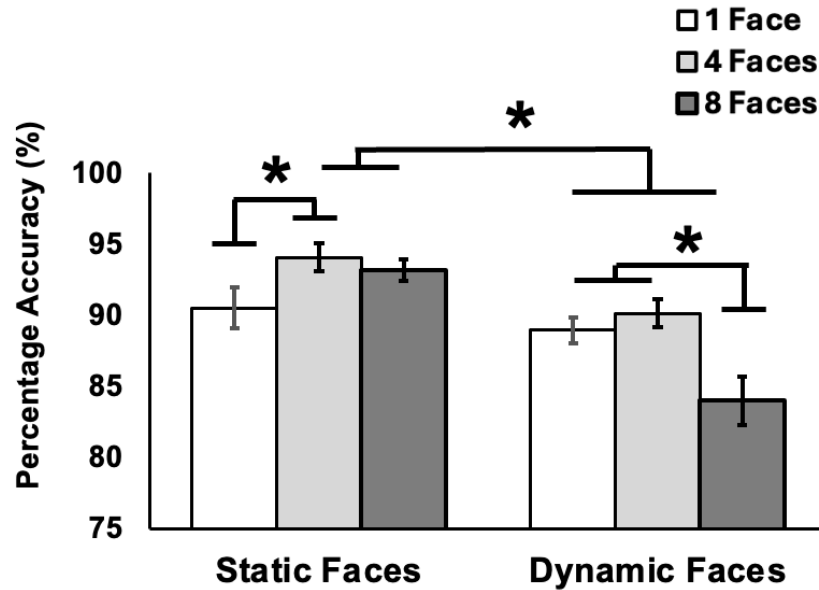
16   presented).

17

1

2 **Figure 3. The results of the significant two-way ANOVA between motion and**
3 **number of faces in the target array. Pairwise comparisons showed that static**
4 **faces were more accurately recognized than dynamic faces when 4 faces (p <**
5 **0.001) and 8 faces (p < 0.001) were presented. For dynamic faces there was a no**
6 **significant difference between when 1 and 4 faces (p = 0.026) and when 1 and 8 (p**
7 **= 0.003) faces were presented.**

8

9 Results (Figure 3) showed significant main effects of motion (F (1,39) = 43.64 , $p <$

10 0.001; partial η^2 = 0.53), number of faces (F (2,78) = 7.74, $p < 0.001$; partial η^2 = 0.17)

11 and a significant two-way interaction (F (2,78) = 7.42, $p = 0.001$; partial η^2 = 0.16).

12 Pairwise comparisons showed that static faces were more accurately recognized than

13 dynamic faces when 4 faces ($p < 0.001$) and 8 faces ($p < 0.001$) were presented. There

14 was no significant difference between dynamic and static faces when 1 face was

15 presented ($p = 0.29$). Further comparisons showed significant differences in the static

16 condition between the 1 target face and 4 target faces condition ($p = 0.02$) but not

17 between the 1 target face and 8 target faces condition ($p = 0.11$) or between the 4 and 8

1   target face condition ($p = 0.98$). In the dynamic condition there was no significant

2   difference between the 1 and 4 target faces ($p = 0.86$) but there were between when 1

3   and 4 faces ($p = 0.03$) and when 1 and 8 ($p = 0.003$) faces were presented.

4

5   Next, to examine whether there were significant differences between the discrimination

6   of dynamic and static facial expression across the target face conditions we performed

7   separate ANOVAs comparing dynamic and static stimuli with expressions for each of the

8   target array conditions (Figure 4).

9

10   1 Target Face: Accuracy data when 1 face was presented were analysed in a two-way

11   ANOVA (Figure 4a). Results did not show a significant main effect of motion ($F_{(1, 39)} =$

12   1.87, $p = 0.18$; partial $\eta^2 = 0.46$). There was a significant main effect of expression ($F_{(4,156)} = 10.28$, $p < 0.001$; partial $\eta^2 = 0.21$) and a significant two-way interaction ($F_{(4,156)} = 5.80$, $p < 0.001$; partial $\eta^2 = 0.13$). Pairwise comparisons showed significant

13    $(4,156) = 10.28$, $p < 0.001$; partial $\eta^2 = 0.21$) and a significant two-way interaction (F

14    $(4,156) = 5.80$, $p < 0.001$; partial $\eta^2 = 0.13$). Pairwise comparisons showed significant

15   differences between dynamic and static faces when the neutral expressions were

16   presented ($p < 0.001$) but not when happy ($p = 0.81$). air-puff ($p = 0.68$), fear ($p = 0.11$)

17   or disgust ($p = 0.47$) expressions were presented.

18

19   4 Target Faces: Accuracy data when 4 faces were presented were analysed in a two-

20   way ANOVA (Figure 4b). Results showed significant main effects of motion ($F_{(1, 39)} =$

21   20.99, $p < 0.001$; partial $\eta^2 = 0.35$) and expression ($F_{(4,156)} = 6.35$, $p < 0.001$; partial

22    $\eta^2 = 0.14$) and a significant two-way interaction ($F_{(4,156)} = 2.62$, $p = 0.037$; partial η^2

1   = 0.06). Pairwise comparisons showed significant differences between dynamic and

2   static faces when the neutral expressions were presented ($p < 0.001$) but not when

3   happy ($p = 0.07$), air-puff ($p = 0.12$), fear ($p = 0.88$) or disgust ($p = 0.31$) expressions

4   were presented.

5

6   <u>8 Target Faces</u>: Accuracy data when 8 faces were presented were analysed in a two-

7   way ANOVA (Figure 4c). Results showed significant main effects of motion ($F_{(1, 39)} =$

8   26.56, $p < 0.001$; partial $\eta^2 = 0.41$) and expression ($F_{(4,156)} = 9.40$, $p < 0.001$; partial

9    $\eta^2 = 0.19$) and a significant two-way interaction ($F_{(4,156)} = 6.04$, $p < 0.001$; partial η^2

10   = 0.13). Pairwise comparisons showed significant differences between dynamic and

11   static faces when happy ($p = 0.003$), air-puff ($p < 0.001$), fear ($p = 0.016$) and disgust (p

12   = 0.03) expressions were presented but not when neutral expressions were presented
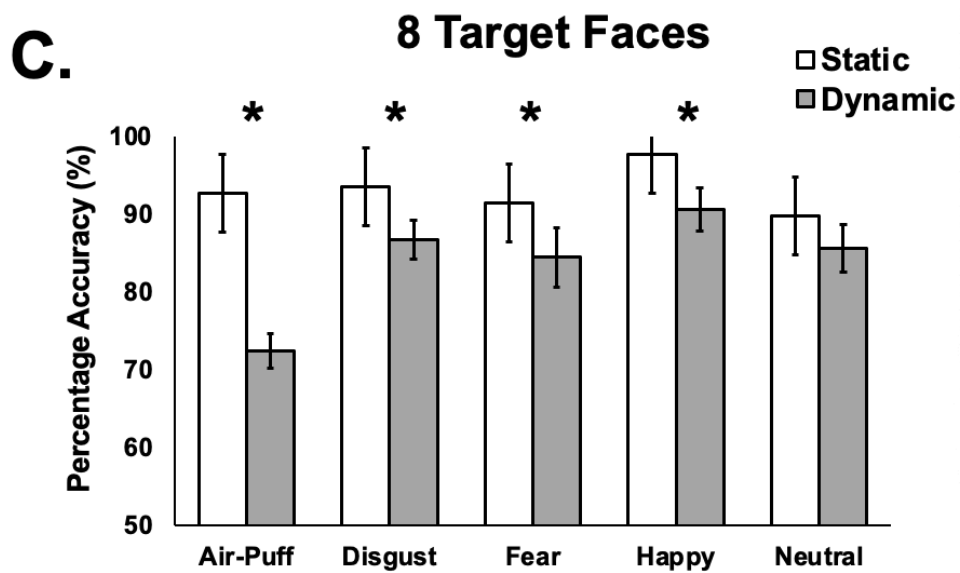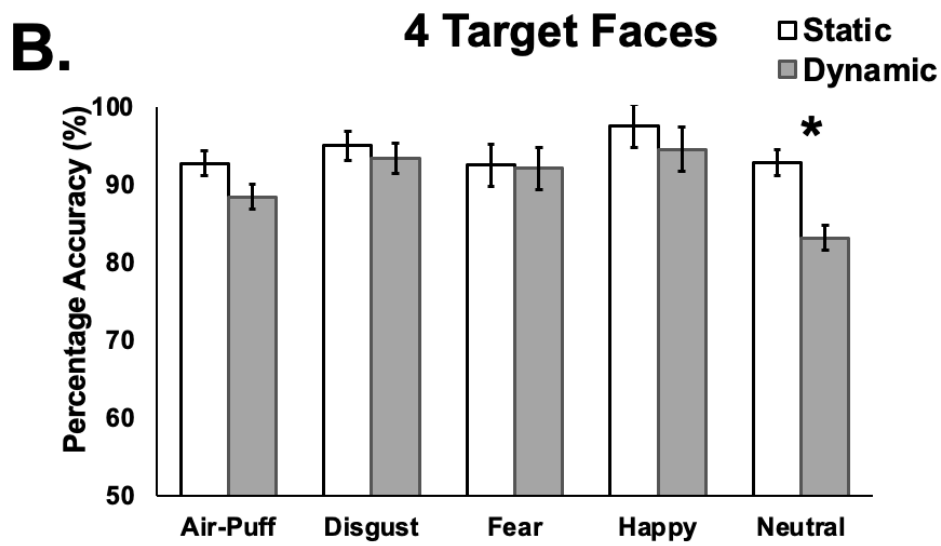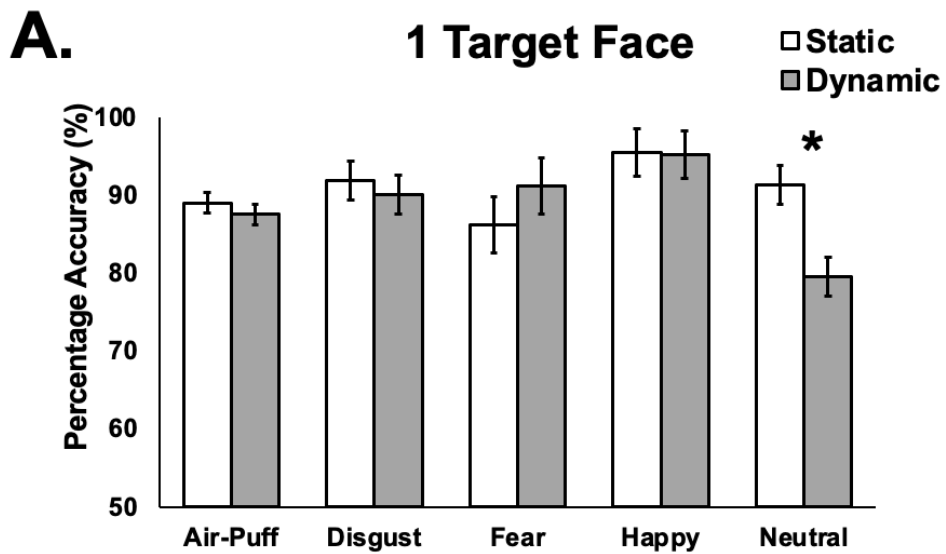
13   ($p = 0.10$).

14

A.

1 Target Face

□ Static
■ Dynamic

Percentage Accuracy (%)

*

Air-Puff   Disgust   Fear   Happy   Neutral

B.

4 Target Faces

□ Static
■ Dynamic

Percentage Accuracy (%)

*

Air-Puff   Disgust   Fear   Happy   Neutral

C.

8 Target Faces

□ Static
■ Dynamic

Percentage Accuracy (%)

*   *   *   *

Air-Puff   Disgust   Fear   Happy   Neutral

1

Figure 4a. The results of the significant ANOVA between motion conditions (dynamic or static) and expression for when 1 target face was presented. Pairwise comparisons showed a significant difference for neutral expressions only (p < 0.001). Figure 4b. The results when 4 target faces were presented. The only significant difference was in the neutral face condition (p < 0.001). Figure 4c. The results when 8 target faces were presented. There were significant differences between dynamic and static when air-puff (p < 0.001), disgust (p = 0.033), fear (p = 0.017) and happy (p = 0.003 expressions were presented but not when neutral expressions were presented (p = 0.1).

Reaction time (RT) data were entered into a two (motion: dynamic and static) by three (number of faces: one face, two faces and four faces) by five (emotion: happy, neutral, air-puff, fear and disgust) repeated measures analysis of variance (ANOVA). We found significant main effects of motion ($F_{(1,39)} = 653.18$, $p < 0.001$; partial $\eta^2 = 0.94$), number of faces ($F_{(2,78)} = 10.62$, $p < 0.001$; partial $\eta^2 = 0.21$) and emotion ($F_{(4,156)} = 9.96$, $p = 0.001$; partial $\eta^2 = 0.20$). Both motion and emotion ($F_{(4,156)} = 5.73$, $p < 0.001$; partial $\eta^2 = 0.13$) and number of faces and emotion $F_{(8,312)} = 2.31$, $p = 0.02$; partial $\eta^2 = 0.06$) combined in significant two-way interactions. There was no two-way interaction between motion and number of faces $F_{(2,78)} = 0.34$, $p = 0.72$; partial $\eta^2 = 0.009$). There was no significant three-way interaction between motion, number of faces and emotion ($F_{(8,312)} = 1.42$, $p = 0.19$.; partial $\eta^2 = 0.035$). Pairwise comparisons of the significant main effects showed that static faces were more quickly recognised than dynamic faces ($p < 0.001$). 8 faces were more quickly recognised than 4 faces ($p = 0.028$ and 1 face ($p < 0.001$). Happy expressions were more quickly recognised than all other expressions ($p < 0.004$). No other tests reached significance ($p > 0.5$).

Discussion

In the current study we investigated whether ensemble encoding of facial expressions would differ when using dynamic and static stimuli. We hypothesised that the redundancy gain reported for static facial expression recognition (Haberman & Whitney, 2007, 2009; Hao et al., 2025; Sama et al., 2025) may not be as robust when using dynamic facial expression stimuli. Our results support this hypothesis. Across two experiments, we demonstrated that while discrimination accuracy for dynamic and static faces was comparable when viewing a single target face, a performance difference emerged as the number of faces in the target array was increased. Specifically, static expressions were recognized significantly more accurately than dynamic expressions when the target array contained two or four faces (Figure 2a) and when the target array contained four or eight faces (Figure 3). Furthermore, Experiment 2 demonstrated that for dynamic stimuli, a target array of eight target faces resulted in significantly worse performance than target arrays containing four faces or one face. This impaired performance for dynamic stimuli was observed for all three expressions (disgust, fear and happy) as well as in the non-emotional motion "air-puff" condition (Figure 4c). This suggests that facial motion in general, rather than motion associated only with facial expressions, disrupts efficient ensemble encoding.

The difference in performance between dynamic and static stimuli was driven by the increase in performance accuracy for the static stimuli when the target array contained two or four faces (Figure 2a and Figure 3). This improvement is consistent with theories proposing a redundancy gain in task performance when two (or more) signals are

coactivated resulting in faster and more accurate task performance (Miller, 1982). Our

results suggest that when two or four faces were presented in the static target arrays

the correspondence of the faces expressing the same emotion increased the strength of

the signal and improved task performance. In Experiment 2 there was an increase in

performance when eight faces were presented in the target array, but the difference was

not statistically significant (Figure 3). This suggests that eight static faces may be

reaching the limit where the increase in signal strength is effective. Crucially, across

both experiments we failed to see this increased accuracy for dynamic stimuli

suggesting that dynamic faces result in a redundancy cost rather than a redundancy

gain. This cost has been reported in prior studies that required participants to match

identity and expression using static stimuli (Yankouskaya et al., 2012; Yankouskaya,

Humphreys, et al., 2014; Yankouskaya, Rotshtein, et al., 2014).

Extracting relevant social information from dynamic faces requires processing the

temporal unfolding of motion, biomechanics, and trajectory of facial components

(Krumhuber et al., 2023). The complex, time-varying information from multiple faces in

the target arrays may have been difficult to integrate into a single, stable ensemble

statistic. In contrast, processing multiple static images of peak expressions may allow

for a more rapid and efficient averaging of features (Haberman & Whitney, 2007, 2009;

Hao et al., 2025; Sama et al., 2025). In our single-face condition, this computational

cost of processing motion is not aggregated, and resources are sufficient for both tasks.

However, when required to form an ensemble using multiple faces, the cognitive

demands of integrating multiple dynamic signals outweighs the potential benefit of

1   redundancy, leading to a redundancy cost. This pattern of results is consistent with

2   cognitive and neurobiological models that propose dissociable pathways for the

3   processing of invariant (static) and changeable (dynamic) facial cues (Allison et al.,

4   2000; Bruce & Young, 1986; Haxby et al., 2000; Pitcher, 2021; Pitcher & Ungerleider,

5   2021). The behavioural dissociation we report suggests that the dynamic processing

6   pathway is more susceptible to disruption when cognition is taxed while the static

7   pathway is more resilient. This suggests that these pathways are not just functionally

8   but are also computationally distinct.

9

10   Our findings help contextualize the dynamic advantage reported in previous studies of

11   facial expression discrimination (Arsalidou et al., 2011; Fiorentini & Viviani, 2011; Jiang

12   et al., 2014; Kamachi et al., 2001; Richoz et al., 2024). This advantage, typically

13   investigated using single faces, may not scale to group or ensemble perception. Our

14   study demonstrates that any benefit of dynamic information for a single face is lost

15   when multiple faces must be processed concurrently. Interestingly a previous study

16   using dynamic faces did report an ensemble advantage for facial expression recognition

17   (Elias et al., 2017). This study did not compare dynamic and static face stimuli but

18   instead contrasted synchronous and asynchronous expressions using dynamic stimuli.

19   In addition, the faces were originally static stimuli (Tottenham et al., 2009) and the

20   dynamic aspect was created by morphing between expressions. These differences may

21   account for the discrepancy between our studies but further experiments will be needed

22   to address this discrepancy.

23

We argue that the informational richness of dynamic cues comes at a computational cost. In real-world conditions, such as trying to track the expressions of multiple people in a crowded room, our data suggest that the complexity of integrating multiple dynamic motion signals becomes a liability. Efficiently summarizing the "average mood" of a crowd may paradoxically rely more on processing static, peak expression cues than on the full, complex dynamic information. This interpretation is bolstered by the results for the neutral faces. In Experiment 1, the static advantage was driven by the neutral faces only (Figure 2b). This pattern also appeared in Experiment 2 in the one and four face target array conditions (Figure 4a and 4b). This was likely due to the nature of the stimuli. The dynamic neutral faces (an actor staring into the camera) may have been perceived as ambiguous, with subtle, non-expressive movements (e.g., blinks) being misinterpreted as suppressed intent or emotion. However, when the eight face target array was shown, this pattern reversed (Figure 4c). The difference for neutral faces was no longer significant, while a significant static face advantage emerged for all three emotional expressions (disgust, fear, and happy) and for the non-emotional motion control (air-puff). This suggests two distinct mechanisms. At low-to-moderate ensemble sizes, the dominant effect was the ambiguity of dynamic-neutral stimuli. At the highest ensemble size, this ambiguity effect was less disruptive than the general computational cost of processing complex facial motion from eight distinct sources, which impaired performance across all dynamic conditions.

Our study provides novel behavioural evidence that forming an ensemble representation of dynamic facial expressions is a computationally demanding process.

1 These findings show the crucial trade-off between the informational richness of dynamic

2 cues and their cognitive computational cost when processed as an ensemble. Real

3 world social situations will sometimes require the processing of many different facial

4 expressions. Our results demonstrate that this is a computationally challenging process

5 and future studies could manipulate real-world social interactions to better understand

6 these processes (Sliwinska et al., 2022). Furthermore, this paradigm could be a

7 powerful tool for investigating atypical social perception in clinical populations known to

8 have difficulties with facial expression recognition.

9

10

11

18

19 ## **References**

20 Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of
21   the STS region. *Trends Cogn Sci, 4*(7), 267-278. https://doi.org/10.1016/s1364-
22   6613(00)01501-1
23 Arsalidou, M., Morris, D., & Taylor, M. J. (2011). Converging evidence for the advantage
24   of dynamic facial expressions. *Brain Topogr, 24*(2), 149-163.
25   https://doi.org/10.1007/s10548-011-0171-4
26 Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3),*
27   305-327. https://www.ncbi.nlm.nih.gov/pubmed/3756376

Calvo, M. G., Avero, P., Fernandez-Martin, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion*, *16*(8), 1186-1200. https://doi.org/10.1037/emo0000192

Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2022). A systematic survey of face stimuli used in psychological research 2000-2020. *Behav Res Methods*, *54*(4), 1889-1901. https://doi.org/10.3758/s13428-021-01705-3

Dobs, K., Bulthoff, I., & Schultz, J. (2018). Use and Usefulness of Dynamic Face Stimuli for Face Perception Studies-a Review of Behavioral Findings and Methodology. *Front Psychol*, *9*, 1355. https://doi.org/10.3389/fpsyg.2018.01355

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble Perception of Dynamic Emotional Groups. *Psychol Sci*, *28*(2), 193-203. https://doi.org/10.1177/0956797616678188

Fiorentini, C., Schmidt, S., & Viviani, P. (2012). The identification of unfolding facial expressions. *Perception*, *41*(5), 532-555. https://doi.org/10.1068/p7052

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *J Vis*, *11*(3). https://doi.org/10.1167/11.3.17

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the Face Processing Network: Optimization of the Functional Localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637-1651. https://doi.org/10.1002/hbm.20630

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci*, *7*(5), 555-562. https://doi.org/10.1038/nn1224

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Curr Biol*, *17*(17), R751-753. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *J Exp Psychol Hum Percept Perform*, *35*(3), 718-734. https://doi.org/10.1037/a0013899

Hao, S., Whitney, D., & Bishop, S. J. (2025). Judging emotion in natural images of crowds. *Emotion*, *25*(1), 57-69. https://doi.org/10.1037/emo0001358

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci*, *4*(6), 223-233. https://www.ncbi.nlm.nih.gov/pubmed/10827445

Jiang, Z., Li, W., Recio, G., Liu, Y., Luo, W., Zhang, D., & Sun, D. (2014). Time pressure inhibits dynamic advantage in the classification of facial expressions of emotion. *PLoS One*, *9*(6), e100162. https://doi.org/10.1371/journal.pone.0100162

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, *30*(7), 875-887. https://doi.org/10.1068/p3131

Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R. (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Curr Biol*, *32*(2), 265-274 e265. https://doi.org/10.1016/j.cub.2021.10.064

Krumhuber, E. G., & Scherer, K. R. (2011). Affect bursts: dynamic patterns of facial expression. *Emotion*, *11*(4), 825-841. https://doi.org/10.1037/a0023856

Krumhuber, E. G., Skora, L. I., Hill, H. C. H., & Lander, K. (2023). The role of facial movements in emotion recognition. *Nature Reviews Psychology*, *2*(5), 283-296. https://doi.org/10.1038/s44159-023-00172-1

Kucuk, E., Foxwell, M., Kaiser, D., & Pitcher, D. (2024). Moving and Static Faces, Bodies, Objects, and Scenes Are Differentially Represented across the Three Visual Pathways. *J Cogn Neurosci*, *36*(12), 2639-2651. https://doi.org/10.1162/jocn_a_02139

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Mem Cognit*, *27*(6), 974-985. https://www.ncbi.nlm.nih.gov/pubmed/10586574

Megreya, A. M., & Burton, A. M. (2006). Recognising faces seen alone or with others: When two heads are worse than one. *Applied Cognitive Psychology*, *20*(7), 957-972. https://doi.org/10.1002/acp.1243

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol*, *14*(2), 247-279. https://doi.org/10.1016/0010-0285(82)90010-x

Nikel, L., Sliwinska, M. W., Kucuk, E., Ungerleider, L. G., & Pitcher, D. (2022). Measuring the response to visually presented faces in the human lateral prefrontal cortex. *Cereb Cortex Commun*, *3*(3), tgac036. https://doi.org/10.1093/texcom/tgac036

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Hochenberger, R., Sogo, H., Kastman, E., & Lindelov, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behav Res Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Pitcher, D. (2014). Facial expression recognition takes longer in the posterior superior temporal sulcus than in the occipital face area. *J Neurosci*, *34*(27), 9173-9177. https://doi.org/10.1523/JNEUROSCI.5038-13.2014

Pitcher, D. (2021). Characterizing the Third Visual Pathway for Social Perception. *Trends Cogn Sci*, *25*(7), 550-551. https://doi.org/10.1016/j.tics.2021.04.008

Pitcher, D. (2025). Neuropsychological evidence of a third visual pathway specialized for social perception. *Nat Commun*, *16*(1), 5774. https://doi.org/10.1038/s41467-025-61396-8

Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *J Neurosci*, *28*(36), 8929-8933. https://doi.org/10.1523/JNEUROSCI.1450-08.2008

Pitcher, D., Pilkington, A., Rauth, L., Baker, C., Kravitz, D. J., & Ungerleider, L. G. (2020). The Human Posterior Superior Temporal Sulcus Samples Visual Space Differently From Other Face-Selective Regions. *Cereb Cortex*, *30*(2), 778-785. https://doi.org/10.1093/cercor/bhz125

Pitcher, D., Sliwinska, M. W., & Kaiser, D. (2023). TMS disruption of the lateral prefrontal cortex increases neural activity in the default mode network when naming facial expressions. *Soc Cogn Affect Neurosci*, *18*(1). https://doi.org/10.1093/scan/nsad072

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends Cogn Sci*, *25*(2), 100-110. https://doi.org/10.1016/j.tics.2020.11.006

Prabhakar, A. T., Roy, A., Margabandhu, K., McKendrick, A. M., Carter, O., & Garrido, M. I. (2025). Double dissociation of dynamic and static face perception provides causal evidence for a third visual pathway. *Nat Commun*, *16*(1), 6513. https://doi.org/10.1038/s41467-025-61395-9

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, *18*(6), 2188-2199. <Go to ISI>://WOS:000072367700026

Richoz, A. R., Stacchi, L., Schaller, P., Lao, J., Papinutto, M., Ticcinelli, V., & Caldara, R. (2024). Recognizing facial expressions of emotion amid noise: A dynamic advantage. *J Vis*, *24*(1), 7. https://doi.org/10.1167/jov.24.1.7

Sama, M. A., Shoura, M., Nestor, A., & Cant, J. S. (2025). Attention and stimulus structure interact during ensemble encoding of facial expression. *Sci Rep*, *15*(1), 18632. https://doi.org/10.1038/s41598-025-03289-w

Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Exp Brain Res*, *194*(3), 465-475. https://doi.org/10.1007/s00221-009-1721-9

Sliwinska, M. W., Bearpark, C., Corkhill, J., McPhillips, A., & Pitcher, D. (2020). Dissociable pathways for moving and static face perception begin in early visual cortex: Evidence from an acquired prosopagnosic. *Cortex*, *130*, 327-339. https://doi.org/10.1016/j.cortex.2020.03.033

Sliwinska, M. W., Elson, R., & Pitcher, D. (2020). Dual-site TMS demonstrates causal functional connectivity between the left and right posterior temporal sulci during facial expression recognition. *Brain Stimul*, *13*(4), 1008-1013. https://doi.org/10.1016/j.brs.2020.04.011

Sliwinska, M. W., Searle, L. R., Earl, M., O'Gorman, D., Pollicina, G., Burton, A. M., & Pitcher, D. (2022). Face learning via brief real-world social interactions induces changes in face-selective brain areas and hippocampus. *Perception*, *51*(8), 521-538. https://doi.org/10.1177/03010066221098728

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res*, *168*(3), 242-249. https://doi.org/10.1016/j.psychres.2008.05.006

van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). Facial expressions: what the mirror neuron system can and cannot tell us. *Soc Neurosci*, *2*(3-4), 179-222. https://doi.org/10.1080/17470910701376878

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annu Rev Psychol*, *69*, 105-129. https://doi.org/10.1146/annurev-psych-010416-044232

Yankouskaya, A., Booth, D. A., & Humphreys, G. (2012). Interactions between facial emotion and identity in face processing: evidence based on redundancy gains. *Atten Percept Psychophys*, *74*(8), 1692-1711. https://doi.org/10.3758/s13414-012-0345-5

Yankouskaya, A., Humphreys, G. W., & Rotshtein, P. (2014). Differential interactions between identity and emotional expression in own and other-race faces: effects of familiarity revealed through redundancy gains. *J Exp Psychol Learn Mem Cogn*, *40*(4), 1025-1038. https://doi.org/10.1037/a0036259

1   Yankouskaya, A., Rotshtein, P., & Humphreys, G. W. (2014). Interactions between
2   Identity and Emotional Expression in Face Processing across the Lifespan:
3   Evidence from Redundancy Gains. *J Aging Res*, *2014*, 136073.
4   https://doi.org/10.1155/2014/136073

5