Enhanced Mirror Neuron Network Activity and Effective Connectivity during Live Interaction

Chun-Ting Hsu[1,*], Wataru Sato[1,*], Takanori Kochiyama[2], Ryusuke Nakai[3], Kohei Asano[3,4], Nobuhito Abe[3], Sakiko Yoshikawa[5]

[1]Psychological Process Research Team, Guardian Robot Project, RIKEN, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan.

[2]Brain Activity Imaging Center, ATR- Promotions, Inc., 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

[3]Kokoro Research Center, Kyoto University, 46 Yoshidashimoadachi-cho, Sakyo-ku, Kyoto, 606-8501 Japan

[4]Department of Children Education, Osaka University of Comprehensive Children Education, 6-chome-4-26 Yuzato, Higashisumiyoshi Ward, Osaka, 546-0013, Japan

[5]Institute of Philosophy and Human Values, Kyoto University of the Arts, 2-116 Uryuyama Kitashirakawa, Sakyo, Kyoto, Kyoto 606-8271, Japan.

* Corresponding Authors: chun-ting.hsu@riken.jp; wataru.sato.ya@riken.jp

Abstract

Facial expressions are indispensable for daily communication. Previous neuroimaging studies investigating facial expression processing presented pre-recorded stimuli and lacked live face-to-face interactions. Our paradigm alternated between presentations of models' real-time performance and pre-recorded videos of dynamic facial expressions to participants. Simultaneous fMRI and facial electromyography activity recordings, as well as post-scan valence and arousal ratings were acquired. Live facial expressions enhanced the subjective valence and arousal ratings as well as facial muscular responses. Live performances showed greater engagement of the right posterior superior temporal sulcus (pSTS) and right inferior frontal gyrus (IFG) more, and modulated the effective connectivity between them. A support vector machine reliably classified multivoxel activation patterns in brain regions involved in dynamic facial expression processing in mentalizing and salience networks. The results suggest that live social interaction modulates the activity and connectivity of the right mirror neuron system and enhances spontaneous mimicry, further facilitating emotional contagion.

Introduction

Facial expressions are an integral part of emotional communication and interaction[1]. They direct the receiver's attention toward, and evoke inferences about, the sender's affective state and expression in communication[2]. Facial expression observation evokes spontaneous facial mimicry[3], which, in the appropriate social context[4], facilitates emotional contagion and can serve as an implicit expression of affective empathy[5–8]. Both motor synchrony and emotional contagion contribute to social alignment and herding behaviors[9]. The extent of spontaneous facial mimicry and emotional contagion, as estimated in forms of subjective affective experiences, is associated with the liveliness of the stimuli. Pre-recorded dynamic emotional facial expressions were shown to elicit stronger subjective arousal experiences and more evident facial mimicry than static pictures of emotional facial expressions[10,11]. In comparison with viewing pre-recorded dynamic facial expressions, dynamic facial expressions observed in live interactions further potentiate subjective emotional experience in both valence and arousal dimensions, as well as the amplitude of spontaneous facial mimicry[12,13].

Previous neuroimaging studies investigating the processing of dynamic facial expressions identified several neural substrates involved, such as the superior temporal sulcus (STS)[14], including the adjacent superior and middle temporal gyri, the fusiform gyrus, the inferior frontal gyrus (IFG) including premotor areas, and limbic regions including the amygdala[15]. These brain regions belong to three neural networks that are particularly relevant to social interaction: the mirror neuron system (MNS), the mentalizing/theory of mind (ToM) network, and the salience network. The MNS, including the IFG and pSTS, is involved in social-motor information processing and supports mimicry behaviors[16,17]. The mentalizing network includes the IFG, anterior temporal lobe (aTL), temporoparietal junction (TPJ), dorsal and ventral medial prefrontal cortex (mPFC), and posterior cingulate cortex (PCC)[18]. This

network is involved in mentalizing or perspective taking about emotion of others, as part of the cognitive empathy process[6]. Furthermore, the salience network, including the limbic regions[19], together with the IFG, is involved in affective empathy[20]. Therefore, individuals are able to form representations of another's feelings by sharing these feelings through embodied simulation[6,21]. These networks and cognitive affective processes are fundamental to emotional contagion and prosocial behaviors, such as motor mimicry[7,22].

However, previous neuroimaging studies presented pre-recorded video stimuli to investigate processing of dynamic facial expressions, with no potential for participants to interact with the observed agents making the expressions. Social neuroscientific studies have emphasized the importance of ecological validity in a face-to-face, real-time interactive design in dyad in recent years[23,24]. In a previous study, we used a video camera relay system to allow participants to view either the model's live or pre-recorded dynamic facial expressions. Our facial electromyogram (EMG) and subjective rating results showed that live facial expressions evoked stronger spontaneous facial mimicry and emotional contagion[13]. However, empirical evidence is lacking regarding neurocognitive processing of live dynamic facial expressions.

In the present study, we tested the effects of face-to-face live interaction using dynamic facial expressions on subjective emotion and facial mimicry in a population of 44 young female participants and two female models. We adapted the video relay system in our previous study for the MR scanner environment (Fig. 1A). We recruited only female participants and models to control for effects of gender differences in facial mimicry responses[25]. The experiment began with a conversation with the model using the video relay system and four trials of practice viewing. Subsequently, each participant viewed pre-recorded or live videos of positive (smiling) or negative (frowning) dynamic facial expressions for 5 runs with 40 trials (10 trials per condition) per run (Fig. 1B). During viewing, we acquired

multiecho echo-planar imaging (EPI)[26] to estimate brain activities. The zygomaticus major (ZM) and corrugator supercilii (CS) EMG activity were also recorded to estimate the magnitude of spontaneous facial mimicry. This was followed by a functional localizer task for dynamic facial expression processing[27], during which participants passively viewed blocks of pre-recorded dynamic facial expressions of smiling and frowning, and dynamic mosaics of these clips. The participants then completed 16 rating trials to assess dimensions of valence and arousal[28] via verbal output with neither fMRI nor EMG data acquisition.

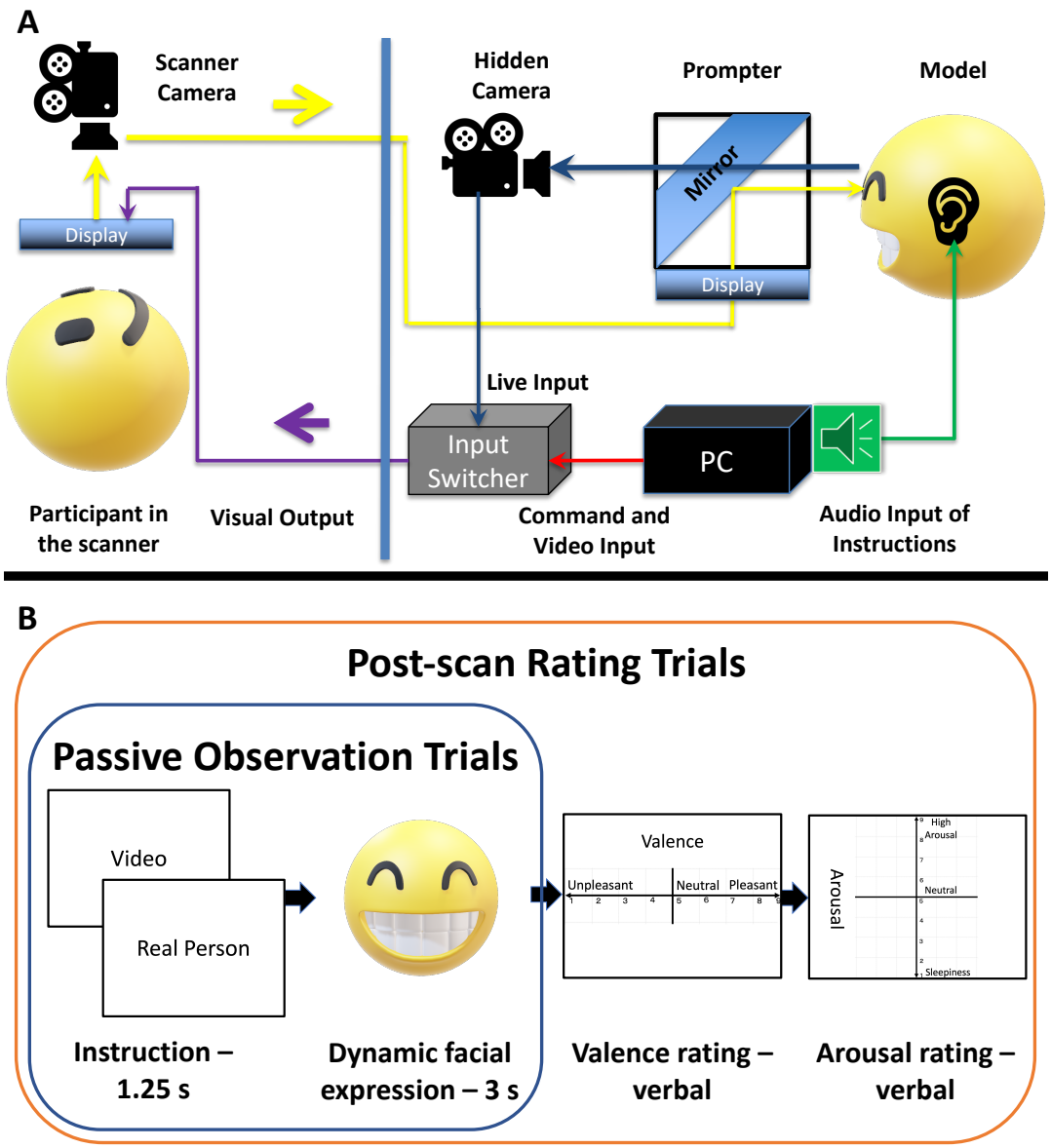

Figure 1. The experimental facilities (A) and paradigm (B). (A) The model outside the scanner faced a prompter. The prompter consisted of a horizontally positioned display and an obliquely positioned transparent mirror. The image of the model captured by a camera hidden behind the mirror was sent to the input switcher (blue route) as one of the visual output options. The PC running the paradigm sent commands (red route) through a serial port to the input switcher to determine the visual output for the participant to view (purple route). During video trials, the PC sent the instructions and pre-recorded video stimulus (red) to the input switcher, and the switcher sent only the input from the PC to the participant's display (purple). During live trials, the switcher first relayed the instructions from the PC to the participant (red), and the PC then transmitted the audio signal to the model via a pair of computer loudspeakers, instructing them to produce dynamic facial expressions (green route) while sending commands (red) to the input switcher to switch to the model's live image input (blue) and to switch back after the duration of the stimulus (3 s). The participant in the scanner could view the visual information on the display at the end of the scanner bore via the reflection by the mirror mounted on the head coil, while the participant's image was video-recorded and live-relayed by a camera at the end of the scanner bore. The participant's image was then sent to the prompter of the model (yellow route). (B) During the passive observation trials, the participant first saw the instruction for either "Video" or "Real Person" followed by 3 s of pre-recorded or live dynamic facial expression, with the fixation cross lasting for the duration of the jittered interstimulus interval. In the post-scan rating trials, the participants saw the instruction followed by the stimulus as in the passive observation trials, and were asked to tell the experimenter their ratings on the valence and arousal via the intercom from inside the scanner room.

We expected to reproduce previous findings of enhanced spontaneous mimicry and emotional contagion under live conditions in EMG and rating data[13]. Based on these findings, we further hypothesized that live interaction enhances mirroring and mentalizing, evoking the MNS and mentalizing network into a more socially engaged state. Such a change in brain activation state could be tested in the form of change in activity amplitude on fMRI in univariate analysis, in the form of changed functional organization of neuronal units in a brain substrate or across the brain presented as multivariate activation pattern change[29], or as change in connectivity within the brain network[30]. We expected to see enhanced activation in univariate analysis, and differential multivoxel activation pattern between live and pre-recorded conditions, as well as enhanced functional connectivity in the MNS and mentalizing networks. This study provides information on how the dynamics of the neural networks involved in social cognition would differ between a "third-person" and a "second-person" approach[23].

Results

Subjective ratings

Valence ratings. The rating and EMG data were analyzed with linear mixed-effects (LME) models using the lme4 package in R. The final model of the highest possible complexity for valence ratings included emotions, presentation conditions, and their interactions as fixed effects, as well as random slopes for emotions, presentation conditions and their interaction over the random factor, "subject". After model diagnostics[31], 638 data points of all 44 participants remained in the model. There were significant main effects of emotion and presentation condition. Higher ratings were given to the positive than to the negative condition, and to the live than to the video condition. There was a significant interaction between emotion and presentation condition (Table 1, Fig. 2A). Simple effects analysis using the emmeans package showed that the difference between positive-live and negative-live conditions (estimate = 3.66, SE = 0.198, df = 44.2, t = 18.501, p < 0.0001) was larger than the difference between positive-video and negative-video conditions (estimate = 2.38, SE = 0.169, df = 43.9, t = 14.061, p < 0.0001). The positive-live stimuli were rated more positively than positive-video stimuli (estimate = 0.959, SE = 0.110, df = 43.8, t = 8.703, p < 0.0001), while the negative-live stimuli were rated more negatively than negative-video stimuli (estimate = -0.317, SE = 0.085, df = 42.5, t = -3.733, p = 0.0006).

Table 1. Statistical summary of valence ratings

Fixed Effects

Effect	Beta	95% CI	SE	df	t-value	Pr(>|t|)	Rsq
Intercept	5.0589	(4.9649, 5.1528)	0.0469	43.890	107.928	< 2e-16	
Emotion	2.1344	(1.8971, 2.3718)	0.1185	44.212	18.016	< 2e-16	0.880
Presentation	0.2268	(0.1383, 0.3151)	0.0441	44.190	5.144	5.93e-06	0.375
Interaction	0.6380	(0.4856, 0.7907)	0.0761	43.394	8.385	1.27e-10	0.618

Random Effects

Group	Effect	Variance	SD	95% CI	Corr. I.	95% CI	Corr. E.	95% CI	Corr. P.	95% CI
Subject	Intercept	0.0836	0.2892	(0.2288, 0.3720)						
	Emotion	0.5914	0.7690	(0.6262, 0.9699)	-0.03	(-0.3370, 0.2915)				
	Presentation	0.0598	0.2445	(0.1795, 0.3287)	-0.05	(-0.4023, 0.3220)	0.37	(0.0141, 0.6446)		
	E:P	0.2031	0.4507	(0.2162, 0.4619)	0.56	(0.2536, 0.7771)	0.23	(-0.1030, 0.5230)	0.37	(-0.0202, 0.6714)
Residual		0.1745	0.4177	(0.3922, 0.4460)						

$n = 44$

Abbreviations: CI: confidence interval; Corr. I.: correlation with the random effect of intercept; Corr. E.: correlation with the random effect of emotion; df: Satterthwaite approximations to degrees of freedom; Rsq: effect size of semi-partial R-squared; SD: standard deviation; SE: standard error

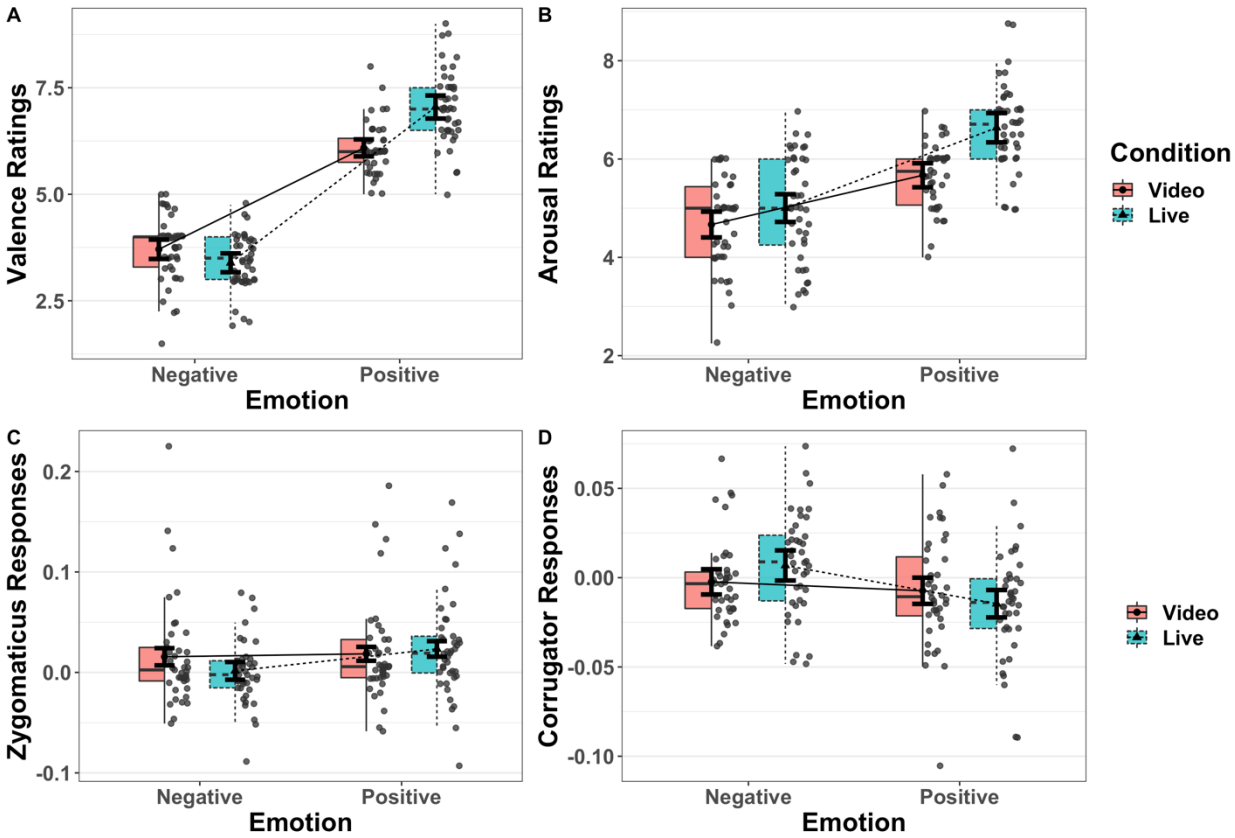

Figure 2. Interaction between presentation conditions and emotions on ratings and facial muscular reactions. (A) valence ratings (n = 44). (B) Arousal ratings (n = 42). (C) Zygomaticus responses (n = 41). (D) Corrugator responses (n = 40). In all cases, significant interactions were observed between emotion (positive, negative) and presentation condition (video, live). For each condition, the right half shows the scattered dots of

participant-wise mean values. The group mean value is shown with a filled dot (video condition) or triangle (live condition) in the middle, accompanied by error bars indicating the within-subject standard error. The box on the left half defines the median, along with the first and third quartiles of the distribution. The upper and lower whiskers extend from the hinge to the most extreme value no further than 1.5 * IQR from the hinge.

Arousal ratings. The final LME model for arousal ratings included random slopes for emotions and presentation conditions and their interaction over the random factor "subject". After model diagnostics, 602 data points of 42 participants remained in the model. There were significant main effects of emotion and presentation condition, and their interaction. Positive conditions were rated as more arousing than negative conditions, and live conditions were rated as more arousing than video conditions (Table 2, Fig. 2B). Simple effects analysis showed that the difference between positive-live and negative-live conditions (estimate = 1.63, SE = 0.225, df = 42.3, t = 7.259, p < 0.0001) was larger than the difference between positive-video and negative-video conditions (estimate = 1.00, SE = 0.191, df = 42.3, t = 5.252, p < 0.0001). The arousal rating difference between positive-live and positive-video stimuli (estimate = 0.967, SE = 0.1285, df = 41.7, t = 7.550, p < 0.0001) was larger than that between negative-live and negative-video stimuli (estimate = 0.335, SE = 0.130, df = 41.2, t = 2.587, p = 0.0133).

Table 2. Statistical summary of arousal ratings

Fixed Effects

Effect	Beta	95% CI	SE	df	t-value	Pr(>|t|)	Rsq
Intercept	5.4941	(5.3432, 5.6454)	0.0753	41.232	72.987	< 2e-16	
Emotion	0.9325	(0.6615, 1.2032)	0.1351	42.270	6.904	1.93e-08	0.530
Presentation	0.4605	(0.3219, 0.5988)	0.0689	39.972	6.682	5.24e-08	0.527
Interaction	0.3160	(0.1473, 0.4856)	0.0843	41.749	3.750	0.000538	0.252

Random Effects

Group	Effect	Variance	SD	95% CI	Corr. I.	95% CI	Corr. E.	95% CI	Corr. P.	95% CI
Subject	Intercept	0.2164	0.4652	(0.3706, 0.5974)						
	Emotion	0.7228	0.8502	(0.6866, 1.0814)	-0.35	(-0.6000, -0.0313)				

	Presentation	0.1566	0.3958	(0.2989, 0.5257)	0.39	(0.0331, 0.6603)	0.09	(-0.2557, 0.4172)		
	E:P	0.2127	0.4612	(0.3378, 0.6224)	-0.07	(-0.4287, 0.2992)	0.28	(-0.0886, 0.5805)	-0.01	(-0.3957, 0.3762)
Residual		0.2882	0.5369	(0.5030, 0.5747)						

n = 42

Abbreviations: CI: confidence interval; Corr. I.: correlation with the random effect of intercept; Corr. E.: correlation with the random effect of emotion; df: Satterthwaite approximations to degrees of freedom; Rsq: effect size of semi-partial R-squared; SD: standard deviation; SE: standard error

Facial EMG

Zygomaticus major. The final LME model for the ZM responses included the random slope of emotion and presentation condition over the random factor, "subject". During model diagnostics, 6574 data points of 41 participants remained in the model. The main effect of emotions was significant. ZM reactions were significantly stronger in the positive condition than in the negative condition. With regard to the significant interactions between emotion and presentation condition (Table 3, Fig. 2C), simple effects analysis showed that the difference in ZM reactions between the positive-live and negative-live conditions (estimate = 0.02179, SE = 0.00544, df = 77.8, t = 4.002, p = 0.0001) was greater than that between positive-video and negative-video conditions (estimate = 0.00292, SE = 0.00541, df = 75.8, t = 0.539, p = 0.5914). ZM contractions did not differ between positive-live and positive-video stimuli (estimate = 0.00484, SE = 0.00477, df = 97.1, t = 1.014, p = .3130), while negative-live stimuli evoked more ZM relaxations than negative-video stimuli (estimate = -0.01403, SE = 0.00480, df = 99.3, t = -2.924, p = 0.0043).

Table 3. Statistical summary of zygomaticus major reactions

Fixed Effects

Effect	Beta	95% CI	SE	df	t-value	Pr(>|t|)	Rsq
Intercept	1.477e-02	(0.0021, 0.0274)	6.292e-03	40.65	2.347	0.02388	
Emotion	8.734e-03	(0.0023, 0.0153)	3.224e-03	38.29	2.709	0.01003	0.161

Presentation	-3.249e-03	(-0.0087, 0.0021)	2.668e-03	38.03	-1.218	0.23069	0.038
Interaction	9.436e-03	(0.0037, 0.0152)	2.944e-03	6463	3.205	0.00136	0.002

Random Effects

Group	Effect	Variance	SD	SD 95% CI	Corr. I.	Corr. 95% CI	Corr. E.	Corr. 95% CI
Subject	Intercept	0.0015318	0.0391	(0.0315, 0.0489)				
	Emotion	0.0002456	0.0157	(0.0115, 0.0158)	0.43	(0.0219, 0.4276)		
	Presentation	0.0001127	0.0106	(0.0032, 0.0171)	-0.63	(-0.7269, -0.1523)	0.26	(-0.4003, 0.3164)
Residual		0.0142216	0.1193	(0.1172, 0.1213)				

$n = 41$

Abbreviations: CI: confidence interval; Corr. I.: correlation with the random effect of intercept; Corr. E.: correlation with the random effect of emotion; df: Satterthwaite approximations to degrees of freedom; Rsq: effect size of semi-partial R-squared; SD: standard deviation; SE: standard error

Corrugator supercilii. The final model for the CS responses included only the random slope for emotions. After model diagnostics, 6477 data points of 40 participants remained in the model. With regard to the main effect of emotion, the CS was more relaxed in the positive condition. There were significant interactions between emotion and presentation condition (Table 4, Fig. 2D). Simple effects analysis showed that the difference between positive-live and negative-live conditions (estimate = -0.02141, SE = 0.00522, df = 66.4, t = 4.104, p = 0.0001) was larger than that between positive-video and negative-video conditions (estimate = -0.00501, SE = 0.00519, df = 64.9, t = 0.9651, p = .3379). There was marginally more CS relaxation toward positive-live than positive-video stimuli (estimate = -0.00724, SE = 0.00426, df = 87.7, t = -1.701, p = 0.0925), while negative-live stimuli evoked more CS contraction than negative-video stimuli (estimate = 0.00916, SE = 0.00436, df = 96.0, t = 2.101, p = 0.0382).

Table 4. Statistical summary of corrugator supercilii reactions

Fixed Effects

Effect	Beta	95% CI	SE	df	t-value	Pr(>|t|)	Rsq
Intercept	-4.361e-03	(-0.0105, 0.0018)	3.085e-03	39.29	-1.414	0.16534	
Emotion	-9.339e-03	(-0.0158, -0.0030)	3.188e-03	37.08	-2.930	0.00578	0.188

Presentation	6.772e-04	(-0.0042, 0.0056)	2.433e-03	37.38	0.278	0.78224	0.002
Interaction	-8.200e-03	(-0.0133, -0.0031)	2.594e-03	6363	-3.161	0.00158	0.002

Random Effects

Group	Effect	Variance	SD	SD 95% CI	Corr. I.	Corr. 95% CI	Corr. E.	Corr. 95% CI
Subject	Intercept	3.110e-04	0.017634	(0.0135, 0.0197)				
	Emotion	2.683e-04	0.016379	(0.0113, 0.0229)	0.43	(0.0090, 0.7518)		
	Presentation	9.978e-05	0.009989	(0.0067, 0.0157)	0.08	(-0.4397, 0.6996)	-0.29	(-0.8464, -0.2350)
Residual		1.087e-02	0.104244	(0.1025, 0.1061)				

$n = 40$

Abbreviations: CI: confidence interval; Corr. I.: correlation with the random effect of intercept; Corr. E.: correlation with the random effect of emotion; df: Satterthwaite approximations to degrees of freedom; Rsq: effect size of semi-partial R-squared; SD: standard deviation; SE: standard error

fMRI Results

Univariate contrast. The univariate average activation pattern of all four conditions in the passive viewing task showed extensive bilateral IFG, supplementary motor area, precentral gyrus, supramarginal gyrus, inferior parietal lobule, hippocampus, and amygdala engagement (Supplementary Table S1, Fig. S1).

The factorial analysis contained two orthogonal factors, emotion (positive vs.negative) and presentation condition (video vs live), resulting in four psychological regressors of interest, i.e., the “video-positive”, “video-negative”, “live-positive”, and “live-negative” conditions. We applied whole-brain cluster-level family-wise error (FWE)-corrected $p < 0.05$, using a cluster-defining threshold of $p = 0.001$. The contrast [live > video] was associated with activities in the MNS: right posterior superior temporal sulcus (pSTS) and right IFG (Fig. 3A). The contrast [video > live] showed the right frontopolar cortex (Supplementary Fig. S2B). For the main effect of emotion, the contrast [positive > negative] showed bilateral occipital poles and right IFG, pars orbitalis (Supplementary Fig. S2C), while the contrast

231 [negative > positive] showed the right occipital pole (Supplementary Fig. S2D). The interaction
232 contrast [(video-positive > video-negative) > (live-positive > live-negative)] showed the left
233 inferior parietal lobule (IPL). There was no significant neural correlate for the interaction
234 contrast of the other direction (Table 5).

235 **Table 5. fMRI GLM results**

H	Regions	Cluster size	p (FWE-corr)*	T	B.A.	[x, y, z]
Live > Video						
R	Posterior STS	676	9.1260e-14	6.29	37	42 -64 2
R	Inferior frontal gyrus	70	0.012	4.46	6	48 2 35
Video > Live						
R	Frontopolar (middle frontal)	188	1.4125e-05	4.36	10	36 53 11
Positive > Negative						
L	Occipital pole	161	4.0784e-05	5.59	17	-24 -100 8
R	IFG, pars orbitalis	138	1.4484e-04	4.48	47	51 11 5
R	Occipital pole	56	0.0284	4.40	18	24 -91 5
Negative > Positive						
R	Occipital pole	87	0.0032	5.16	18	27 -88 -13
[(Video-Positive > Video-Negative) > (Live-Positive > Live-Negative)]						
L	Inferior parietal lobule	56	0.0225	4.15	7	-30 -61 41
Positive Correlation with Congruency-weighted Zygomaticus minus Corrugator Activities (Whole-brain)						
R	Postcentral gyrus	254	1.5267e-06	5.43	4	48 -7 35
	Precentral gyrus			4.78	4	51 -13 41
	Precentral gyrus			4.34	6	57 2 29
R	Superior occipital gyrus	127	5.9972e-04	5.27	18	21 -91 8
R	Posterior MTG	228	4.6853e-06	5.02	22	66 -40 2
	Posterior MTG			4.84	21	54 -52 5
	Posterior MTG			4.66	22	51 -34 8
L	Middle occipital gyrus	176	5.0676e-05	4.88	17	-18 -97 8
L	Postcentral gyrus	70	0.0167	4.73	4	-54 -10 47
	Postcentral gyrus			3.83	3	-45 -19 38

Positive Correlation with Congruency-weighted Zygomaticus minus Corrugator Activities (Small Volume Correction)						
R	IFG cluster (peak at Precentral gyrus)	27	0.0132	3.65	6	51 -1 38
R	Posterior MTG	463	0.0044	4.88	21	54 -52 5

* Cluster-level FWE-corrected for the whole-brain, cluster-forming threshold: uncorrected p = .001. Peak-level FWE-corrected for the small volume correction.

** Multiple peaks were listed for one cluster if the cluster spans across both hemispheres.

*** *n* = 44, except for the parametric modulation where N = 43.

Abbreviations: IFG = Inferior frontal gyrus; MTG = middle temporal gyrus; STS = superior temporal sulcus; H = hemisphere; L = left; R = right; p = p-value; T = T-value; B.A. = Brodmann area; x, y, z = MNI coordinates

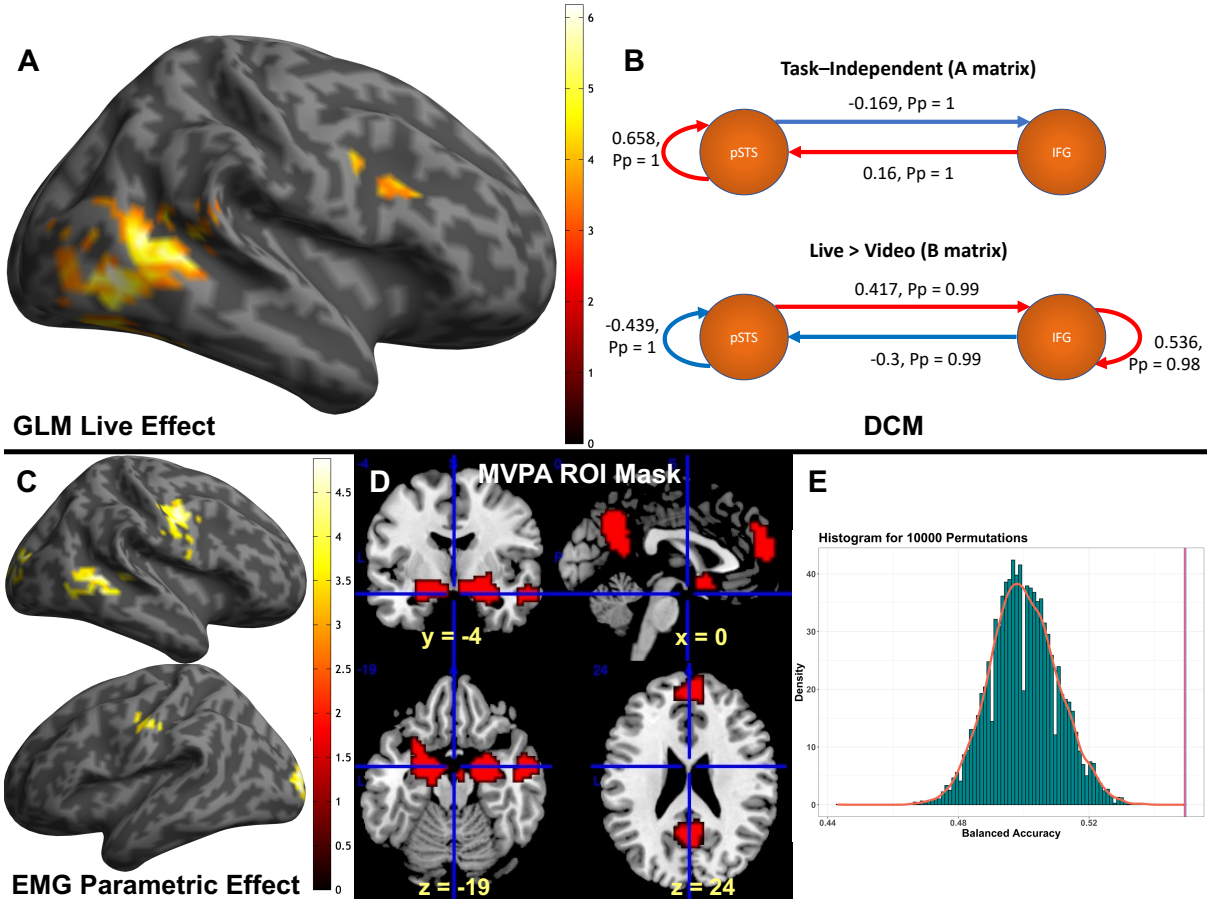

Figure 3. fMRI findings. (A) Statistical parametric maps of the main effects of presentation condition: [Live > Video], showing right IFG and pSTS (*n* = 44). (B) DCM results of right inferior frontal gyrus (IFG) and posterior superior temporal sulcus (pSTS). The circles indicate two ROIs (nodes). Red lines indicate positive effective connectivity, and blue lines indicate negative effective connectivity. The numbers indicate the strength of the connectivity (in Hz). The upper panel represents the task-independent baseline connectivity, estimated in the DCM as the A matrix. The lower panel represents the task modulatory effective connectivity, estimated in the DCM as the B matrix. In the design matrix, the parametric modulatory effect of live conditions was dummy-coded as 1,

while the effect of video conditions was dummy-coded as -1 (n = 44). (C) Statistical parametric maps of the parametric effect of congruency-weighted contrast of zygomaticus and corrugator activities (n = 43). The upper panel presents the lateral view of the right hemisphere showing activations in the right IFG, pre- and postcentral gyri, posterior superior and middle temporal gyri, and occipital pole. The lower panel presents the lateral view of the left hemisphere showed activations in the left postcentral gyrus and occipital pole. (D) The ROI mask of aTL, mPFC, PCC, amygdala, and hippocampus: the upper left, upper right, and lower left panels show corona, sagittal and transverse section views with the cross-hair at MNI [0 -4 -19], respectively. The lower right panel shows the transverse section with the cross hair at MNI [0 -4 24]. (E) Histogram (turquoise) and density plot (coral) of null distributions of the dataset based on permutations with 10,000 repetitions. The violet line shows the true classification accuracy value at 54.94%. The classifier successfully decoded multivoxel activation patterns between live vs. video conditions above the chance level in the ROIs (n = 44). The color bars indicate voxel t-values. Abbreviations: Pp = posterior probability of parameter > 0.

Parametric modulation by facial mimicry. To investigate whether activities in the right IFG and pSTS (showing the main effect of presentation condition) were associated with spontaneous facial mimicry, we applied a general linear model (GLM) analysis with one psychological regressor of interest, "dynamic facial expressions", and a parametric modulator of trial-wise congruently mimicking EMG responses. Congruently mimicking EMG responses in positive conditions included a positive ZM response (contraction) and a negative CS response (relaxation), and vice versa for the negative conditions[32,33]. Congruency was coded as either 1 (congruent) or 0 (incongruent) separately for ZM and CS responses. For each trial, the mimicking EMG response was operationalized as congruency-weighted normalized ZM response minus congruency-weighted normalized CS response. At the whole-brain level, the congruent EMG responses were correlated with activities in the bilateral pre- and postcentral gyri, occipital poles, and right posterior temporal lobe. The small volume correction[34] was significant for the clusters of right IFG (with the peak located in the precentral gyrus) and pSTS, which showed the main effect of presentation condition [live > video] (Table 5, Fig. 3C).

Dynamic causal modeling. We then used dynamic causal modeling (DCM)[35] and group-level parametric empirical Bayes (PEB) and Bayesian model reduction analysis[36] to investigate the

274 effective connectivity between two brain regions associated with the main effect of
275 presentation condition, i.e., the right pSTS and right IFG (Fig. 3A). Group DCM results (Fig.
276 3B) showed that the baseline effective connectivity consisted of negative effective connectivity
277 from right pSTS to right IFG (-0.169 Hz, posterior probability [Pp] = 1) and positive effective
278 connectivity from right IFG to right pSTS (0.16 Hz, Pp = 1). The modulatory effect of live over
279 video conditions consisted of the positive effective connectivity from right pSTS to right IFG
280 (0.417 Hz, Pp = 0.99) and negative effective connectivity from right IFG to right pSTS (-0.3
281 Hz, Pp = 0.99).

282 ***Functional localizer.*** We analyzed the functional localizer task for dynamic facial expression
283 processing using a GLM containing two psychological regressors of interest, the "dynamic
284 face" and "dynamic mosaic" conditions. The contrast of [dynamic face > dynamic mosaic] was
285 compatible with results in a previous study using more stimulus blocks[27]. Significant clusters
286 included bilateral amygdala, right posterior temporal cortex including the fusiform gyrus, and
287 the mentalizing network, such as the right anterior temporal lobe (aTL), mPFC, and PCC
288 (Supplementary Table S2, Fig. S3). A functional mask containing the significant clusters of
289 right aTL, bilateral mPFC, PCC, amygdala, and hippocampus was created for multivariate
290 pattern analysis (MVPA, Fig. 3D).

291 ***Multivariate pattern analysis.*** Among regions associated with dynamic facial expression
292 processing, regions in the mentalizing network and salience network did not show differences
293 between live and video conditions on univariate analysis. We further examined whether the
294 presentation condition modulated regions involved in dynamic facial expression processing but
295 found no amplitude difference between live vs. video conditions (i.e., right aTL, bilateral
296 mPFC, PCC, amygdala, and hippocampus, Fig. 3D) in forms of the multivariate voxel
297 activation pattern. We used a binary support vector machine (SVM) in the Pattern Recognition

298 for Neuroimaging Toolbox (PRoNTo)[37], which yielded the group-level classification accuracy
299 of 54.94%. The permutation test[38] with 10,000 repetitions yielded a p-value of 9.999×10^{-5},
300 indicating that the accuracy was significantly above the 50% chance level (Fig. 3E).

301 We then performed a whole-brain multiple kernel learning (MKL) implemented in
302 PRoNTo to further investigate anatomical regional contribution on the differentiable
303 multivariate patterns. MKL was performed using *a priori* knowledge about brain anatomy
304 defined by the Automated Anatomical Labeling (AAL)[39]. MKL estimated the linear
305 combination of each AAL regional basis kernel by a contribution vector. Brain regions could
306 therefore be ranked according to the contribution vector, which indicated the percentage of the
307 whole brain normalized weights explained by each region[40,41]. The whole-brain group-level
308 classification accuracy was 55.06%, and the contribution ranking of AAL ROIs that contributed
309 > 1% to the whole-brain weight included right PCC (33.99%), right anterior cingulate cortex
310 (ACC, 23.09%), right putamen (22.82%), left pallidum (2.39%), right pallidum (2.09%), left
311 putamen (2.02%), cerebellum vermis (1.56%), and right cerebellum (1.25%).

312 Discussion

313 Using a video camera relay system, we investigated effects of live interaction on
314 subjective experiential, muscular, and neural networks important for social cognition. Under
315 the live presentation condition, positive expressions were rated as more positive and arousing,
316 and elicited stronger congruent motor responses in ZM and CS (Table 1-4, Fig. 2). These results
317 were consistent with those of our previous behavioral study[13], and indicated that live
318 interactions enhanced spontaneous facial mimicry and emotional contagion.

319 In the fMRI data, passive observations of both prerecorded and live dynamic facial
320 expressions involved brain regions classically associated with facial expression processing,

321 including the bilateral visual cortex, the MNS including the bilateral frontoparietal network
322 and right posterior temporal lobe[17], the motor system including the bilateral motor cortex and
323 medial supplementary motor area, as well as the salience network including bilateral amygdala,
324 hippocampus and right thalamus[19] (Supplementary Table S1, Fig. S1). The whole-brain
325 activation pattern was compatible with findings of previous neuroimaging studies investigating
326 facial expression processing (see the Introduction). These observations confirmed the
327 participants' engagement in the passive observation task.

328 The main effect of presentation condition in the fMRI data showed stronger activation
329 in the right MNS (pSTS and IFG) when observing live performances (Table 5, Fig. 3A).
330 Furthermore, the activity in pSTS and IFG was parametrically correlated with measures of
331 spontaneous facial mimicry (congruent ZM and CS responses, Table 5, Fig. 3C). The live effect
332 also modulated effective connectivity within the right MNS (enhanced from pSTS to IFG, and
333 reduced from IFG to pSTS, Fig. 3B). These results were consistent with findings of previous
334 neuroimaging studies on dynamic facial expression processing reporting heightened activity in
335 the right pSTS and IFG as well as effective connectivity between the right pSTS and IFG[42,43].
336 In addition, similar results were reported in a functional near-infrared spectroscopy
337 hyperscanning study comparing the effects between dyad eye-to-eye (live eye contact) vs. eye-
338 to-picture gaze (eye contact with the eyes in the picture) between two participants. The live
339 eye-to-eye condition evoked greater responses in the left IFG pars opercularis and pre- and
340 supplementary motor cortices. Psychophysiological interaction analysis showed greater
341 functional connectivity between the left IFG and left superior temporal gyrus[44]. The authors
342 postulated that eye gaze alone does not evoke as much face recognition, emotion processing,
343 and social inferencing as dynamic facial expressions and, hence, they observed left-lateralized
344 results[44]. The parametric modulatory effect of mimicking EMG activity in the MNS in the
345 present study was also observed in a study of simultaneous fMRI-EMG acquisition

346 investigating dynamic vs. static facial expression processing; the positive correlation between
347 ZM and fMRI region of interest (ROI) activity of the happy-dynamic condition was significant
348 in the right pSTS and showed a trend in right IFG, while the negative correlation between CS
349 and ROI activity of the happy dynamic condition was significant in both the left pSTS and left
350 IFG. In the anger-dynamic condition, there was a positive trend between CS and ROI activity
351 in the right IFG[45]. Anatomical evidence also showed that the IFG and pSTS are directly
352 connected via the arcuate fasciculus and indirectly via other regions, including the IPL and
353 posterior temporal lobe[46]. The results were also consistent with the theoretical proposal that, in
354 the MNS, the direct pSTS-IFG pathway is involved in mimicry behaviors[17]. Taken together
355 with these data, our results of the heightened activity in the right pSTS and IFG, the parametric
356 correlation with mimicking amplitude involving these regions, and the modulated effective
357 connectivity between these regions in the live condition suggested that, during perception of
358 live facial expressions, the neural circuit consisting of the pSTS and IFG in the right
359 hemisphere more strongly activated mirroring or mimicry processing between individuals
360 compared with the observation of pre-recorded videos of other individuals.

361 In the present study, although the mentalizing and salience network involved in
362 dynamic facial expression processing (right aTL, bilateral mPFC, PCC, amygdala, and
363 hippocampus, Fig. 3D), as identified by the functional localizer task, did not show differences
364 on univariate analysis between different presentation conditions in the passive observation task,
365 the multivoxel activation patterns in these regions could be distinguished between presentation
366 conditions using the SVM (Fig. 3E). MKL analysis further indicated that such differentiation
367 was mainly due to the activation patterns in the right ACC and PCC of the mentalizing network.
368 The differentiable activation pattern in the mentalizing network could be explained by
369 activation of the metacognitive "we-mode"[47], the automatic taking account of the knowledge
370 and intentions of others, in the live conditions. Alternatively, the "watching eyes effect"[48]

371 suggested that knowledge of the presence of others heightened individuals' attention and self-
372 referential processing, resulting in greater expression of mimicry and engagement in social-
373 reputation management for the purpose of increasing likeability and social rapport[4]. Conty et
374 al. proposed that mentalizing computations resulting from the belief of being watched occurred
375 in the pSTS/MTG, mPFC, and OFC. Self-referential processing resulted in self-awareness and
376 pro-social behaviors involved medial cortical structures, such as the ACC and dorsal mPFC
377 (dmPFC)[48].

378 The right lateral frontopolar cortex (FPl or FPC), Brodmann area (BA) 10, was more
379 active in the video than in the live condition (Table 5, Supplementary Fig. S2B). The FPl has
380 been cognitively associated with exploration, rapid learning, relational reasoning[49], and
381 management of competing goals during multitasking[50]. Specifically in social cognition,
382 patients with lesions in BA10 showed deficits in complex ToM tasks[51]. Recently, FPl has also
383 been associated with emotion regulation[52] and management of social context-dependent
384 private-public mappings (private state vs. public actions)[53]. We found FPl relevant to social
385 context processing in the present study. The video condition is more ambiguous in terms of
386 social interaction context than the live condition. Participants felt less as though they were
387 being observed, and showed a more private state of social processing in the video condition. It
388 could then become more cognitively complex in mentalizing to simulate the mental state of a
389 counterpart in a video prerecorded at an earlier time. This also corresponded with the finding
390 of the left IPL, part of the MNS, in the interaction contrast [(video-positive > video-negative)
391 > (live-positive > live-negative)].

392 In terms of the main effects of emotion condition, we found the right IFG to be more
393 active in the positive than the negative condition (Table 5, Supplementary Fig. S2C). According
394 to a previous meta-analysis[54], the IFG is valence-general, i.e., it responds to both positive and

395 negative stimuli. In the present study, the positive stimuli, regardless of pre-recorded or live
396 performances, were generally rated as further away from the neutral level and more arousing
397 than the negative stimuli (Tables 1 and 2, Fig. 2A, B). It is plausible that the positive stimuli
398 engaged more valence-associated processing in the right IFG, which is particularly associated
399 with emotional dynamic facial expression processing, especially in the context of social
400 interaction[55]. Both directions of comparisons between positive and negative stimuli showed
401 significant occipital involvement (Supplementary Fig. S2C, D). However, the significant
402 cluster for the positive > negative contrast was higher in position (larger MNI z-axis value)
403 than the cluster in the negative > positive contrast. As the position of visual salience in the
404 positive stimuli (the moving mouth) was lower than the position of visual salience in the
405 negative stimuli (the moving eyebrows), and the inverse retinotopy in the primary visual cortex,
406 these results likely reflected the effects of visual features in dynamic facial expressions.

407 Spontaneous mimicry, emotional contagion, emotion and cognitive empathy, and
408 social conformity have been incorporated under the framework of social alignment for dyadic
409 interaction and group herding[9]. Motor synchrony, emotional contagion, and cognitive
410 synchrony reciprocally influence one another and contribute to social alignment. Under this
411 model, the dmPFC and dACC (mentalizing system), and anterior insula, served as error
412 detectors for social alignment. When misalignment was detected, the human MNS (IFG, IPL,
413 pSTS, and premotor cortex), functioning in observation-execution, served as the alignment
414 system to align the emotion and motor behavior with the other person(s). The aligned and
415 synchronized motor, emotion, and cognitive status reduced the prediction error in social
416 alignment and herding[9]. Our findings in spontaneous facial mimicry (EMG) and emotional
417 contagion (ratings) suggested that there is more social alignment effort under the second-person
418 condition (live condition). Our contrasting neuroimaging findings between live and pre-
419 recorded condition further suggested that social alignment might engage the misalignment

420 detection and alignment enhancement networks in different ways neurocognitively. Within the
421 MNS for alignment enhancement, the response to live interaction is presented as enhanced
422 univariate activation amplitude, which was correlated with the amplitude of mimicking
423 muscular activities (Table 5, Fig. 3C) and the enhanced effective connectivity within the MNS
424 (Fig. 3B). In the mentalizing network for misalignment detection, however, we did not observe
425 difference in response amplitude. Instead, the differentiable co-activation patterns of voxels
426 suggested different functional organization for misalignment detection between live vs pre-
427 recorded conditions (Fig. 3D and E). Note that while the observation-execution process of the
428 MNS was explicitly described as a consequence of alignment error detection in the
429 abovementioned framework, the mirroring activity could also provide an embodied copy[56] of
430 the observed motor, emotion, or cognitive status for gap detection in the mentalizing network.
431 Such iterative alignment processes have been proposed and computationally implemented in
432 the model of "hierarchical modular selection and identification for control" (HMOSAIC)[57]. In
433 HMOSAIC, the observer's internal social model infers the hidden mental state of the observed
434 person by predicting possible actions and their motor commands in the MNS and motor system,
435 and comparing the predicted outcome with the observed action, likely to occur in the
436 mentalizing network, leading to updated predictions about the internal model of the observed
437 person and possibly leading to the alignment of both the observer's internal model and the
438 simulated internal model regarding the observed person. While our neuroimaging data did not
439 directly address the computational nature of different networks, our findings corroborated
440 previous findings that the knowledge of the mere presence of another individual could enhance
441 emotion and autonomic synchronization[58] for social alignment. The present study has also
442 provided empirical neuroscientific evidence for the ecological validity of the second-person
443 approach for social neuroscience.

444 During social alignment, emotional contagion changed the trajectory of an agent's

445 emotional state[59]. As both motor mirroring and emotional empathy facilitated emotional
446 contagion[32,60], the alignment process would involve emotion appraisal[61] and constructive
447 processes[62], and the resulting emotional state would be empathic, vicarious, and simulated[61].
448 Indeed, while our affective ratings and EMG results showed emotion-by-presentation condition
449 interactions, the univariate activity and effective connectivity in the MNS were modulated by
450 the presentation condition. Furthermore, the MKL multivariate results suggested that the
451 mentalizing network (right ACC and PCC) and right putamen contributed 79.9% of the whole-
452 brain differentiable multivoxel pattern between presentation conditions. According to the
453 appraisal-by-content model, the mentalizing network, particularly the dmPFC, has also been
454 proposed to appraise others' mental status, intentionality, and traits[63]. Vividly demonstrated in
455 a previous fMRI study, when participants read short events involving four protagonists, the
456 multivoxel activation pattern in the dmPFC could differentiate between representations of
457 personality types[64]. The lateral prefrontal cortex (lPFC), including the IFG, generally reflected
458 an explicit appraisal of an ongoing emotional state based on context, goals, or task demands.
459 Specifically, the lPFC is involved in the valuation component of appraisal and emotion
460 regulatory strategy selection[63]. On the other hand, the putamen has been shown to be associated
461 with negative emotion suppression[65], and also anger suppression for the purpose of cooperative
462 behavior[66]. Bilateral putamen volume has also been reported to be negatively associated with
463 accurate recognition of fearful faces[67]. This paralleled the role of the mentalizing network in
464 social alignment error detection and the role of MNS in alignment enhancement.

465 This study had some limitations. To avoid confounding factors, such as models
466 making more spontaneous facial expressions and improvising facial expressions in live trials,
467 the models posed with easily reproducible facial expressions. Consequently, the live and
468 videotaped performances could not be differentiated without instruction[13], and we could more
469 confidently attribute the observed effect to participants' knowledge of being in a live interaction.

470 However, in real-life interactions, facial expressions are usually embedded in the narrative flow
471 of the conversation, and seldom used in isolation[2,68]. Posed, stereotyped facial expressions
472 could also be perceived as less natural than evoked facial expressions[69]. A more naturalistic
473 social context and interactive activities, such as conversation or games, would help the
474 participants to feel closer to real-life social interactions. Such designs could possibly evoke
475 more natural emotional facial expressions and facial mimicry in dyadic interactions. In addition,
476 to avoid effects of age and gender differences on spontaneous facial mimicry[4,25], we recruited
477 only young, female participants. Future studies investigating the effect of age, gender and
478 socioeconomic status of dyadic pairs could be helpful in determining the generalizability of
479 our findings.

480 In conclusion, our results provided clear empirical evidence for the theoretical claims
481 about the ecological validity of switching from a "third-person" approach to a "second-person"
482 approach in social neuroscience studies[23]. Our findings demonstrated changes in activation
483 dynamics and functional organization at both network and regional levels, suggesting that the
484 "second-person" approach could bring additional neurocognitive insight for neuroimaging
485 studies in social neuroscience.

486 Methods

487 **Participants.** Fifty-one women from the Japanese cities of Kyoto and Osaka were recruited to
488 this study in February and March 2021 via email lists and through online advertisements and
489 flyers. This study was approved by the ethics committee of the Unit for Advanced Studies of
490 the Human Mind, Kyoto University (approval number: 30-P-7), and performed in accordance
491 with the ethical standards of the committee. Written informed consent was obtained from all
492 participants prior to their participation in the study. All participants consented to the publication
493 of their data, and to being recorded on video. All participants were rewarded monetarily. After

494 excluding five participants who fell asleep in more than 3 out of 5 runs of the passive viewing
495 fMRI tasks and two who did not finish the fMRI tasks, 44 participants were included in the
496 ratings and fMRI analysis (mean ± SD age = 22.82 ± 2.59 years; range: 20-29 years). EMG
497 recordings of one of the 44 participants failed due to equipment setup error whereby neither
498 scanner pulse nor experimental condition timing markers were stamped in the data, and it was
499 impossible to reconstruct precisely, leaving 43 participants for the EMG and fMRI parametric
500 modulation analysis.

501 We used the simr package[70], based on the zygomaticus major EMG linear mixed-effect
502 model analysis from a previous study using the same paradigm[13], to simulate the power of the
503 interaction effect between emotion and presentation condition with extended participant
504 numbers up to 55. The simulation demonstrated that for the interaction effect of zygomaticus
505 major EMG data to reach a power of 80%, more than 17 participants were needed. No *a prior*
506 power estimation was done for the sample size required for the fMRI statistics of the passive
507 viewing task, as the effect size of the presentation condition in this paradigm could not be
508 approximated by previous neuroimaging studies. However, we opted for 50, and finally
509 included 44, which was at the top 20% in terms of sample size for recent experimental fMRI
510 studies[71]. Using NeuroPowerTools[72], we used the statistical t-map of a previous study with the
511 functional localizer task as in the present study. Power calculation showed that a sample size
512 of 30 had a power of 0.85. Note that the NeuroPowerTools only considers peak inference, and
513 the power for cluster inference was estimated assuming there was no peak-level correction.

514 **Materials**

515 ***Pre-recorded videos.*** Similar to a previous study[13], two female models each recorded more
516 than 40 video clips of positive (smiling) and negative (frowning) dynamic facial expressions.
517 The appearance of the models was kept consistent between the video clips and the live stimuli.

The clips lasted 3 s, and consisted of a neutral expression, gradual dynamic changes, and sustained maximal emotional facial expression, for 1 s each. Fifty smiling and fifty frowning clips were used for the passive viewing component; two per condition were used for the practice ratings, and four per condition were used for the actual ratings. Ten clips were chose at random, per condition, for each run of passive viewing.

Facilities The model outside the scanner faced a prompter with a concealed VIXIA HF R800 camera (Canon Inc., Tokyo, Japan). A camera in the scanner captured the "live-relay images" of the participant to the model. Presentation version 21 (Neurobehavioral Systems, Inc., Berkeley, CA, USA) was used to run the experiments on a Precision T3500 computer (Dell Inc., Round Rock, TX, USA) with the Windows 7 Professional operating system (Microsoft Corp., Redmond, WA, USA). Serial signals were sent to an SL-41C switcher (Imagenics Co., Ltd., Tokyo, Japan), to switch the visual input of the scanner screen between the computer and the model's camera. Presentation software transmitted sounds to the models via a pair of computer loudspeakers, instructing them to produce dynamic facial expressions. The resolution (1280 × 960) and height-to-width ratio of the video output of the Presentation software and live-relay images were kept consistent by the switcher. The BrainAmp ExG MR amplifier (Brain Products, Munich, Germany) and BrainVision Recorder software (Brain Products) were used to record EMG from the ZM and CS of the participants.

Paradigm and procedures As in a previous study[13], a 2 × 2 study design was used, with factors of presentation (video vs. live) and emotion (positive vs. negative), resulting in four conditions. After providing written informed consent and the attachment of EMG electrodes, the participant was placed inside the scanner with a microphone. The participant and model engaged in audiovisual conversation for 3 min through the prompter and interphone. Subsequently, the structural T1 image was acquired.

542 ***Passive Viewing Task.*** The participants fixated on a cross (mean intertrial interval = 4475 ms;
543 range: 2950 - 10150 ms) until they were shown the instructions for either "Video" or "Real
544 Person" for 1.25 s. During the video trials, one pre-recorded video clip was presented
545 immediately after the instructions. During the live trials, the models performed dynamic facial
546 expressions according to the instructions given across the loudspeaker. The screen displayed
547 the fixation cross again after the facial expressions had been shown (Fig. 1B). The participants
548 performed eight practice trials without data acquisition. Five runs of functional task runs were
549 performed, during which multi-echo EPI and EMG data were acquired. Each run included 10
550 passive viewing trials per condition and lasted 361.6 s (113 volumes). Due to input error in text
551 files indicating the condition sequences, in some runs there were 11 "positive-live" trials and
552 9 "negative-live" trials. This occurred in 2 out of 5 runs for 14 participants, and in 1 out of 5
553 runs for another 14 participants. The data acquisition was stopped, and participants could take
554 a break between runs. A total of 200 trials were acquired per participant. For each participant,
555 the sequence of conditions during the trial was pseudorandomized, and the presentation
556 sequence of the pre-recorded videos per condition was randomized. The pseudorandomized
557 presentation sequences and jittered intertrial intervals were optimized using optseq2 based on
558 the cost function of the average variance reduction factor after 300,000 simulations[73]. Both the
559 model and the participants were videotaped during data acquisition.

560 ***Functional Localizer Task.*** The functional localizer task was modified from Sato et al.[27] using
561 identical video clips of dynamic facial expressions (happy and angry) and dynamic mosaic
562 images. Each clip lasts 1.52 s. In each epoch, the participant saw eight clips of the same
563 condition with fixation cross presented for 1.68 s between clips. The onset of each clip was
564 synchronized with the scanner pulse signal. Each epoch lasted for 25.6 s, and four epochs per
565 condition were presented. The first fixation cross lasted for 35.2 s (11 TRs) until the onset of
566 the first epoch. The inter-epoch interval and the duration of the final fixation cross were both

19.2 s (6 TRs). This task lasted for 752 s (235 volumes).

Valence and Arousal Ratings. Each trial began in the same manner as described for the passive viewing task, and valence and arousal ratings[28] were given verbally from inside the scanner by the participant after presentation of the dynamic facial expressions. Valence and arousal were rated on Likert scales (range: 1–9) according to the following instruction: "When viewing the images, please rate the emotion that you feel." Four practice rating trials were followed by 16 test rating trials. No fMRI or EMG data were acquired during this period. Subsequently, the participant was removed from the scanner and electrodes were removed. The participants were then monetarily rewarded and dismissed.

MRI data acquisition Data were acquired using a 3T Siemens MAGNETOM Verio scanner with a 12-channel phased array coil. We acquired an MPRAGE scan with T_1 weighted contrast [176 ascending sagittal slices with A/P phase encoding direction; voxel size = 1mm isotropic; FOV = 256 mm; TR = 2250 ms; TE = 3.06 ms; acquisition time = 314 s; flip angle = 9°; GRAPPA in-plane acceleration factor = 2; complete brain coverage for cerebrum, cerebellum, and brainstem]. After the T_1, we acquired five functional runs of T_2* weighted echo planar sequence images of three echoes [40 interleaved axial slices with A/P phase encoding direction; voxel size = 3×3×3 mm; FOV = 192 mm; TR = 3200 ms; TEs = 20, 37.29, 54.58 ms; 113 volumes; acquisition time (TA) = 361.6 s; flip angle = 90°; GRAPPA in-plane acceleration factor = 2; complete brain coverage for cerebrum, cerebellum, and brainstem]. The MR sequence was compatible with the specific absorption rate requirement of BrainAmp ExG for simultaneous physiological-fMRI measurements (one pulse per slice, flip angle not larger than 90°, TR larger than 2000 ms, and no more than 25 slices per 2000 ms). The same multi-echo sequence was used to acquire 235 volumes (TA = 752 s) for the functional localizer task.

Live stimuli validation We recorded and visually inspected the live performances of the

models during the passive viewing trials, to ensure valid performance of dynamic facial expressions. Three trials were excluded as the models performed incorrect facial expressions.

Participant's scanner performance check We recorded and visually inspected video recordings of participants inside the scanner. Runs in which the participant closed their eyes for more than 50% of the time were excluded from EMG and fMRI data analysis. If three or more runs had to be excluded, the participant was completely excluded from EMG and fMRI data analysis. Five participants were excluded for this reason. Among the remaining 44 participants, four runs for eight participants and three runs for seven participants were included in the analysis.

EMG data pre-processing Raw EMG data were first pre-processed using BrainVision Analyzer 2. MR gradient artifacts were removed based on the pulse trigger, and data were down-sampled to 500 Hz. Data were then imported and pre-processed using the EEGLAB MATLAB toolbox (version 2019.1; Swartz Center for Computational Neuroscience, San Diego, CA, USA). A high-pass filter was applied at 20 Hz. For each trial, the signal was detrended and the baseline was corrected for low frequency drift. The baseline for this step was defined as the mean of values from 3 s before to 1 s after stimulus onset, which was 2 s before and after the onset of the trial/instruction. For each signal, the absolute or rectified value plus one was natural log-transformed to correct for the right-skewness of the raw data distribution.

Statistical analyses of behavioral data LME models, model comparisons, and influence diagnostics were performed using R software (version 4.1.1; R Core Team. Vienna, Austria) using the packages lme4 1.1.27.1, lmerTest 3.1-3, HLMdiag 0.5.0, emmeans 1.6.3, simr 1.0.5, and optimizer BOBYQA. The dependent variables included valence ratings, arousal ratings, and the differences in EMG (ZM and CS) between the neutral (0-1 s after stimulus onset) and maximal phases (2.5-3.5 s after stimulus onset) of the dynamic facial expression. For each

LME model, emotion, presentation condition, and their interactions were fixed effects, while the participant was a random factor. Model complexity, in terms of the number of random effects, was increased stepwise; each model was compared with the previous (i.e., less complicated) model using F-tests [lmerTest::anova()] until the more complex model was no longer significantly superior to the simpler one. When the model estimation resulted in a singular fit, the model iteration process was considered complete, and the simpler model was chosen as the final model. Initially, random intercepts for participants were included; the inclusion of by-subject random slopes for the effects of emotion, presentation condition, and their interaction term depends on the results of the model comparison process.

Homogeneity of residual variance is an *a priori* assumption to ensure valid inferences from LME models [31,74]. For LME model diagnostics, upward residual and influence analyses were performed using the HLMdiag package[75]. Trials with an absolute standardized residual larger than 3, or a Cook's distance larger than 1.5 × the interquartile range (IQR) above the third quartile ($Q_3 + 1.5 \times IQR$), were excluded. After the reduced data had been re-fitted, the Cook's distance was checked to exclude highly influential participants.

Satterthwaite's formula was used to evaluate the degrees of freedom (df). Effect size estimation of the semi-partial R-squared (variance explained) of fixed effects was carried out using r2glmm 0.1.2 with the Kenward-Roger approach[76–78]. Interactions between emotion categories and presentation conditions (live vs. video) were evaluated by simple effects analysis using the emmeans function. Two-tailed statistical *p*-values are reported throughout the text.

fMRI data preprocessing DICOM files were imported into 4D Nifti files using dcm2niix_afni v1.0.20181125. Multi-echo EPI preprocessing were performed in tedana v0.0.8 (https://doi.org/10.5281/zenodo.4725985) and SPM12 v7771

(http://www.fil.ion.ucl.ac.uk/spm). Realignment parameters were estimated for the first echo volumes of all runs and using the participant's first echo volume of the first functional run as the reference. Slice timing was corrected toward the first slice (acquired at the mid-point of the TR) of each frame. Realignment parameters estimated using the first echo volumes were then applied to all echo volumes of the same run and resliced while retaining the original resolution (3×3×3 mm).

The TE-dependence processing workflow of tedana[79,80] (https://doi.org/10.5281/zenodo.5227302) was as follows. For each run, an initial mask was generated from the first echo using nilearn's compute_epi_mask function. An adaptive mask was then generated, in which each voxel's value reflected the number of echoes with "good" data. A monoexponential model was fit to the data at each voxel using log-linear regression to estimate T2* and S0 maps. For each voxel, the value from the adaptive mask was used to determine which echoes would be used to estimate T2* and S0. Multi-echo data were then optimally combined using the "t2s" combination method[81]. Global signal regression was applied to the multi-echo and optimally combined datasets. Principal component analysis followed by the Kundu component selection decision tree[80] was applied to the optimally combined data for dimensionality reduction. Independent component analysis was then used to decompose the dimensionally reduced dataset. A series of TE-dependence metrics were calculated for each ICA component, including Kappa, Rho, and variance explained. Next, component selection was performed to identify blood-oxygen-level-dependent (BOLD, TE-dependent), non-BOLD (TE-independent), and uncertain (low-variance) components using the Kundu decision tree (v2.5)[80]. T1c global signal regression was then applied to the data to remove spatially diffuse noise. This workflow used numpy[82], scipy (http://www.scipy.org/), pandas[83], scikit-learn[84], nilearn, nibabel (http://doi.org/10.5281/zenodo.3233118), and the Dice similarity index[85,86] in a conda environment with python 3.6.12.

Using SPM12, the structural image was then coregistered to the first frame of the denoised EPI of the first functional run, and segmented[87] into gray matter and white matter. Forward deformation parameters to the MNI 152 space were estimated using the geodesic shooting registration[88] to a geodesic shooting template provided by the Computational Anatomy Toolbox 12. Images were normalized with the 4th degree B-Spline Interpolation algorithm in the original resolution (3×3×3 mm) and further smoothed with a Gaussian kernel of 6 mm full-width-at-half-maximum.

fMRI data analysis GLM and DCM analysis were performed with SPM12. In the GLM analysis, the design matrix contained four psychological regressors of interest, the "video-positive", "video-negative", "live-positive", and "live-negative" conditions, specifying the onsets of the instruction ("Video" or "Real Person"). The duration of each trial was 4.25 s including 1.25 s of the instruction and 3 s of video/live dynamic facial expression. We included six motion parameters as regressors of no interest. We then applied a high-pass filter with a cutoff period of 128 s, and the temporal autocorrelation was accounted for with the FAST option in SPM12[89]. Then, we calculated fixed effects across all runs for each subject. Contrasts of the average beta-images for all four conditions, as well as contrasts of [positive > negative], [live > video] and [(live-positive > live-negative) > (video-positive > video-negative)] were calculated at the first level for each subject. At the group-level, four random-effect one-sample *t* tests (n = 44) were performed for the average activation across all conditions, for the main effects of emotion and presentation condition, and the interaction effect. We applied whole-brain cluster-level FWE-corrected $p < 0.05$, using a cluster-defining threshold of $p = 0.001$ to control for multiple comparisons[90,91].

To investigate whether activities in the right IFG and pSTS modulated by the presentation conditions were associated with facial mimicry, we estimated the parametric

effects of congruency-weighted contrast of ZM and CS responses. Congruent EMG responses for positive expressions included contraction (positive response) of ZM and relaxation (negative response) of CS, and vice versa for negative expressions[32,33]. For each trial, congruent ZM and CS responses were coded as 1, while incongruent responses were coded as 0. The trial-wise mimicking EMG responses (for positive expressions) were estimated as congruency-weighted contrast of zygomaticus and corrugator responses, and operationalized as [ZM congruency (1 or 0) × the run-wise normalized ZM responses] − [CS congruency (1 or 0) × the run-wise normalized CS responses]. In the GLM analysis, the design matrix contained one psychological regressor of interest, the "dynamic facial expressions", including trials from all four conditions. The duration of each trial was 4.25 s. The trial-wise congruent EMG responses were used as the parametric modulator, including all trials in runs that were included in fMRI analysis. We included six motion parameters as regressors of no interest. We then applied a high-pass filter with a cutoff period of 128 s, and the FAST option to account for the temporal autocorrelation[89]. Then, the beta maps of the parametric effect across all runs for each subject were taken to the group-level for the random-effect one-sample *t* tests (n = 43). We applied whole-brain cluster-level FWE-corrected $p < 0.05$, using a cluster-defining threshold of $p = 0.001$ to control for multiple comparisons. We then specifically tested the two clusters significant for the [live > video] contrast, the right IFG and pSTS, using small volume correction (SVC)[34] with functional masks of each of the cluster, applying peak-level FWE-corrected $p < 0.05$ to control for multiple comparisons.

In the DCM analysis[92,93] (n = 44), we concatenated all runs of functional data per subject, and defined a single-run subject-level GLM model. The design matrix included one condition of interest, "faces", including all trials of stimuli presentation of 4.25 s, and one parametric regressor "presentation condition", coding "live" trials as 1 and "video" trials as -1. Regressors of no interest included six motion parameters and run regressors to partial out

effects of different runs. Two regions of interest, the pSTS and the IFG, were defined based on the functional mask of the group GLM result of the main effect of presentation condition [live > video]. The eigenvariate of time series in each ROI was extracted. In the full deterministic DCM model, we defined mutual bilinear connectivity as existent between two ROIs (A matrix). The condition "faces" served as the psychological input to the pSTS (C matrix). The parametric modulator "presentation condition" (live vs. video) was set to modulate both between-region connections and self-connections (B matrix). The full DCM model of each subject was inverted for parameter estimation[35]. The group mean of the estimated full DCM model parameters were used as empirical priors to re-estimate the parameters to avoid local maxima for higher accuracy[36]. At the group-level, a parametric empirical Bayes (PEB) model was set up to evaluate the posterior density over model parameters from single subject-model inversions. Bayesian model reduction[94] then provided the evidence and parameter distributions of all nested DCM models. Model evidence of all nested models were compared, and Bayesian model averaging[95] yielded the weighted-average model of winning nested models. Only parameters with posterior probability of parameter different from zero larger than 0.95 are reported.

GLM analysis was performed for the localizer task. We included 30 participants who were included in the task fMRI analysis and were awake in the localizer task (i.e., 14 of 44 participants who were included in the task fMRI analysis fell asleep for more than 50% of the time during the localizer scan). The design matrix contained four psychological regressors of interest, the "dynamic smiling expressions", "dynamic frowning expressions", "dynamic mosaic of smiling expressions", and "dynamic mosaic of frowning expressions" conditions, specifying the onset of the video clips. The duration of each trial was 1.52 s. We included six motion parameters as regressors of no interest, applied a high- pass filter with a cutoff period of 128 s to the first-level fixed effect analysis, and accounted for the temporal autocorrelation

738 with the FAST option in SPM12. The contrast [dynamic expressions > dynamic mosaics] was
739 calculated at the first level for each subject. At the group-level, a random-effect one-sample *t*
740 test (n = 30) for the contrast was performed. We applied whole-brain cluster-level FWE-
741 corrected $p < 0.05$, using a cluster-defining threshold of $p = 0.001$.

742 We used the Pattern Recognition for Neuroimaging Toolbox (PRoNTo)[37] v2.1 to test
743 whether voxel activation covariation patterns within regions involved in dynamic facial
744 expression processing, but showing no amplitude difference between live vs. video conditions
745 (i.e., right aTL, bilateral mPFC, PCC, amygdala and hippocampus, Fig. 5A), could be
746 differentiated between the live vs. video condition. The ROI was defined by one functional
747 mask of the group-level GLM contrast [dynamic face > dynamic mosaic] in the functional
748 localizer task, and excluding right IFG, right posterior temporal lobe and bilateral visual areas
749 which were also observed in the univariate result of the contrast [live > video] in the passive
750 observation task. For all 44 participants, non-smoothed functional imaging data within the ROI
751 mask were linearly detrended and mean centered. A design matrix containing "video" and "live"
752 conditions were used for volume selection. Under the assumption of a hemodynamic response
753 delay of 6 s and dispersion of 6 s, volumes in which the BOLD signal corresponded to more
754 than one condition were excluded from this analysis. A binary SVM was implemented to
755 classify volumes containing BOLD signal corresponds to "live" vs "video" conditions, and a
756 5-fold cross-validation on subjects (with 9, 9, 9, 9, and 8 subjects per fold) was employed.
757 Nested 5-fold cross-validation on blocks within each subject was further implemented for the
758 SVM C parameter optimization, and, as a result, C = 0.1 was adopted. A permutation of 10,000
759 repetitions was performed to determine whether the balanced classification accuracy over the
760 whole group was higher than the level of chance (50%).

761 To infer the contributions of different anatomical regions to the differentiable

762 multivariate patterns further, we exploratorily performed a whole-brain multiple kernel
763 learning (MKL) implemented in PRoNTo with the SimpleMKL package[96]. MKL first used *a*
764 *priori* knowledge about brain anatomy, defined by the Automated Anatomical Labeling
765 (AAL)[39], to segment the whole-brain multivariate pattern into regional patterns. The voxels of
766 each of the m brain regions of the AAL constituted a basis kernel, and the whole-brain kernel
767 was a linear combination of m basis kernels by the contribution vector, d_m, in the size of m,
768 specifying the contribution of each basis kernel to the model. All elements of d_m were ≥ 0, and
769 the elements of d_m summed to 1. SimpleMKL simplified the optimization problem to an SVM
770 machine solving the whole-brain kernel and gradient descents to solve d_m. Brain regions were
771 then ranked according to the d_m, i.e., the percentage of the whole-brain normalized weights
772 explained by each region[40,41].

773 Acknowledgement

774 This project was generously supported by the Japan Science and Technology Agency CREST
775 (JPMJCR17A5) and JST Mirai Program (JPMJMI20D7). The authors would like to thank
776 Mami Fujikura, Yuko Kuroda, Kazusa Minemoto, Mei Nakabayashi, and Masaru Usami (in
777 alphabetic order) for their assistance. This study was conducted using the MRI scanner and
778 related facilities of the Kokoro Research Center, Kyoto University.

779 Author Contribution

780 C.-T. H., W.S. T.K. and S.Y. conceived the experiment. C.-T. H., W.S., K.A., N.A. and R.N.
781 prepared the experimental setup. C.-T. H. conducted the experiments and analyzed the results.
782 C.-T. H., W.S. T.K., K.A., N.A., R.N. and S.Y prepared the manuscript.

783 Competing Interests

784 The authors declare no competing financial or non-financial interests.

785 Data availability

The following data tables and codes are available on the Open Science Framework (https://osf.io/vybwm/?view_only=89cc060a240a473588cbbfeb2e6173dc): (1) Rating data table; (2) EMG response tables with extracted EMG response values and cleaned up by excluding runs in which participants fell asleep and trials with inaccurate live facial expression performances; and (3) R codes for statistical analyses of rating and EMG data. The MRI datasets generated during and/or analyzed during the present study are not publicly available because the ethical approval and the informed consent did not include permission for neuroimaging data to be publicly deposited but are available from the corresponding author W.S. (wataru.sato.ya@riken.jp) on reasonable request.

References

1. Adolphs, R. Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav Cogn Neurosci Rev* **1**, 21–62 (2002).
2. Fernández-Dols, J.-M. Natural facial expression: A view from psychological constructionism and pragmatics. in *The science of facial expression* 457–475 (Oxford University Press, 2017).
3. Dimberg, U. Facial reactions to facial expressions. *Psychophysiology* **19**, 643–7 (1982).
4. Hess, U. & Fischer, A. Emotional Mimicry as Social Regulation. *Pers Soc Psychol Rev* **17**, 142–157 (2013).
5. Arnold, A. J. & Winkielman, P. The Mimicry Among Us: Intra- and Inter-Personal Mechanisms of Spontaneous Mimicry. *J Nonverbal Behav* **44**, 195–212 (2020).
6. Gonzalez-Liencres, C., Shamay-Tsoory, S. G. & Brüne, M. Towards a neuroscience of empathy: Ontogeny, phylogeny, brain mechanisms, context and psychopathology. *Neurosci Biobehav Rev* **37**, 1537–1548 (2013).
7. Prochazkova, E. & Kret, M. E. Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neurosci Biobehav Rev* **80**, 99–114 (2017).
8. Herrando, C. & Constantinides, E. Emotional Contagion: A Brief Overview and Future Directions. *Front. Psychol.* **12**, 712606 (2021).
9. Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z. & Gvirts, H. Z. Herding Brains: A

Core Neural Mechanism for Social Alignment. *Trends Cogn. Sci.* **23**, 174–186 (2019).

10. Sato, W. & Yoshikawa, S. Enhanced Experience of Emotional Arousal in Response to Dynamic Facial Expressions. *J Nonverbal Behav* **31**, 119–135 (2007).

11. Sato, W. & Yoshikawa, S. Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* **104**, 1–18 (2007).

12. Hietanen, J. K., Kylliäinen, A. & Peltola, M. J. The effect of being watched on facial EMG and autonomic activity in response to another individual's facial expressions. *Sci Rep* **9**, 14759 (2019).

13. Hsu, C.-T., Sato, W. & Yoshikawa, S. Enhanced emotional and motor responses to live versus videotaped dynamic facial expressions. *Sci Rep* **10**, 16825 (2020).

14. Allison, T., Puce, A. & McCarthy, G. Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* **4**, 267–278 (2000).

15. Zinchenko, O., Yaple, Z. A. & Arsalidou, M. Brain Responses to Dynamic Facial Expressions: A Normative Meta-Analysis. *Front. Hum. Neurosci.* **12**, 227 (2018).

16. Iacoboni, M. & Dapretto, M. The mirror neuron system and the consequences of its dysfunction. *Nat Rev Neurosci* **7**, 942–51 (2006).

17. Hamilton, A. F. de C. Emulation and Mimicry for Social Interaction: A Theoretical Approach to Imitation in Autism. *Q. J. Exp. Psychol.* **61**, 101–115 (2008).

18. Schurz, M. *et al.* Toward a hierarchical model of social cognition: A neuroimaging meta-

analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* (2020) doi:10.1037/bul0000303.

19. Seeley, W. W. *et al.* Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* **27**, 2349–56 (2007).
20. Walter, H. Social Cognitive Neuroscience of Empathy: Concepts, Circuits, and Genes. *Emot Rev* **4**, 9–17 (2012).
21. Gallese, V. & Sinigaglia, C. What is so special about embodied simulation? *Trends Cogn. Sci.* **15**, 512–519 (2011).
22. Elfenbein, H. A. The many faces of emotional contagion. *Organ. Psychol. Rev.* **4**, 326–362 (2014).
23. Redcay, E. & Schilbach, L. Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat Rev Neurosci* **20**, 495–505 (2019).
24. Shamay-Tsoory, S. G. & Mendelsohn, A. Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspect Psychol Sci* **14**, 841–859 (2019).
25. Seibt, B., Mühlberger, A., Likowski, K. U. & Weyers, P. Facial mimicry in its social setting. *Front. Psychol.* **6**, (2015).
26. Kundu, P. *et al.* Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals. *NeuroImage* **154**, 59–80 (2017).
27. Sato, W. *et al.* Widespread and lateralized social brain activity for processing dynamic

852 facial expressions. *Hum. Brain Mapp.* **40**, 3753–3768 (2019).

853 28. Russell, J. A., Weiss, A. & Mendelsohn, G. A. Affect Grid: A Single-Item Scale of Pleasure

854 and Arousal. *J Pers Soc Psychol* **57**, 493–502 (1989).

855 29. Weaverdyck, M. E., Lieberman, M. D. & Parkinson, C. Tools of the Trade Multivoxel

856 pattern analysis in fMRI: a practical introduction for social and affective neuroscientists.

857 *Soc Cogn Affect Neurosci* **15**, 487–509 (2020).

858 30. Friston, K. J. Functional and effective connectivity: a review. *Brain Connect* **1**, 13–36

859 (2011).

860 31. Baayen, R. H. & Milin, P. Analyzing reaction times. *Int. j. psychol. res.* **3**, 12–28 (2010).

861 32. McIntosh, D. N. Spontaneous facial mimicry, liking and emotional contagion. *Pol. Psychol.*

862 *Bull.* **37**, 31–42 (2006).

863 33. McIntosh, D. N., Reichmann-Decker, A., Winkielman, P. & Wilbarger, J. L. When the

864 social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial

865 expressions in autism. *Dev Sci* **9**, 295–302 (2006).

866 34. Worsley, K. J. *et al.* A unified statistical approach for determining significant signals in

867 images of cerebral activation. *Hum. Brain Mapp.* **4**, 58–73 (1996).

868 35. Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *NeuroImage* **19**, 1273–

869 302 (2003).

870 36. Friston, K. J. *et al.* Bayesian model reduction and empirical Bayes for group (DCM) studies.

NeuroImage **128**, 413–31 (2016).

37. Schrouff, J. *et al.* PRoNTo: Pattern Recognition for Neuroimaging Toolbox. *Neuroinformatics* (2013) doi:10.1007/s12021-013-9178-1.

38. Golland, P. & Fischl, B. Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies. in *Information Processing in Medical Imaging* (eds. Taylor, C. & Noble, J. A.) vol. 2732 330–341 (Springer Berlin Heidelberg, 2003).

39. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–89 (2002).

40. Schrouff, J. *et al.* Localizing and Comparing Weight Maps Generated from Linear Kernel Machine Learning Models. in *2013 International Workshop on Pattern Recognition in Neuroimaging* 124–127 (2013). doi:10.1109/PRNI.2013.40.

41. Schrouff, J. *et al.* Embedding Anatomical or Functional Knowledge in Whole-Brain Multiple Kernel Learning Models. *Neuroinformatics* **16**, 117–143 (2018).

42. Sato, W., Kochiyama, T., Uono, S. & Yoshikawa, S. Amygdala integrates emotional expression and gaze direction in response to dynamic facial expressions. *NeuroImage* **50**, 1658–1665 (2010).

43. Sato, W., Toichi, M., Uono, S. & Kochiyama, T. Impaired social brain network for processing dynamic facial expressions in autism spectrum disorders. *BMC Neurosci.* **13**,

99 (2012).

44. Hirsch, J., Zhang, X., Noah, J. A. & Ono, Y. Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *NeuroImage* **157**, 314–330 (2017).

45. Rymarczyk, K., Żurawski, Ł., Jankowiak-Siuda, K. & Szatkowska, I. Neural Correlates of Facial Mimicry: Simultaneous Measurements of EMG and BOLD Responses during Perception of Dynamic Compared to Static Facial Expressions. *Front. Psychol.* **9**, 52 (2018).

46. Friederici, A. D. Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* **13**, 175–181 (2009).

47. Frith, C. D. The role of metacognition in human social interactions. *Phil. Trans. R. Soc. B* **367**, 2213–2223 (2012).

48. Conty, L., George, N. & Hietanen, J. K. Watching Eyes effects: When others meet the self. *Conscious Cogn* **45**, 184–197 (2016).

49. Boschin, E. A., Piekema, C. & Buckley, M. J. Essential functions of primate frontopolar cortex in cognition. *P Natl Acad Sci USA* **112**, E1020–E1027 (2015).

50. Mansouri, F. A., Koechlin, E., Rosa, M. G. P. & Buckley, M. J. Managing competing goals - a key role for the frontopolar cortex. *Nat. Rev. Neurosci.* **18**, 645–657 (2017).

51. Roca, M. *et al.* The role of Area 10 (BA10) in human multitasking and in social cognition:

A lesion study. *Neuropsychologia* **49**, 3525–3531 (2011).

52. Bramson, B. *et al.* Human Lateral Frontal Pole Contributes to Control over Emotional Approach–Avoidance Actions. *J. Neurosci.* **40**, 2925–2934 (2020).

53. Bang, D., Ershadmanesh, S., Nili, H. & Fleming, S. M. Private–public mappings in human prefrontal cortex. *eLife* **9**, e56477 (2020).

54. Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J. & Barrett, L. F. The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cereb. Cortex* **26**, 1910–1922 (2016).

55. Liu, M., Liu, C. H., Zheng, S., Zhao, K. & Fu, X. Reexamining the neural network involved in perception of facial expression: A meta-analysis. *Neurosci Biobehav Rev* **131**, 179–191 (2021).

56. Wood, A., Rychlowska, M., Korb, S. & Niedenthal, P. Fashioning the Face: Sensorimotor Simulation Contributes to Facial Expression Recognition. *Trends Cogn. Sci.* **20**, 227–240 (2016).

57. Wolpert, D. M., Doya, K. & Kawato, M. A unifying computational framework for motor control and social interaction. *Phil. Trans. R. Soc. Lond. B* **358**, 593–602 (2003).

58. Golland, Y., Arzouan, Y. & Levit-Binnun, N. The Mere Co-Presence: Synchronization of Autonomic Signals and Emotional Responses across Co-Present Individuals Not Engaged in Direct Interaction. *PLoS ONE* **10**, e0125804 (2015).

59. Gross, J. J. Emotion Regulation: Current Status and Future Prospects. *Psychol. Inq.* **26**, 1–26 (2015).

60. Hatfield, E., Bensman, L., Thornton, P. D. & Rapson, R. L. New Perspectives on Emotional Contagion: A Review of Classic and Recent Research on Facial Mimicry and Contagion. *Interpersona* **8**, 159–179 (2014).

61. Paulus, F. M., Müller-Pinzler, L., Westermann, S. & Krach, S. On the distinction of empathic and vicarious emotions. *Front. Hum. Neurosci.* **7**, (2013).

62. Gross, J. J. & Feldman Barrett, L. Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View. *Emot Rev* **3**, 8–16 (2011).

63. Dixon, M. L., Thiruchselvam, R., Todd, R. & Christoff, K. Emotion and the prefrontal cortex: An integrative review. *Psychol. Bull.* **143**, 1033–1081 (2017).

64. Hassabis, D. *et al.* Imagine All the People: How the Brain Creates and Uses Personality Models to Predict Behavior. *Cereb. Cortex* **24**, 1979–1987 (2014).

65. Vanderhasselt, M.-A., Kühn, S. & De Raedt, R. 'Put on your poker face': neural systems supporting the anticipation for expressive suppression and cognitive reappraisal. *Soc Cogn Affect Neurosci* **8**, 903–910 (2013).

66. Eimontaite, I. *et al.* Left Amygdala and Putamen Activation Modulate Emotion Driven Decisions in the Iterated Prisoner's Dilemma Game. *Front. Neurosci.* **13**, 741 (2019).

67. Uono, S. *et al.* Putamen Volume is Negatively Correlated with the Ability to Recognize

Fearful Facial Expressions. *Brain Topogr* **30**, 774–784 (2017).

68. Bavelas, J. & Chovil, N. Some pragmatic functions of conversational facial gestures. *GEST* **17**, 98–127 (2018).

69. Faso, D. J., Sasson, N. J. & Pinkham, A. E. Evaluating Posed and Evoked Facial Expressions of Emotion from Adults with Autism Spectrum Disorder. *J Autism Dev Disord* **45**, 75–89 (2015).

70. Green, P. & MacLeod, C. J. SIMR : an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol* **7**, 493–498 (2016).

71. Szucs, D. & Ioannidis, J. PA. Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage* **221**, 117164 (2020).

72. Durnez, J. *et al. Power and sample size calculations for fMRI studies based on the prevalence of active peaks*. http://biorxiv.org/lookup/doi/10.1101/049429 (2016) doi:10.1101/049429.

73. Dale, A. M. Optimal experimental design for event-related fMRI. *Hum Brain Mapp* **8**, 109–114 (1999).

74. Loy, A. & Hofmann, H. Diagnostic tools for hierarchical linear models: Diagnostic tools for hierarchical linear models. *WIREs Comp Stat* **5**, 48–61 (2013).

75. Loy, A. & Hofmann, H. HLMdiag: A Suite of Diagnostics for Hierarchical Linear Models

in R. *J. Stat. Soft.* **56**, (2014).

76. Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F. & Schabenberger, O. An R^2 statistic for fixed effects in the linear mixed model. *Statist. Med.* **27**, 6137–6157 (2008).

77. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* **4**, 133–142 (2013).

78. Jaeger, B. C., Edwards, L. J., Das, K. & Sen, P. K. An R^2 statistic for fixed effects in the generalized linear mixed model. *J. Appl. Stat.* **44**, 1086–1105 (2017).

79. Kundu, P., Inati, S. J., Evans, J. W., Luh, W. M. & Bandettini, P. A. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage* **60**, 1759–1770 (2012).

80. Kundu, P. *et al.* Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc Natl Acad Sci USA* **110**, 16187–16192 (2013).

81. Posse, S. *et al.* Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magn. Reson. Med.* **42**, 87–97 (1999).

82. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).

83. McKinney, W. Data Structures for Statistical Computing in Python. in 56–61 (PROC. OF THE 9th PYTHON IN SCIENCE CONF., 2010). doi:10.25080/Majora-92bf1922-00a.

84. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**, 2825–

2830 (2011).

85. Dice, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **26**, 297–302 (1945).

86. Sørensen, T. J. *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons.* (I kommission hos E. Munksgaard, 1948).

87. Ashburner, J. & Friston, K. J. Unified segmentation. *NeuroImage* **26**, 839–51 (2005).

88. Ashburner, J. & Friston, K. J. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage* **55**, 954–967 (2011).

89. Corbin, N., Todd, N., Friston, K. J. & Callaghan, M. F. Accurate modeling of temporal correlations in rapidly sampled fMRI time series. *Hum. Brain Mapp.* **39**, 3884–3897 (2018).

90. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *P Natl Acad Sci USA* **113**, 7900–7905 (2016).

91. Flandin, G. & Friston, K. J. Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Hum. Brain Mapp.* **0**, (2017).

92. Zeidman, P. *et al.* A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *NeuroImage* **200**, 174–190 (2019).

93. Zeidman, P. *et al.* A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage* **200**, 12–25 (2019).

94. Friston, K. J., Li, B., Daunizeau, J. & Stephan, K. E. Network discovery with DCM. *NeuroImage* **56**, 1202–21 (2011).

95. Trujillo-Barreto, N. J., Aubert-Vazquez, E. & Valdes-Sosa, P. A. Bayesian model averaging in EEG/MEG imaging. *NeuroImage* **21**, 1300–1319 (2004).

96. Rakotomamonjy, A., Bach, F. R., Canu, S. & Grandvalet, Y. SimpleMKL. *J Mach Learn Res* **9**, 2491–2521 (2008).