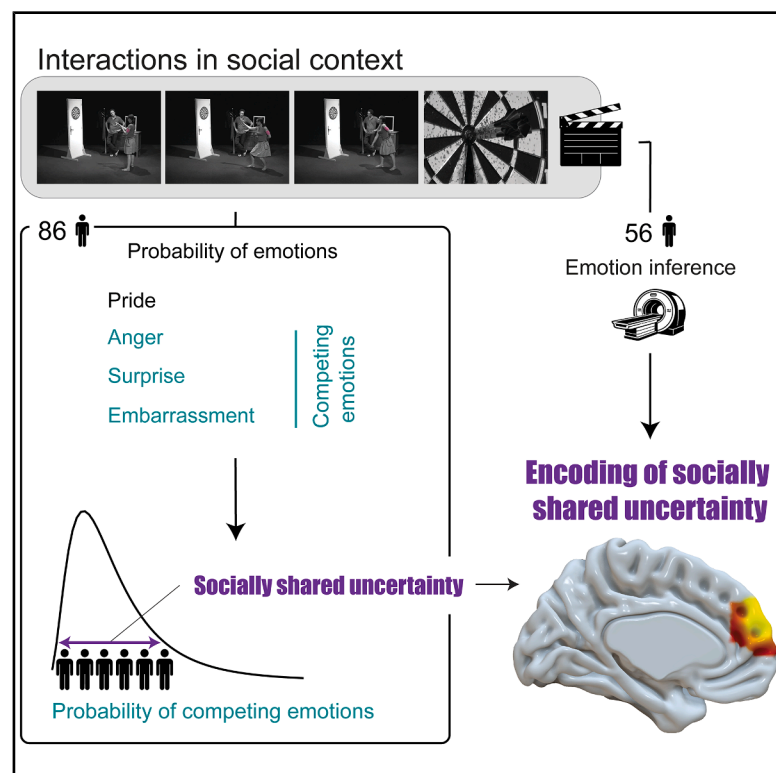


# Socially shared emotions shape the activity of the medial prefrontal cortex during inference of others' emotional states

## Graphical abstract



## Authors

Marine Le Petit, Pierre Gagnepain, Alexandre Bejanin, Vincent de La Sayette, Francis Eustache, Mickaël Laisney

## Correspondence

marine.lepetit@uliege.be (M.L.P.),  
mickael.laisney@ephe.psl.eu (M.L.)

## In brief

People differ in how they interpret social situations and infer others' emotions. This fMRI study shows that the medial prefrontal cortex integrates this social variability to infer others' emotional states, and indicates how aging affects this process.

## Highlights

- Inferring others' emotion during complex social scenes involves the mPFC
- mPFC integrates social variability of emotional predictions to guide emotional inference
- Social consensus on emotional conflict enhances representational similarity in mPFC
- Age reduces processing of emotional competition in the mPFC



## Article

# Socially shared emotions shape the activity of the medial prefrontal cortex during inference of others' emotional states

Marine Le Petit,<sup>1,2,3,\*</sup> Pierre Gagnepain,<sup>1</sup> Alexandre Bejanin,<sup>1,4,5</sup> Vincent de La Sayette,<sup>1</sup> Francis Eustache,<sup>1</sup> and Mickaël Laisney<sup>1,6,\*</sup>

<sup>1</sup>Normandie University, UNICAEN, PSL Université Paris, EPHE, INSERM, U1077, CHU de Caen, Centre Cyceron, Neuropsychologie et Imagerie de La Mémoire Humaine, 14000 Caen, France

<sup>2</sup>GIGA Research, CRC Human Imaging, University of Liège, 4000 Liège, Belgium

<sup>3</sup>Psychology and Cognitive Neuroscience, Department of Psychology, University of Liège, 4000 Liège, Belgium

<sup>4</sup>Sant Pau Memory Unit, Department of Neurology, Hospital de La Santa Creu I Sant Pau, Biomedical Research Institute Sant Pau, Universitat Autònoma de Barcelona, 08025 Barcelona, Spain

<sup>5</sup>Center of Biomedical Investigation Network for Neurodegenerative Diseases (CIBERNED), 28029 Madrid, Spain

<sup>6</sup>Lead contact

\*Correspondence: [marine.lepetit@uliege.be](mailto:marine.lepetit@uliege.be) (M.L.P.), [mickael.laisney@ephe.psl.eu](mailto:mickael.laisney@ephe.psl.eu) (M.L.)

<https://doi.org/10.1016/j.celrep.2026.117506>

## SUMMARY

The medial prefrontal cortex (mPFC) weighs options during decision-making, but its role in inferring competing emotional states and how this process changes with age, remain unclear. We recorded brain activity with functional MRI while 20 young and 40 older adults inferred the emotions of actors shown in ecological yet controlled social interactions. Socially shared representations about the timing, accuracy, and uncertainty of emotional inferences were defined using response distributions from an independent sample viewing the same scenes. mPFC activity is sensitive to emotional inference and scales with the social uncertainty of alternative emotional hypotheses. Social consensus regarding the emotional conflict elicited by the actors' interaction is associated with increased representational similarity in the mPFC, with this effect being modulated by age. Together, these results indicate that the mPFC integrates the variability of social representations to infer emotions in others and provide an account of age-related changes in social cognition.

## INTRODUCTION

Mindreading, the ability to attribute mental states, such as thoughts, intentions, and emotions to oneself or others,<sup>1,2</sup> involves inferring covert mental states in other agents based on decoding multiple cues from the social environment and using past experience and existing knowledge.<sup>3</sup> According to Bayesian models of the brain, this inference process entails assigning probabilities to hypotheses that best explain incoming information, allowing one to choose between different possibilities.<sup>4,5</sup> These models allow the brain to represent and resolve uncertainty through probabilistic inference.<sup>4,6</sup> Applied to social cognition, this framework posits that during social interactions, the brain forms probabilistic priors about others' future behavior, including about others' mental states such as emotions or intentions.<sup>7,8</sup> Because these hypotheses coexist as probability distributions, mindreading is accompanied by uncertainty about the possible mental states inferred. Critically, in the brain, such predictive signals are defined not only by their accuracy—the expected value of the prediction—but also by their associated variance reflecting uncertainty. Despite the ubiquity of this

uncertainty in everyday life social interactions,<sup>9</sup> how the brain represents the content of these predictions and uncertainty during mindreading remains poorly understood.

One brain region consistently involved in predictive inference and uncertainty processing is the medial prefrontal cortex (mPFC). The mPFC has been linked to predictive processes through the generation of hierarchical predictions,<sup>10</sup> and plays a central role in encoding uncertainty across several cognitive domains.<sup>11</sup> For instance, it contributes to resolving uncertainty during decision-making,<sup>12,13</sup> and executing uncertain decisions such as those encountered in card games with varying probabilities of reward.<sup>14,15</sup> These functions are consistent with Bayesian accounts of neural computation.<sup>16</sup>

Converging evidence further suggests that computations in mPFC may support decisions through competitive interaction between neural representations of alternative options.<sup>17–19</sup> Within this framework, representations associated with different options are thought to compete via mutual inhibition, allowing the system to compare alternatives and select a representation. Such competition-based dynamics provide a plausible mechanism through which the mPFC could resolve situations in which



multiple (affective) predictions are simultaneously activated under conditions of uncertainty. Much of the evidence supporting these competition-based mechanisms derives from electrophysiological, stimulation, or computational approaches, and it remains unclear how such representational competition is reflected in large-scale neural activity measured with functional magnetic resonance imaging (fMRI). However, this notion of competition in the mPFC extends naturally to mindreading. Affective mental states experienced during social interactions involve the co-activation of multiple emotions.<sup>20–22</sup> Consequently, mindreading requires activating and selecting between dominant and their alternative competing emotions.

From a computational perspective, the probability assigned to potential emotions can be aggregated across social agents to characterize two complementary signals forming the basis of socially shared representations.<sup>16,23</sup> First, the mean probability across individuals reflects the degree to which one emotion is inferred, which we refer to as the “accuracy” of emotional inference. However, the attribution of an emotion can significantly vary between individuals.<sup>24</sup> The variance of these probabilities across individuals hence captures the extent to which observers diverge in their inferences, reflecting a second computation signal, referred to as “social uncertainty” (i.e., inverse of precision). The influence of socially shared representations during the activation of competing emotions seems particularly relevant to understand the mPFC, given its role in processing multiple and competing signals. While people often agree on a situation’s dominant emotion, they may diverge on alternative interpretations—suggesting that greater social uncertainty about competing emotions may increase mPFC recruitment. Moreover, the role of socially shared representations in the mPFC is particularly relevant in light of evidence showing that this region stores schemas derived from shared social experiences, including shared beliefs,<sup>25</sup> as well as normative,<sup>26</sup> cultural,<sup>27</sup> and collective knowledge.<sup>28</sup> Such schematic representations of social inputs in the mPFC may support the integration of multiple sources of information to generate predictions about others’ emotions and behavior.

However, although brain regions supporting mindreading have been well characterized,<sup>29,30</sup> the neural representation of accuracy and uncertainty inherent to social inferences during everyday social situations is less well understood. Evidence points to a role of the mPFC in social prediction. The mPFC has been shown to anticipate others’ future mental states,<sup>8</sup> and to support social inference through probabilistic, Bayesian-like computations.<sup>16,31–35</sup> These studies suggest that the mPFC assigns values to competing hypotheses about social inferences and encodes participants’ confidence in their own social judgments. This body of work supports the compatibility between Bayesian models of brain function and mental state inference processes.<sup>16,31,36</sup> However, the influence of social accuracy and uncertainty on emotional inferences has, to our knowledge, never been systematically investigated.

A further challenge in addressing this question lies in the frequent conflation of objective uncertainty and meta-cognitive subjective confidence in the literature. While both concepts are related, several authors highlight a conceptual distinction.<sup>16,37</sup> Uncertainty can be defined as a property of the environment or

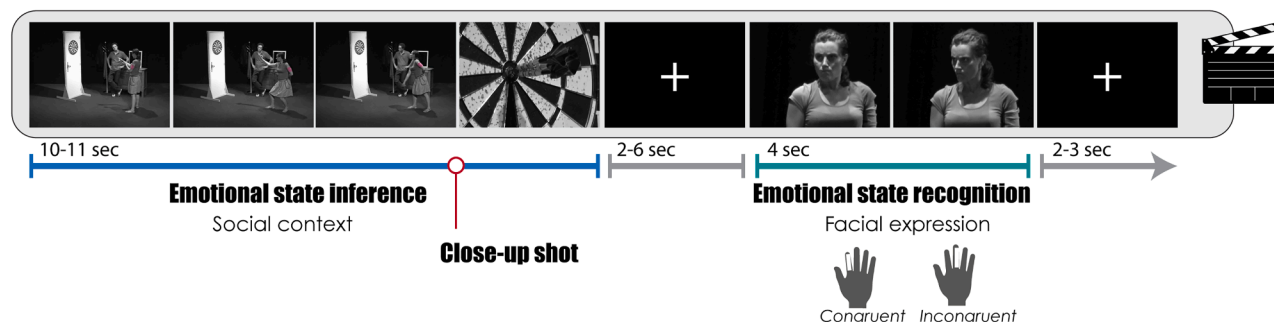
stimuli, such as ambiguity or variability in possible outcomes, whereas confidence reflects an individual’s metacognitive evaluation of their own decision accuracy. In the context of emotional mental state inference, uncertainty can be conceptualized as a socially shared index, representing the probability distribution of potential choices across individuals, in line with Bayesian inference frameworks.<sup>23,38</sup> This distinction between confidence and uncertainty allows uncertainty to be quantified independently of individual confidence reports, thereby isolating their respective contributions to mPFC activity. Importantly, this Bayesian approach captures implicit and automatic estimates of uncertainty, distinct from explicit, conscious metacognitive evaluations. To date, only one study has shown that mPFC recruitment is modulated by ambiguity inherent to daily life social scenarios during mindreading.<sup>39</sup> Hence, a formal demonstration that the accuracy and uncertainty of social inferences arise from the probabilistic structure of competing outcomes, rather than from metacognitive processes remains critically lacking.

Moreover, older individuals exhibit reduced performance in mindreading tasks compared with younger participants.<sup>40–42</sup> However, the nature of these changes remains subject to debate, especially with respect to the functions of the mPFC.<sup>41–43</sup> Age-related modifications in executive functions could account for the decrease observed in mindreading performances in this population.<sup>44,45</sup> In this study, we suggest a new account of the aging social brain and propose that aging may affect the encoding of competing emotional representations during mental state inference, influencing the computations of accuracy and uncertainty during social inference.

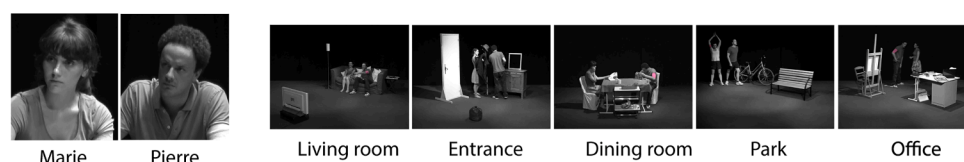
In the present study, we examined whether the mPFC responds to competing emotional representations during the inference of affective mental states arising from complex and natural social scenes. We hypothesized that the effect of emotional competition on the response of the mPFC during mindreading would depend on the statistical regularities (i.e., accuracy and uncertainty) of socially shared emotional inferences. Brain activity was recorded while participants inferred affective mental states in 20 young (mean age =  $24 \pm 2.7$  years), 20 middle-aged (mean age =  $40 \pm 5.9$  years), and 20 older (mean age =  $63.3 \pm 6.7$  years) healthy adults.

To address the inherent challenges posed by the natural variations and complexities of social inferences, our study combined several approaches: 1) we used an ecological task wherein participants inferred the putative affective mental states of two protagonists depicted in videos portraying everyday life social situations; and experiencing four possible emotions (i.e., surprise, anger, pride, and embarrassment) (Caillaud et al.<sup>46</sup>; Figure 1). For examples of two trials, see Video S1. For each video, one of the four emotions was considered as the dominant emotion (i.e., that the actor was supposed to feel) while the other three were regarded as competing emotions; (2) in the main study, we recorded brain activity using fMRI while participants performed affective mental states inferences throughout the course of these videos; (3) two separate pre-experimental behavioral studies involving 86 and 56 participants were conducted. These pre-experimental studies were designed to investigate the natural variations of possible emotional inferences associated with each affective mental state and

## A Affective mental state attribution task



## B Actors and places



**Figure 1. Affective mental state attribution task**

(A) Experimental design of the affective mental state attribution task (Caillaud et al., 2020).

This task involved an experimental design wherein videos were segmented into two parts.

The initial 10–11 “emotional state inference” seconds induced an affective mental state in one of the actors, requiring participants to infer this emotion.

This section was followed by 4-s “emotional state recognition” segment, showing the emotional facial expression of the target actor, which could be either congruent or incongruent with the preceding section. During that period, participants had to realize a congruency judgment by determining if the mental state expressed during the second segment matched, or not, with the preceding social context.

The second segment of the video, with the behavioral task, ensured that the participants were actively engaged in mindreading during the presentation of the context which is the main focus of this study.

(B) Actors and places. The task featured two protagonists, Marie and Pierre, presented as roommates. The scenes occurred in five different places (dining and living rooms, entrance, park, and office).

thus estimate the corresponding social accuracy and uncertainty signals (pre-study 1), along with the temporal distribution of the critical inference time-points in the videos (pre-study 2) (Figure 2 and STAR Methods). During the first pre-study, participants had to characterize the mental state felt by the protagonist by estimating the intensity associated with each possible emotion. The rating distribution across participants in this first study served as a proxy for socially shared representations of emotions felt by the protagonist. For each video and type of emotion (dominant and competing), the accuracy and uncertainty of inferred emotions were estimated using the mean and standard deviation of the distribution of emotional intensity across the tested population, respectively (Figure 2A). Four indices were computed, crossed in terms of parameters (i.e., accuracy and uncertainty) and type of emotion (i.e., dominant and competing). The second pre-study allowed to extract the temporal distribution of the critical inference time-points in the videos by asking participants to indicate the precise moment at which the affective mental state inference could be made. By combining accuracy and uncertainty estimates (pre-study 1) with the temporal distribution of mental state inferences (pre-study 2), we constructed predictors of the dynamics of neural activity reflecting those social computations (Figure 2B). In a first analysis, these estimates parametrically modulated the canonical

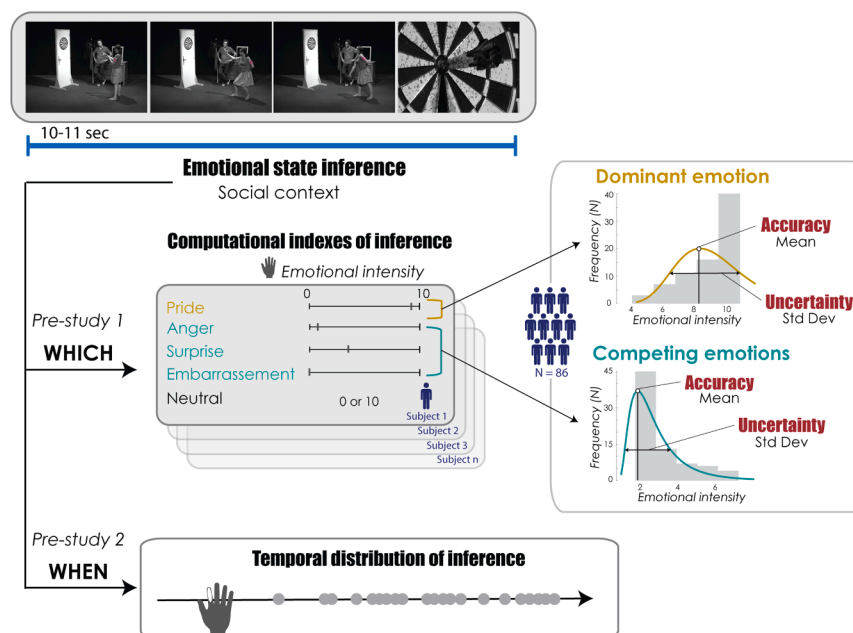
hemodynamic response function (HRF), from the main study ( $N = 56$ ), which was used to analyze the fMRI activity associated with the dynamic content of dominant and competing emotions. In a second analysis, using representational similarity analysis (RSA<sup>47,48</sup>), we sought to further elucidate the mechanisms underpinned by the mPFC by clarifying what aspect of emotional competition is represented in the mPFC. Specifically, we tested whether mPFC activity reflects a signal indexing the overall magnitude of competition among co-activated emotions (i.e., the interference caused by competing emotions), or instead captures the emotional content structure of that competition. In a third analysis, using psychophysiological interaction (PPI), we further investigated the central role of the mPFC by assessing its condition-dependent connectivity with other brain regions during emotional inference.

## RESULTS

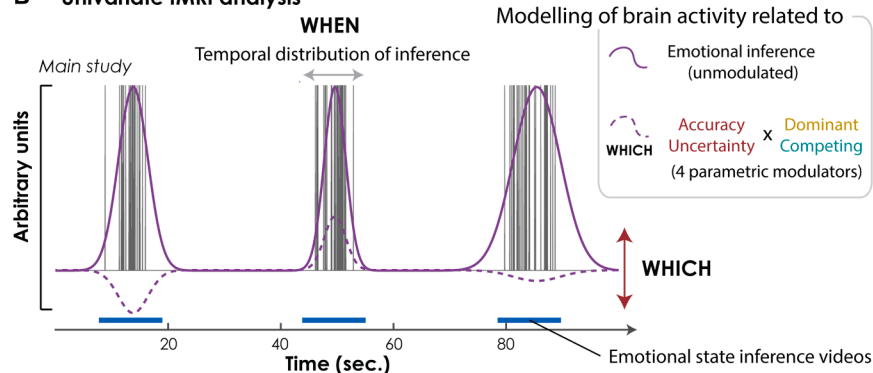
### Behavioral differences between young and older individuals

The congruency judgment measures, in addition to ensuring participants' engagement in the task, provided an indirect assessment of their ability to attribute affective mental states. The total number of responses, the correct response rate, and response

## A Pre-experimental studies



## B Univariate fMRI analysis



### Figure 2. Methodology of univariate analysis

(A) Measures extracted from the pre-experimental studies.

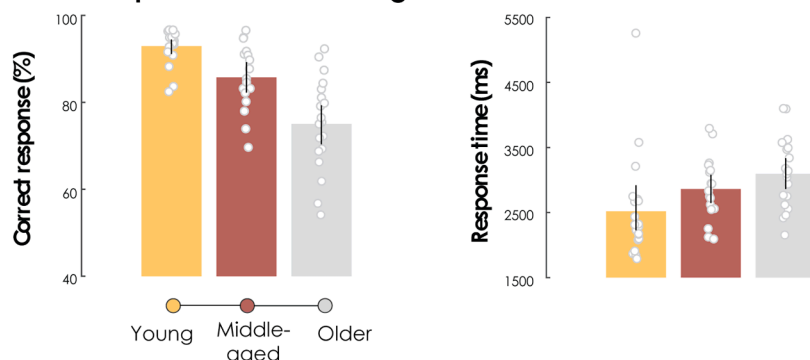
During the emotional state inference segment of the video, two separate behavioral experiments were performed to study the variation of possible emotional inferences associated with each affective mental state (“which”, pre-study 1), and their temporal distribution (“when”, pre-study 2). In pre-study 1, participants had to estimate the intensity associated with each possible emotion, allowing to estimate the probability of the dominant and competing affective states. Four indices were computed divided in accuracy (accuracy of dominant emotion and accuracy of competing emotions) and uncertainty (uncertainty of dominant emotion and uncertainty of competing emotions). Accuracy corresponded to the mean values of probability and uncertainty represented the associated standard deviations. The temporal distribution of inference was derived from the second pre-study wherein participants rated the precise moment at which the affective mental state inference could be made. Std Dev, standard deviation.

(B) Univariate fMRI analysis. A first analysis assessed brain activity associated with emotional inference. This model included regressors encoding the temporal distribution of inference (when) for both emotional and neutral video contexts, without further modulation by accuracy or uncertainty (solid purple line). A second analysis investigated brain activity associated with accuracy and uncertainty. This model combined the temporal distribution of inference for emotional contexts (when) together with regressors parametrically modulated by the accuracy and uncertainty of both dominant and competing emotions (which; dotted purple line). Hence, we constructed four predictors of the dynamics of neural activity that parametrically modulated the stimulus function used to analyze the fMRI activity associated with the dynamic content of mental state inference videos. A fifth unmodulated regressor represented the emotional inference temporal distribution.

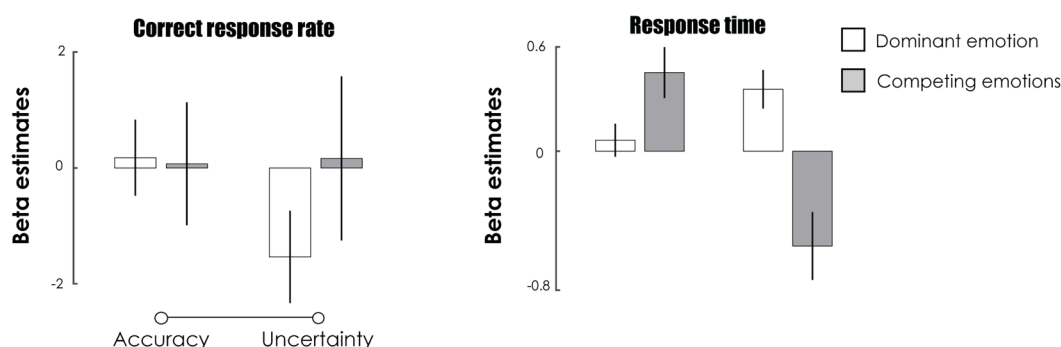
times (RTs) were compared across groups (Figure 3A) using the Kruskal-Wallis nonparametric analysis of variance tests (ANOVA). Post-hoc analyses were conducted using Mann-Whitney tests. A significant main effect of the group was

highlighted concerning the total number of responses ( $p = 0.017$ ,  $H(2,53) = 8.20$ ). Older subjects exhibited significantly fewer responses than younger ones ( $p = 0.009$ ). However, differences between the young and middle-aged groups ( $p = 0.066$ )

### A Behavioral performances during affective mental states attribution



### B Relationship between behavioral performances and statistical regularities



**Figure 3. Behavioral performance and relationships with statistical regularities**

(A) Behavioral performances across groups during the congruency judgment. Correct response rate (in percent) on the left and response time (in millisecond) on the right.

(B) Modulation of behavioral performances by computational indices of inference.

Beta estimates show the relationship between computational indices and correct response rates at the congruency judgment on the left and response time on the right.

Error bars reflect confidence intervals.

and between middle-aged and older subjects ( $p = 0.192$ ) were not statistically significant. Given the observed discrepancy in response rates between older and younger subjects, correct response rate variable was defined as the number of correct responses in the congruency judgment corrected by the total number of responses. Overall, all subjects performed above the chance level (i.e., 50%). A significant main effect of the group was found ( $p < 0.001$ ,  $H(2,53) = 29.76$ ). Young subjects demonstrated significantly higher total correct response rate at the congruency judgment than middle-aged ( $p = 0.002$ ) and older subjects ( $p < 0.001$ ). Middle-aged participants also exhibited higher correct response rates compared to older subjects ( $p = 0.002$ ). Finally, the main effect of the group on RT was significant ( $p = 0.002$ ,  $H(2,53) = 12.20$ ). The young group showed shorter RTs than both the middle-aged ( $p = 0.013$ ) and older ( $p < 0.001$ ) groups. However, no significant difference was observed between the middle-aged and older groups ( $p = 0.257$ ).

In sum, our findings confirmed that participants from each group properly understood the task and likely engaged in emotion attribution. Additionally, we observed that elderly participants and middle-aged individuals encountered difficulties

compared to young participants regarding social inferences, as reflected by lower correct response rate and extended reaction times.

### Modulation of behavioral performances by accuracy and uncertainty

We then sought to characterize the relationships between social inference abilities, indirectly measured through congruency judgment performance realized in the later segment of the video (i.e., correct response rate and RTs), and the computational indices computed during affective mental state inferences (i.e., accuracy and uncertainty crossed with dominant and competing emotions) (Figure 3B). We expected correct response rate to be positively modulated by accuracy of dominant emotions and negatively modulated by accuracy of competing emotion as well as uncertainty. Moreover, we hypothesized that the accuracy of dominant and competing emotions would respectively negatively and positively modulate RTs during congruency judgment, as an increase in the accuracy of the competing emotion would slow down the identification of the dominant emotion. Regarding uncertainty, we hypothesized that an increase in

uncertainty would slow down emotional inference, thus positively impacting RTs. A mixed-effects model using binomial distribution family with logit link function for correct response rate and a full linear mixed-effects model for RT, were implemented. Both models included the four computational indices, their interactions with age group, and participant-level random intercepts and slopes.

The model predicting correct response rates revealed that neither accuracy (dominant:  $\beta = 0.18$ ,  $p = 0.60$ ; competing:  $\beta = 0.07$ ,  $p = 0.89$ ) nor uncertainty of competing emotions ( $\beta = 0.16$ ,  $p = 0.82$ ) significantly influenced correct response rate. Only the uncertainty of the dominant emotion significantly reduced correct response rate ( $\beta = -1.54$ ,  $p < 0.001$ ). No age effects were highlighted. The model predicting RTs did not reveal any significant effect of accuracy of dominant emotion ( $\beta = 0.063$ ,  $p = 0.19$ ). By contrast, higher accuracy of competing emotions was associated with slower responses ( $\beta = 0.45$ ,  $p < 0.001$ ). Greater uncertainty of the dominant emotion was associated with slowed RTs ( $\beta = 0.36$ ,  $p < 0.001$ ), whereas greater uncertainty surrounding competing emotions showed the opposite effect ( $\beta = -0.55$ ,  $p < 0.001$ ). Age groups differed in the magnitude of these effects: compared to younger group, older adults showed a lower effect of uncertainty of dominant emotions on slowing of RTs ( $\beta = 0.30$ ,  $p = 0.009$ ). A less pronounced effect on RT when competing emotions were highly uncertain was also observed for the older compared to young group ( $\beta = -0.46$ ,  $p = 0.025$ ).

Overall, computational indices predicted behavior during social inference. The uncertainty of dominant emotions selectively reduced correct response rate, an effect not influenced by aging. RTs reflected a complex interplay between those computations, suggesting that competing emotions interfered less with emotional inference when they were themselves less accurate and highly uncertain. Moreover, behavior of older adults was less sensitive to emotional uncertainty than young adults. Taken together, these findings suggest that socially shared representations of emotional states, and their corresponding probability distribution, influenced later social inference, especially decision time.

### Brain activity associated with emotional inference

We first sought to determine the effect of emotion during inference on brain activity. We designed a first-level fMRI model including regressors encoding the temporal distribution of inference for both emotional and neutral video contexts, without further modulation by accuracy or uncertainty. Analyses were first restricted to the mPFC, using a mask based on a meta-analysis of the mindreading network that encompasses the dorsal and ventral mPFC<sup>49</sup> to perform small volume correction (SVC). The contrast between emotional and neutral videos, irrespective of age ( $N = 56$ ), revealed a large cluster in the mPFC (dorsal: peak coordinates [6 56 23],  $z < 8.00$ ,  $p_{\text{FWE-SVC}} < 0.001$ ; ventral: peak coordinates [-6 56 -4],  $z = 5.03$ ,  $p_{\text{FWE-SVC}} < 0.001$ ; Figure 4A). An exploratory whole-brain analysis was conducted, revealing a network associated with emotional inference, including the mPFC, anterior cingulate cortex, the bilateral temporal parietal junction, the bilateral middle temporal gyrus (MTG), and the amygdalo-hippocampal complex (whole-brain  $p_{\text{FWE}} < 0.05$ ,

see Table S1; Figure S1A). We then contrasted young and older adults using two-sample  $t$  test (combining middle-aged and older groups to enhance the statistical power,  $N = 36$ ). This analysis revealed that, compared with younger, older adults showed significantly reduced effects of emotion during inference on activity in the mPFC (Figure 4A; dorsal: peak coordinates [-3 62 11],  $z = 3.82$ ,  $p_{\text{FWE-SVC}} = 0.018$ ; ventral: peak coordinates [3 47 -7],  $z = 3.69$ ,  $p_{\text{FWE-SVC}} = 0.027$ ). Additional whole-brain analysis revealed a similar effect in the hippocampus (Figure S1A; peak coordinates [27 -13 -16],  $z = 4.73$ ,  $p_{\text{FWE}} = 0.015$ ).

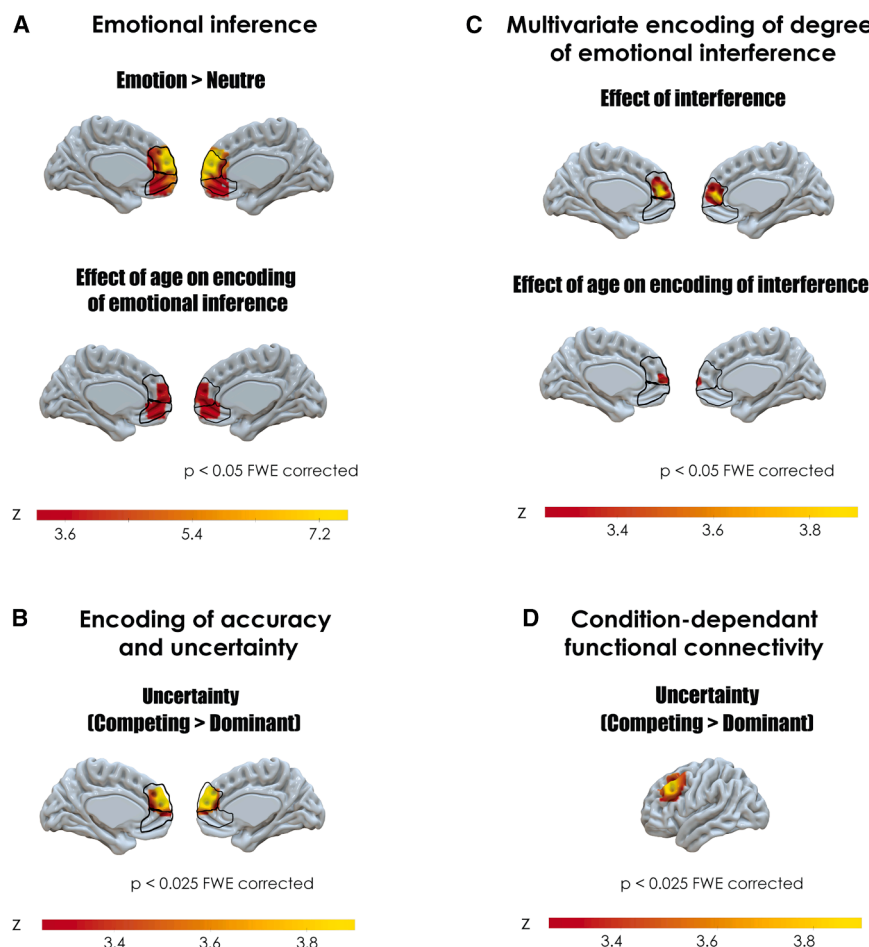
### Brain activity associated with accuracy and uncertainty

Having established that activity in the mPFC increased during the temporal window in which emotional inference is maximal, we then sought to determine how estimates of accuracy and uncertainty may further modulate brain activity. We designed a first-level fMRI model including regressors encoding the temporal distribution of inference for emotional contexts together with regressors parametrically modulated by the accuracy and uncertainty of both dominant and competing emotions.

To examine the role of the mPFC in distinguishing between alternative emotions, the main effect of the type of emotion was first investigated (i.e., across both groups). This analysis involved contrasting dominant and competing emotions for both computational indices. For accuracy, the mPFC did not show any distinct response between dominant and competing emotions. The dorsal mPFC (dmPFC) demonstrated higher activations for the uncertainty of competing emotions compared to the uncertainty of dominant emotions (Figure 4B; Left: peak coordinates [3 59 23],  $z = 4.84$ ,  $p_{\text{FWE-SVC}} = 0.005$ ; Right: peak coordinates [-3 59 23],  $z = 4.55$ ,  $p_{\text{FWE-SVC}} = 0.016$ ). These clusters also survived correction for multiple comparisons across the whole brain. We also tested if the mPFC responded to the main effect of accuracy and uncertainty, irrespectively of emotional content (i.e., dominant or competing emotions), but no significant effect was observed. Moreover, the effect of emotional competition that we observed for uncertainty did not interact with age.

As an exploratory analysis, the effects of interest were further investigated at the whole brain level (Table S1; Figure S1B). Analyzing brain regions outside the mPFC revealed increased activity bilaterally in the MTG and the left inferior frontal gyrus pars opercularis in response to the accuracy of dominant emotion compared with the accuracy of competing emotions. The opposite contrast (competing > dominant) did not highlight significant clusters. Uncertainty of dominant vs. competing emotions was associated with activations in visual regions. Finally uncertainty of competing vs. dominant emotion was associated with activity in brain regions of the mentalizing network, namely the right temporal pole, right temporoparietal junction, and dmPFC.

To summarize, our findings demonstrated that activity in the dmPFC, as well as in other regions central to mentalizing process, increased when emotional competition is socially uncertain. Conversely, the accuracy of emotions did not influence activity in the mPFC but rather regions related to social knowledge and simulation of other's mental states.



**Figure 4. Brain activity during emotional inference (N=56)**

(A) Encoding of emotional inference in the mPFC.

Brain regions within the mask of interest demonstrating a significant effect of the inference of emotional state compared to neutral (top) and effect of group on this encoding of emotional inference (bottom). mPFC was activated during the inference of emotional states (vs. neutral), and this effect decreased with age. For illustration purpose, activation maps are shown at a threshold of  $p < 0.001$ , but only clusters that survived  $p < 0.05$ , corrected for multiple comparisons across the mask of interest are shown.

(B) Univariate encoding of accuracy and uncertainty of dominant and competing emotions. Activity in the mPFC significantly increased with the uncertainty of competing emotions compared to dominant emotions. For illustration purpose, activation maps are shown at a threshold of  $p < 0.001$ , but only clusters that survived  $p < 0.025$ , corrected for multiple comparisons across the mask of interest are shown.

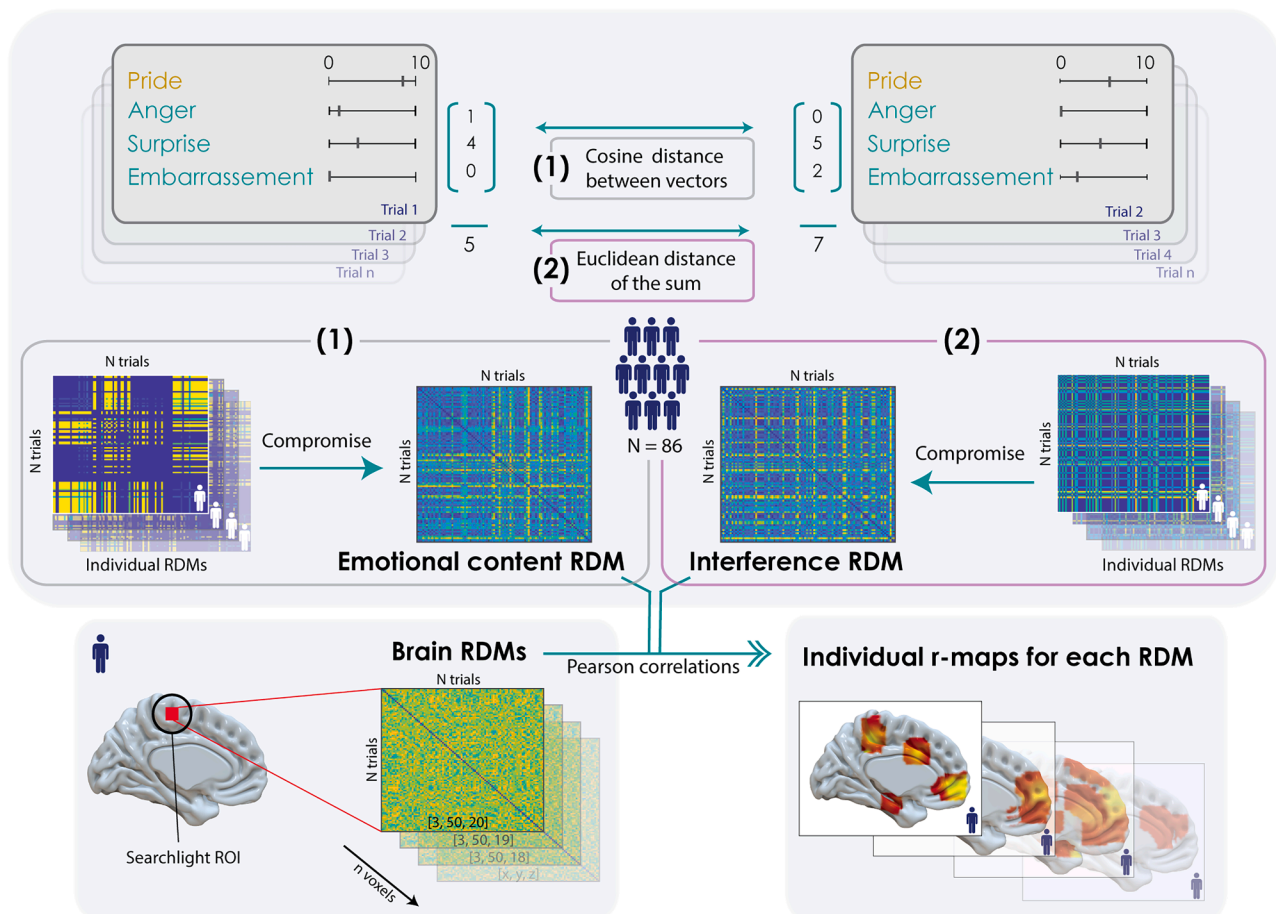
(C) Multivariate encoding of the degree of emotional interference due to competing emotions. Brain regions within the mask of interest demonstrating a significant effect of the interference (top) and an effect of age on this spatial encoding (bottom). Interference was spatially encoded in the mPFC, an effect that was affected by age (20 young and 36 older participants). For illustration purpose, activation maps are shown at a threshold of  $p < 0.001$ , but only clusters that survived  $p < 0.05$ , corrected for multiple comparisons across the mask of interest are shown.

(D) Condition-dependent functional connectivity of the mPFC. vmPFC showed increased functional coupling with the left dorsolateral prefrontal cortex for uncertainty of competing vs. dominant emotions. For illustration purpose, activation maps are shown at a threshold of  $p < 0.001$ , but only clusters that survived  $p < 0.025$ , corrected for multiple comparisons across the mask of interest are shown. mPFC: medial prefrontal cortex.

### RSA of the competition between emotional states

Our findings showed that the effect of emotional competition on the activity of the dmPFC is driven by the social uncertainty of inferred emotions. However, these findings do not allow to disentangle which aspect of competing emotional inference is represented in the mPFC and gives rise to this uncertainty signal. In particular, emotional inferences may differ either in the specific emotional alternatives that compete (e.g., anger vs. embarrassment) or in the overall strength of emotional interference,

reflected by the total probability of competing emotions, independently of which emotions are involved. To test those two hypotheses, we conducted RSA<sup>47,48</sup> (see Figure 5 for the methodology). RSA allows testing these hypotheses about the structure of neural representations by comparing neural similarity patterns to model representational dissimilarity matrices (RDMs) derived from behavioral data. Importantly, these analyses do not test whether mPFC represents accuracy or uncertainty, which were examined in previous analyses. Instead,



**Figure 5. Methodology of RSA**

First, model RDMs were computed.

(1) Emotional content RDM represented the pairwise relationships between trials based on the content of the emotional competition.

(2) Interference RDM represented the pairwise relationship between trials based on the overall probabilities of competing emotions.

For each participant in the pre-experimental study ( $N = 86$ ), individual model RDMs were generated. First individual RDMs (1) were generated by computing the cosine distance between every trial pair using the vectors of intensities of competing emotions for emotional content. Second, individual RDMs (2) were generated by computing the Euclidean distance between every trial pair in terms of the sum of intensities of competing emotions. Then, the shared structure of representations across those individuals was captured by seeking optimal agreement or compromise across individual RDMs of the same dimension, with the DISTATIS method (Abdi et al., 2012), leading to the (1) emotional content RDM and (2) interference RDM. The brain RDMs were also computed. For each participant in the fMRI study ( $N = 56$ ), a searchlight analysis was carried-out. More precisely, a spherical ROI with a radius of 8 mm was moved through voxels of the individual beta maps. For each voxel, a brain RDM was computed by comparing neural activity across voxels encompassed within the searchlight ROI for every trial pair. Finally, the correspondences between the brain RDM, centered on a specific voxel, and each of the two model RDMs were specified using Pearson correlations. At this step, variance created by potential confounds (timing of inference and intensity of the dominant emotion) were regressed-out. The resulting correlation values between the model RDMs and the similarity of brain activity (brain RDM) in the area of the sphere were stored in individual r-maps; leading in two r-maps per participants.

they test what component of emotional competition—emotional content or the magnitude of interference between interpretations—drives the neural encoding of social uncertainty.

Two hypotheses were tested: (1) the mPFC represents the specific emotional alternatives that compete when individuals interpret a social situation. Under this account, neural similarity between situations should reflect the similarity of their competing (i.e., non-dominant) emotional profiles across the population; (2) the mPFC represents the overall magnitude of interference generated by the amount of probability summed across competing emotions, independently of their specific content.

Under this hypothesis, neural similarity should depend on the overall amount of competition, such that situations producing comparable levels of emotional interference evoke similar neural patterns regardless of which emotions are involved.

To operationalize these hypotheses, we constructed two model RDMs based on behavioral data from the pre-experimental study. Each RDM captured a different aspect of emotional competition across social situations. The emotional content RDM captured differences between social situations according to the content of their competing (i.e., non-dominant) emotions (Figure 5). For each participant, dissimilarity between

pairs of situations was computed as the cosine distance between the probability vectors of competing emotions. Under this model, situations were considered more similar when they involved similar profiles of competing emotional inferences. The interference RDM captured differences between situations according to the overall magnitude of emotional competition. For each participant, interference was defined as the summed probability of the competing emotions for a given situation. Dissimilarity between situations was then computed as the Euclidean distance between these interference values. Under this model, situations were considered more similar when they produced comparable levels of emotional interference, irrespective of which specific emotions were involved. Individual RDMs were then combined across participants using the DISTATIS compromise procedure,<sup>28,50</sup> which extracts the similarity structure shared across individuals (i.e., the social consensus). The resulting group-level RDMs therefore capture the socially shared organization of emotional competition across situations. To control for potential confounds, two additional RDMs capturing the timing of inference and the probability of the dominant emotion were included as nuisance regressors in the analysis. We then performed a searchlight RSA: for each voxel, a local brain RDM was computed from the similarity of neural activity patterns across trials and correlated with the model RDMs, producing individual correlation maps (*r*-maps). Group-level analyses were performed using one-sample *t* tests to assess the main effects of emotional content and interference, and two-sample *t* tests to examine age-group differences. Consistent with the univariate analyses, middle-aged and older participants were combined into a single older group for these comparisons.

The results from the multivariate RSA are summed in Figure 4C. The mPFC, mostly in its dorsal part, exhibited a significant encoding of the interference caused by competing emotions when inferring emotional states. In contrast, the mPFC did not represent the content of emotional competition. Group comparisons revealed a decrease of encoding of the interference due to competition in the dmPFC, in old compared to young group. These effects were robust, as sensitivity analyses using searchlight radii of 6 mm and 10 mm yielded convergent results (data not shown).

Following these findings, we examined whether individual differences in executive function could account for the observed group effect in dmPFC. Executive functions variables included inhibitory control (stroop interference time), cognitive flexibility (trail making test B–A time), verbal fluency (mean of animal, P, and V), and working memory (forward and backward visuospatial span). Correlational analyses between these variables and the RSA rho estimates capturing the encoding of interference in dmPFC showed no significant relationships across all participants (all  $\rho < 0.25$  and  $p_{\text{FDR}} > 0.307$ ).

Finally, we explored whole brain encoding of emotional content and interference (Table S1; Figure S1C). Emotional content was spatially encoded in the left MTG and the left anterior supramarginal gyrus. Degree of interference showed no significant cluster.

To summarize, these findings demonstrated that the dmPFC represents the emotional competition by spatially encoding the social consensus regarding emotional interference (or conflict)

caused by the overall probability of competing emotions. Within this region, pairs of social situations with similar level of emotional interference had increased similarity of neural activity compared with pairs of distinct degree of interference. This finding suggests that uncertainty of inferred mental states is driven by the social consensus on the magnitude of interference caused by the overall probability of emotional hypotheses about unfolding social events in the mPFC. It also suggests that this computation is altered in aging, independently of executive function decline.

### Functional connectivity of the mPFC during inference of emotions

To further understand the computations taking place in the mPFC, we conducted additional psychophysiological interaction (PPI) analyses to investigate how the mPFC functionally interact with other brain regions during emotional inference. Seed regions for the ventral mPFC (vmPFC) and dmPFC were the same as the masks used in previous analyses. PPI regressors were computed using the main general linear model (GLM), which included accuracy and uncertainty for both dominant and competing emotions. All analyses were thresholded at  $p < 0.05$  (corrected for family-wise error [FWE],  $k = 10$ ). One-sample *t* tests assessed condition effects on functional connectivity for each seed, and two-sample *t* tests tested for group differences.

No suprathreshold clusters emerged for the dmPFC for any condition. For the vmPFC (Figure 4D), the comparison between uncertainty for competing versus dominant emotions revealed increased vmPFC coupling with the left dorsolateral prefrontal cortex (dlPFC; peak coordinates  $[-45, 32, 38]$ ,  $z = 4.48$ ,  $p_{\text{FWE}} = 0.019$ ). No other condition effects on vmPFC connectivity reached significance. For both vmPFC and dmPFC, no group effects were highlighted.

In sum, uncertainty related to competing emotions was associated with stronger functional connectivity between the vmPFC and the left dlPFC—a region typically involved in interference resolution.<sup>51,52</sup> This pattern aligns with our RSA findings showing that the mPFC encodes the degree of interference caused by emotional competition. Hence, when emotional competition induces uncertainty, the vmPFC likely enhances its coupling with dlPFC, supporting the resolution of interference caused by competing affective signals.

### Relationships between behavioral performances and brain estimates of accuracy and uncertainty in the mPFC

Finally, we sought to highlight the relationships between brain estimates within the regions of interest (ROIs) corresponding to ventral and dorsal mPFC and behavioral performances, across all participants. Spearman's correlations were applied between the correct response rate at the congruency judgment and (1) the estimates of accuracy and uncertainty in the univariate analyses, and (2) the correlation coefficients extracted from the RSA with the interference model.

Regarding the univariate analysis, correct response rate showed significant correlations with brain estimates of uncertainty of dominant emotion in the dorsal and ventral mPFC (vmPFC:  $\rho = -0.28$   $p_{\text{FDR}} = 0.035$ ; dmPFC:  $\rho = -0.32$   $p_{\text{FDR}} = 0.033$ ). Brain estimates

of the other computational indices did not show any significant relationship with behavioral performances. Additionally, rho coefficients extracted from the interference RSA exhibited a significant relationship with correct response rate in the dmPFC ( $\rho = 0.31$ ,  $p_{FDR} = 0.040$ ), while such correlation was not observed in the vmPFC ( $\rho = 0.20$ ,  $p_{FDR} = 0.13$ ).

In sum, the encoding of the social consensus on emotional interference in the dmPFC was related to behavioral performances associated with mindreading.

## DISCUSSION

This study aimed to investigate how accuracy and uncertainty of inferred emotional mental states linked with socially shared representations influence the activity of the mPFC, and how the encoding of these statistical regularities in the mPFC may change with aging. Actors were filmed experiencing social situations that elicited different emotions in an ecologically valid yet highly controlled setting. Pre-experimental data ( $N = 86$ ) enabled to determine the expected probabilities of both dominant and competing emotions during emotional state inference. The mean and standard deviation of these pre-experimental ratings were employed to estimate the accuracy and uncertainty of emotional inference, respectively. These estimates served as parametric modulators in the affective mental state attribution task, during which brain activity was recorded using fMRI in young ( $N = 20$ ) and older adults ( $N = 36$ ). Findings revealed that behavior during the task was particularly associated with social uncertainty related to emotional inference. This behavioral sensitivity to uncertainty was affected by age, with older adults showing a reduced influence of emotional uncertainty during inferences. At the neural level, activity in the dmPFC was not modulated by the probability of competing emotions (i.e., accuracy), but by the variability of possible emotional outcomes, a factor that makes emotional inference more uncertain from a social perspective. Representational similarity analyses further showed that the effect of emotional competition in this same region was not driven by emotional content itself but instead by the social consensus about the strength of emotional competition (i.e., interference), an effect that was attenuated in older compared to younger adults. Functional connectivity analyses revealed a coupling between the vmPFC and the left dlPFC that was increased when multiple competing emotions were uncertain, thereby suggesting that social emotional uncertainty increased interaction between regions supporting interference resolution. Finally, neural activity associated with emotional interference in the dmPFC predicted subsequent social behavior, highlighting the functional relevance of these neural computations for mindreading.

Our findings revealed that within the mentalizing network—engaged during emotional inference—socially shared representations about the emotional states of others were related to mPFC activity in two ways: (1) the social variability (i.e., uncertainty) of alternative emotional hypotheses enhanced activity of the dmPFC, and (2) the social consensus regarding the emotional conflict elicited by the actors' interaction (i.e., characterized by an increased overall emotional competition) enhanced pattern similarity within the dmPFC. In our study, uncertainty of

emotional inference is operationalized as the variance of shared representations across individuals. Those socially shared representations, presumably acquired through experiencing others' emotions during social interactions, would guide emotional hypotheses established at the individual level. Consistent with this view, the mPFC is involved in the representation of shared social experiences, including shared beliefs,<sup>25</sup> norms,<sup>26</sup> as well as cultural<sup>27</sup> and collective knowledge.<sup>28</sup> Furthermore, the integration of information related to past experiences from the mPFC<sup>53</sup> plays a role in attributing an affective value to future options.<sup>54</sup> This integrative function of the mPFC could therefore also support the integration of competing and alternative hypotheses learned during social interactions, which have a potential value in explaining others' emotions.

Interestingly, this interpretation aligns with the central role of the mPFC within the brain valuation system (BVS). The BVS encodes the value of alternative options together with confidence about these estimates.<sup>37,55,56</sup> Beyond economic choice, computational models indicate that social value also guides behavior during social inference.<sup>16,31,32</sup> Within this framework, inferring others' emotions can be conceived as relying on the representation of competing learned perspectives derived from prior social experiences, which are then used to interpret others' emotional states. Emotional perspective learned through social interactions would thus be assigned a goal-value during inference, reflecting how well it may explain the current social context. The emotion with the highest value would be selected as the most probable<sup>37,57</sup>; however, the coexistence of multiple plausible emotional perspectives would increase uncertainty and, in turn, mPFC activity. This interpretation would also explain why the mPFC activity patterns are sensitive to the degree of interference rather than to the specific competing emotional content, as valuation signals encode the relative value of options independently of their semantic labels. Conversely, the accuracy of dominant emotion was bilaterally encoded in the MTG, a region implicated in social knowledge retrieval,<sup>58</sup> and RSA further revealed that the content of emotional competition was also represented in this region. This finding likely reflects the fact that information about the alternative emotional contents might be represented within a distributed semantic network, while the mPFC integrates the social value of competing emotional representations during the inferential process.

However, although the parallel with the BVS is appealing as it provides a parsimonious account of the mPFC, it must be emphasized that emotional uncertainty as computed here is different from the confidence signal in value-based choice.<sup>37,59,60</sup> Our methodological approach specifically isolated socially based uncertainty rather than individual confidence. Uncertainty indices were derived from a population independent of the scanned participants, capturing population-level variance rather than metacognitive judgments. As such, uncertainty reflects an implicit, Bayesian-like estimate grounded in socially shared priors.<sup>23,38</sup> In the brain valuation system, subjective confidence has been consistently shown to positively correlate with mPFC activity.<sup>37,56</sup> If our uncertainty index merely reflected the inverse of confidence, one would therefore expect a negative association with mPFC signals. Instead, we observed a positive encoding of uncertainty in mPFC. This result aligns with Jenkins and

Mitchell<sup>39</sup> work, showing greater mPFC activation when inferring mental states from ambiguous social scenarios—where the scenario itself prevented a determinate inference—compared to non-ambiguous scenarios. Moreover, the absence of explicit confidence ratings in our task further supports the interpretation that uncertainty is computed automatically, even when behavior does not require explicit consideration of alternative emotional states. Together, these findings extend the role of mPFC beyond metacognitive confidence, showing that it encodes probabilistic uncertainty of competing emotions in the context of complex social inference.

The mPFC activity associated with uncertainty of competing emotions may reflect multiple underlying computational processes. One possibility is that heightened mPFC responses under uncertainty reflect prediction errors generated during such inferences.<sup>61</sup> This is consistent with evidence that the mPFC supports automatic prediction and updating of social expectations.<sup>8,32,61,62</sup> Within predictive coding frameworks, uncertainty and prediction error are tightly coupled, such that less precise priors increase prediction error and drive model updating.<sup>63</sup> In this view, mPFC activation might signal the need for model revision rather than uncertainty per se. Another possibility is that the activity of the mPFC reflects the confounding effect of the effort invested when inferring emotions, instead of uncertainty. This assumption, however, seems highly unlikely for several reasons. First, the regressors we defined reflect the temporal window when the maximum of information for inference is available, presumably requiring minimal effort. Second, activity in the mPFC is parametrically modulated by uncertainty, which is defined by socially shared representations independently of the effort engaged to reach such inference. Third, the brain region associated with the representation of effort mobilization (i.e., dorsal anterior cingulate cortex) is more dorsal.<sup>64</sup> Finally, our results more strongly support a third account in which mPFC activity reflects the conflict generated by competing emotional representations. In our data, RSA revealed that mPFC patterns track similarity as a function of the strength of emotional competition that we interpret as an overall emotional interference. This interpretation aligns with models describing the mPFC as monitoring and resolving affective conflict.<sup>65</sup> This view is further supported by our observation of increased functional coupling between the mPFC and dlPFC under uncertainty related to competition. Given the established role of the dlPFC in cognitive control and interference resolution,<sup>51,52</sup> these findings suggest that mPFC–lateral prefrontal interactions support the resolution of uncertainty arising from the interference due to competing emotions. Our RSA analysis did not, however, confirm that the vmPFC encodes competing emotional content, as previously suggested by studies at the neuronal level.<sup>17–19</sup> This discrepancy may indicate that the population code associated with emotional content in the vmPFC is too weak to be reliably detected in the fMRI signal relative to noise. In contrast, we found that the magnitude of emotional competition, independent of the specific emotional content, was represented in the dmPFC. This suggests that the dmPFC monitors competition processes during emotional inference.

Our findings revealed a reduced engagement of the mPFC during emotional inference in older compared to younger adults,

in line with converging evidence from previous neuroimaging studies.<sup>41–43</sup> Beyond mPFC alterations, we also observed a significant effect of age on hippocampal function during emotional inference. The hippocampus plays a central role in the mental simulation of future events,<sup>66,67</sup> a process that is highly relevant to the present task, as participants were required to anticipate others' future emotional states based on a social context. Importantly, hippocampal integrity and function are known to decline with aging,<sup>68</sup> suggesting that age-related hippocampal changes may constrain the simulation of future emotional events and thereby contribute to altered emotional inference in older adults. Crucially, our results further indicate age-related alterations in the mPFC encoding of socially shared interference arising from competing emotional hypotheses, alongside a reduced behavioral sensitivity to uncertainty in older adults. These effects were not explained by individual differences in executive functioning. Instead, interference-related activity in the dmPFC was selectively associated with mindreading performance, suggesting a mechanism more specifically tied to social-cognitive processing. These findings were consistent with previous work suggesting that older adults underestimate uncertainty<sup>69,70</sup> and rely more strongly on prior knowledge than exploration of new uncertain information during decision-making, a shift linked to age-related changes in mPFC function.<sup>13,69</sup> Together, these results support the view that age-related changes in social inference arise from altered mental simulation of future emotional states and strategies changes when facing uncertain non-prototypical social situations. These changes could lead older adults to down-weight alternative emotional hypotheses and thereby underestimate social uncertainty.

### Limitations of the study

This study has several limitations and opens up new research perspectives and questions. Our task captured mindreading through dynamic social situations resembling real-life interactions. This design necessarily introduced ambiguity and engaged multiple cognitive processes. While this allowed us to characterize both accuracy and uncertainty and to identify the role of the mPFC in encoding emotional interference, the behavioral evidence linking emotional inference—through congruency judgments—to prediction and uncertainty remained indirect. Second, although our sample size was sufficient to support the main uncertainty-related effects independent of age ( $N = 56$ ), it remains modest and may have limited sensitivity to detect more subtle or fine-grained age-related differences. Future studies with larger and more age-diverse samples, together with appropriate MRI sequences to disentangle age-related vascular effects,<sup>71</sup> will be required to better understand the changes of social cognition during aging. Moreover, directly linking neural markers of uncertainty to both the specific content of inferred mental states and subjective confidence ratings would help disentangle neural representations of uncertainty from metacognitive confidence and clarify their respective contributions across the lifespan. Finally, social schemas and emotional concepts vary across cultures and shape how emotions are inferred from context.<sup>72,73</sup> While intercultural data collection was beyond the scope of the current study, future work should test the generalizability of these findings in culturally diverse populations.

## RESOURCE AVAILABILITY

### Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Mickaël Laisney ([mickael.laisney@ephe.psl.eu](mailto:mickael.laisney@ephe.psl.eu)).

Postal address: U1077 Neuropsychologie et Imagerie de la Mémoire Humaine (NIMH), Pôle des Formations et de Recherche en Santé, 2 Rue des Rochambelles, F-14032 CAEN Cedex CS 14032, France.

### Materials availability

The affective mental state inference task material is available from the corresponding author on request. The MATLAB implementation based on SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>) of the univariate, psychophysiological interaction, and representational similarity analyses is publicly available on GitHub, with an archived version accessible via Zenodo (DOI: <https://doi.org/10.5281/zenodo.20036054>).

### Data and code availability

- All raw behavioral and imaging data are archived at the GIP Cyceron Center in Caen. The raw data are not publicly available due to legal restrictions imposed by the sponsor of this study but are available from the corresponding authors upon request.
- All data reported in this paper will be shared by the lead contact upon request.
- The MATLAB implementation based on SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>) of the univariate, psychophysiological interaction and representational similarity analyses is publicly available on GitHub, with an archived version accessible via Zenodo (DOI: <https://doi.org/10.5281/zenodo.20036054>).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## ACKNOWLEDGMENTS

This paper is dedicated to the memory of our colleague Beatrice Desgranges who made substantial contributions to the conception and design of the task used in this study and to the acquisition of data. The French Ministry of Health supported this research (PHRC, ID-RCB: 2011-A00681-40). Dr. Bejanin acknowledges support from Instituto de Salud Carlos III and co-funded by the European Union through the Miguel Servet grant (CP20/00038) and Fondo de Investigaciones Sanitarias (PI22/00307), the Alzheimer's Association (AARG-22-923680) and the Ajuntament de Barcelona, in collaboration with Fundació La Caixa (23S06157-001). The authors would like to thank the participants of this study as well as Marie Caillaud, Harmony Duclos, Rémi Laillier, Noémie Guesney, and Charlotte Toso for their involvement in recruiting participants and collecting neuropsychological and MRI data. We are also grateful to Antoine Nourry, Nicolas Delcroix, Vincent Mouton, Salhi Karim, Malo Gaubert, Armelle Viard, Patrice Clochon, Mikael Naveau, Shailendra Segobin, and Joy Perrier for their support to validate the pre-experiments, implement the MRI experiment and analyze the data. The authors are also grateful to the actors (Brice Carrois and Pauline Bertani) and moviemakers (Baptiste Trolong-Bailly and Victor Royer) of the mental state attribution task. We thank the municipality of Chedigny and its mayor Pierre Louault for allowing us to film the scene in the municipal hall, as well as Isabelle Bejanin for her precious assistance before and during the shooting.

## AUTHOR CONTRIBUTIONS

Conceptualization, M.L.P., P.G., F.E., and M.L.; methodology, M.L.P., P.G., and M.L.; formal analysis, M.L.P. and P.G.; investigation, A.B. and V.d.L.S.; writing—original draft, M.L.P. and P.G.; writing—review and editing, P.G., A.B., F.E., and M.L.; visualization, M.L.P. and P.G.; supervision, F.E. and M.L.; funding acquisition, F.E. and M.L.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) used ChatGPT in order to improve the readability and language of the manuscript. After using this tool or service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
- **METHOD DETAILS**
  - Participants
  - Procedure
  - Statistical analysis of behavioral performances
  - MRI acquisition parameters
  - Preprocessing
  - First-level univariate analysis
  - Univariate brain analysis
  - Representational similarity analysis
  - Psychophysiological interaction analysis
  - Relationships between behavioral performances and brain estimates
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplementary data related to this article can be found online at <https://doi.org/10.1016/j.celrep.2026.117506>.

Received: June 25, 2025

Revised: December 19, 2025

Accepted: May 15, 2026

Published: June 4, 2026

## REFERENCES

1. Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. <https://doi.org/10.1017/S0140525X00076512>.
2. Apperly, I.A. (2010). *Mindreaders: The Cognitive Basis of "Theory of Mind"* (Psychology Press). <https://doi.org/10.4324/9780203833926>.
3. Duclos, H., Desgranges, B., Eustache, F., and Laisney, M. (2018). Impairment of social cognition in neurological diseases. *Rev. Neurol. (Paris)* 174, 190–198. <https://doi.org/10.1016/j.neurol.2018.03.003>.
4. Shi, Z., and Burr, D. (2016). Predictive coding of multisensory timing. *Curr. Opin. Behav. Sci.* 8, 200–206. <https://doi.org/10.1016/j.cobeha.2016.02.014>.
5. Doya, K., Ishii, S., Pouget, A., and Rao, R.P.N. (2006). *Bayesian Brain: Probabilistic Approaches to Neural Coding* (The MIT Press). <https://doi.org/10.7551/MITPRESS/9780262042383.001.0001>.
6. Lin, C.H.S., and Garrido, M.I. (2022). Towards a cross-level understanding of Bayesian inference in the brain. *Neurosci. Biobehav. Rev.* 137, 104649. <https://doi.org/10.1016/j.neubiorev.2022.104649>.
7. Diaconescu, A.O., Mathys, C., Weber, L.A.E., Daunizeau, J., Kasper, L., Lomakina, E.I., Fehr, E., and Stephan, K.E. (2014). Inferring on the Intentions of Others by Hierarchical Bayesian Learning. *PLoS Comput. Biol.* 10, e1003810. <https://doi.org/10.1371/JOURNAL.PCBI.1003810>.
8. Thornton, M.A., Weaverdyck, M.E., and Tamir, D.I. (2019). The Social Brain Automatically Predicts Others' Future Mental States. *J. Neurosci.* 39, 140–148. <https://doi.org/10.1523/JNEUROSCI.1431-18.2018>.

9. FeldmanHall, O., and Shenhav, A. (2019). Resolving uncertainty in a social world. *Nat. Hum. Behav.* 3, 426–435. <https://doi.org/10.1038/s41562-019-0590-x>.
10. Alexander, W.H., and Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–1344. <https://doi.org/10.1038/nn.2921>.
11. Rouault, M., Lebreton, M., and Pessiglione, M. (2023). A shared brain system forming confidence judgment across cognitive domains. *Cereb. cortex* 33, 1426–1439. <https://doi.org/10.1093/CERCOR/BHAC146>.
12. Trudel, N., Scholl, J., Klein-Flügge, M.C., Fouragnan, E., Tankelevitch, L., Wittmann, M.K., and Rushworth, M.F.S. (2021). Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. *Nat. Hum. Behav.* 5, 83–98. <https://doi.org/10.1038/s41562-020-0929-3>.
13. Spreng, R.N., and Turner, G.R. (2021). From exploration to exploitation: a shifting mental mode in late life development. *Trends Cogn. Sci.* 25, 1058–1071. <https://doi.org/10.1016/j.tics.2021.09.001>.
14. Brand, M., Labudda, K., and Markowitsch, H.J. (2006). Neuropsychological correlates of decision-making in ambiguous and risky situations. *Neural Netw.* 19, 1266–1276. <https://doi.org/10.1016/j.neunet.2006.03.001>.
15. Stern, E.R., Gonzalez, R., Welsh, R.C., and Taylor, S.F. (2010). Updating beliefs for a decision: Neural correlates of uncertainty and underconfidence. *J. Neurosci.* 30, 8032–8041. <https://doi.org/10.1523/JNEUROSCI.4729-09.2010>.
16. De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., and Love, B.C. (2017). Social Information Is Integrated into Value and Confidence Judgments According to Its Reliability. *J. Neurosci.* 37, 6066–6074. <https://doi.org/10.1523/JNEUROSCI.3880-16.2017>.
17. Hunt, L.T., Kolling, N., Soltani, A., Woolrich, M.W., Rushworth, M.F.S., and Behrens, T.E.J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nat. Neurosci.* 15, 470–476. <https://doi.org/10.1038/nn.3017>.
18. Hämmerer, D., Bonaiuto, J., Klein-Flügge, M., Bikson, M., and Bestmann, S. (2016). Selective alteration of human value decisions with medial frontal tDCS is predicted by changes in attractor dynamics. *Sci. Rep.* 6, 25160. <https://doi.org/10.1038/SREP25160>.
19. Strait, C.E., Blanchard, T.C., and Hayden, B.Y. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron* 82, 1357–1366. <https://doi.org/10.1016/J.NEURON.2014.04.032>.
20. Thornton, M.A., and Tamir, D.I. (2017). Mental models accurately predict emotion transitions. *Proc. Natl. Acad. Sci. USA* 114, 5982–5987. <https://doi.org/10.1073/PNAS.1616056114>.
21. Grossmann, I., Huynh, A.C., and Ellsworth, P.C. (2016). Emotional complexity: Clarifying definitions and cultural correlates. *J. Pers. Soc. Psychol.* 111, 895–916. <https://doi.org/10.1037/PSP0000084>.
22. Quidbach, J., Gruber, J., Mikolajczak, M., Kogan, A., Kotsou, I., and Norton, M.I. (2014). Emodiversity and the emotional ecosystem. *J. Exp. Psychol. Gen.* 143, 2057–2066. <https://doi.org/10.1037/A0038025>.
23. Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. <https://doi.org/10.1038/nrn2787>.
24. Conway, J.R., Catmur, C., and Bird, G. (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. *Psychon. Bull. Rev.* 26, 798–812. <https://doi.org/10.3758/S13423-018-1559-X>.
25. Ereira, S., Hauser, T.U., Moran, R., Story, G.W., Dolan, R.J., and Kurth-Nelson, Z. (2020). Social training reconfigures prediction errors to shape Self-Other boundaries. *Nat. Commun.* 11, 3030. <https://doi.org/10.1038/s41467-020-16856-8>.
26. Pegado, F., Hendriks, M.H.A., Amelynck, S., Daniels, N., Bulthé, J., Masson, H.L., Boets, B., and de Beeck, H.O. (2018). Neural Representations Behind “Social Norm” Inferences In Humans. *Sci. Rep.* 8, 12943. <https://doi.org/10.1038/S41598-018-31260-5>.
27. Broom, T.W., Iyer, S., Courtney, A.L., and Meyer, M.L. (2024). Loneliness corresponds with neural representations and language use that deviate from shared cultural perceptions. *Commun. Psychol.* 2, 40. <https://doi.org/10.1038/s44271-024-00088-3>.
28. Gagnepain, P., Vallée, T., Heiden, S., Decorde, M., Gauvain, J.L., Laurent, A., Klein-Peschanski, C., Viader, F., Peschanski, D., and Eustache, F. (2020). Collective memory shapes the organization of individual memories in the medial prefrontal cortex. *Nat. Hum. Behav.* 4, 189–200. <https://doi.org/10.1038/s41562-019-0779-z>.
29. Molenberghs, P., Johnson, H., Henry, J.D., and Mattingley, J.B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neurosci. Biobehav. Rev.* 65, 276–291. <https://doi.org/10.1016/j.neubiorev.2016.03.020>.
30. Van Overwalle, F., and Baetens, K. (2009). Understanding others’ actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage* 48, 564–584. <https://doi.org/10.1016/j.neuroimage.2009.06.009>.
31. Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., and Rushworth, M.F.S. (2008). Associative learning of social value. *Nature* 456, 245–249. <https://doi.org/10.1038/nature07538>.
32. Diaconescu, A.O., Mathys, C., Weber, L.A.E., Kasper, L., Mauer, J., and Stephan, K.E. (2017). Hierarchical prediction errors in midbrain and septum during social learning. *Soc. Cognit. Affect Neurosci.* 12, 618–634. <https://doi.org/10.1093/scan/nsw171>.
33. Reiter, A.M.F., Kanske, P., Eppinger, B., and Li, S.C. (2017). The Aging of the Social Mind-Differential Effects on Components of Social Understanding. *Sci. Rep.* 7, 11046. <https://doi.org/10.1038/s41598-017-10669-4>.
34. Zhang, L., and Gläscher, J. (2020). A brain network supporting social influences in human decision-making. *Sci. Adv.* 6, eabb4159. <https://doi.org/10.1126/sciadv.abb4159>.
35. Hampton, A.N., Bossaerts, P., and O’Doherty, J.P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. USA* 105, 6741–6746. <https://doi.org/10.1073/PNAS.0711099105>.
36. Yoshida, W., Seymour, B., Friston, K.J., and Dolan, R.J. (2010). Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* 30, 10744–10751. <https://doi.org/10.1523/JNEUROSCI.5895-09.2010>.
37. Lebreton, M., Abitbol, R., Daunizeau, J., and Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nat. Neurosci.* 18, 1159–1167. <https://doi.org/10.1038/nn.4064>.
38. Knill, D.C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. <https://doi.org/10.1016/J.TINS.2004.10.007>.
39. Jenkins, A.C., and Mitchell, J.P. (2010). Mentalizing under Uncertainty: Dissociated Neural Responses to Ambiguous and Unambiguous Mental State Inferences. *Cerebr. Cortex* 20, 404–410. <https://doi.org/10.1093/CERCOR/BHP109>.
40. Duval, C., Piolino, P., Bejanin, A., Eustache, F., and Desgranges, B. (2011). Age effects on different components of theory of mind. *Conscious. Cognit.* 20, 627–642. <https://doi.org/10.1016/j.concog.2010.10.025>.
41. Lailier, R., Viard, A., Caillaud, M., Duclos, H., Bejanin, A., de La Sayette, V., Eustache, F., Desgranges, B., and Laisney, M. (2019). Neurocognitive determinants of theory of mind across the adult lifespan. *Brain Cognit.* 136, 103588. <https://doi.org/10.1016/j.bandc.2019.103588>.
42. Moran, J.M. (2013). Lifespan development: The effects of typical ageing on theory of mind. *Behav. Brain Res.* 237, 32–40. <https://doi.org/10.1016/j.bbr.2012.09.020>.
43. Kemp, J., Després, O., Sellal, F., and Dufour, A. (2012). Theory of Mind in normal ageing and neurodegenerative pathologies. *Ageing Res. Rev.* 11, 199–219. <https://doi.org/10.1016/J.ARR.2011.12.001>.
44. Zamarian, L., Sinz, H., Bonatti, E., Gamboz, N., and Delazer, M. (2008). Normal Aging Affects Decisions Under Ambiguity, but Not Decisions Under Risk. *Neuropsychology* 22, 645–657. <https://doi.org/10.1037/0894-4105.22.5.645>.

45. Fein, G., McGillivray, S., and Finn, P. (2007). Older Adults Make Less Advantageous Decisions than Younger Adults: Cognitive and Psychological Correlates. *J. Int. Neuropsychol. Soc.* 13, 480–489. <https://doi.org/10.1017/S135561770707052X>.
46. Caillaud, M., Bejanin, A., Laisney, M., Gagnepain, P., Gaubert, M., Viard, A., Clochon, P., de La Sayette, V., Allain, P., Eustache, F., and Desgranges, B. (2020). Influence of emotional complexity on the neural substrates of affective theory of mind. *Hum. Brain Mapp.* 41, 139–149. <https://doi.org/10.1002/hbm.24794>.
47. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4. <https://doi.org/10.3389/NEURO.06.004.2008>.
48. Viviani, R. (2021). Overcoming bias in representational similarity analysis. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2102.08931>.
49. Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>.
50. Abdi, H., Williams, L.J., Valentin, D., and Bennani-Dosse, M. (2012). STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling. *WIREs Computational Stats.* 4, 124–167. <https://doi.org/10.1002/WICS.198>.
51. Nee, D.E., Wager, T.D., and Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognit. Affect. Behav. Neurosci.* 7, 1–17. <https://doi.org/10.3758/CABN.7.1.1/METRICS>.
52. Li, Q., Yang, G., Li, Z., Qi, Y., Cole, M.W., and Liu, X. (2017). Conflict detection and resolution rely on a combination of common and distinct cognitive control networks. *Neurosci. Biobehav. Rev.* 83, 123–131. <https://doi.org/10.1016/J.NEUBIOREV.2017.09.032>.
53. Preston, A.R., and Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Curr. Biol.* 23, R764–R773. <https://doi.org/10.1016/J.CUB.2013.05.041>.
54. Benoit, R.G., Szpunar, K.K., and Schacter, D.L. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proc. Natl. Acad. Sci. USA* 111, 16550–16555. <https://doi.org/10.1073/pnas.1419274111>.
55. Bartra, O., McGuire, J.T., and Kable, J.W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76, 412–427. <https://doi.org/10.1016/j.neuroimage.2013.02.063>.
56. Clairis, N., and Pessiglione, M. (2022). Value, confidence, deliberation: a functional partition of the medial prefrontal cortex demonstrated across rating and choice tasks. *J. Neurosci.* 42, 5580–5592. <https://doi.org/10.1523/JNEUROSCI.1795-21.2022>.
57. De Martino, B., Fleming, S.M., Garrett, N., and Dolan, R.J. (2013). Confidence in value-based choice. *Nat. Neurosci.* 16, 105–110. <https://doi.org/10.1038/NN.3279>.
58. Lin, N., Wang, X., Xu, Y., Wang, X., Hua, H., Zhao, Y., and Li, X. (2018). Fine Subdivisions of the Semantic Network Supporting Social and Sensory-Motor Semantic Processing. *Cerebr. Cortex* 28, 2699–2710. <https://doi.org/10.1093/CERCOR/BHX148>.
59. Katayama, R., Shiraki, R., Ishii, S., and Yoshida, W. (2024). Belief inference for hierarchical hidden states in spatial navigation. *Commun. Biol.* 7, 614. <https://doi.org/10.1038/s42003-024-06316-0>.
60. White, T.P., Engen, N.H., Sørensen, S., Overgaard, M., and Shergill, S.S. (2014). Uncertainty and confidence from the triple-network perspective: Voxel-based meta-analyses. *Brain Cognit.* 85, 191–200. <https://doi.org/10.1016/J.BANDC.2013.12.002>.
61. Ondobaka, S., Kilner, J., and Friston, K. (2017). The role of interoceptive inference in theory of mind. *Brain Cognit.* 112, 64–68. <https://doi.org/10.1016/J.BANDC.2015.08.002>.
62. Koster-Hale, J., and Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron* 79, 836–848. <https://doi.org/10.1016/J.NEURON.2013.08.020>.
63. Henson, R.N., and Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus* 20, 1315–1326. <https://doi.org/10.1002/hipo.20857>.
64. Clairis, N., and Pessiglione, M. (2024). Value estimation versus effort mobilization: a general dissociation between ventromedial and dorsomedial prefrontal cortex. *J. Neurosci.* 44, e1176232024. <https://doi.org/10.1523/JNEUROSCI.1176-23.2024>.
65. Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* 15, 85–93. <https://doi.org/10.1016/J.TICS.2010.11.004>.
66. Thakral, P.P., Madore, K.P., Kalinowski, S.E., and Schacter, D.L. (2020). Modulation of hippocampal brain networks produces changes in episodic simulation and divergent thinking. *Proc. Natl. Acad. Sci. USA* 117, 12729–12740. [https://doi.org/10.1073/PNAS.2003535117/SUPPL\\_FILE/PNAS.2003535117.SAPP.PDF](https://doi.org/10.1073/PNAS.2003535117/SUPPL_FILE/PNAS.2003535117.SAPP.PDF).
67. Martin, V.C., Schacter, D.L., Corballis, M.C., and Addis, D.R. (2011). A role for the hippocampus in encoding simulations of future events. *Proc. Natl. Acad. Sci. USA* 108, 13858. <https://doi.org/10.1073/pnas.1105816108>.
68. Bettio, L.E.B., Rajendran, L., and Gil-Mohapel, J. (2017). The effects of aging in the hippocampus and cognitive decline. *Neurosci. Biobehav. Rev.* 79, 66–86. <https://doi.org/10.1016/J.NEUBIOREV.2017.04.030>.
69. Nassar, M.R., Bruckner, R., Gold, J.I., Li, S.C., Heekeren, H.R., and Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nat. Commun.* 7, 11609. <https://doi.org/10.1038/ncomms11609>.
70. Reiter, A.M.F., Diaconescu, A.O., Eppinger, B., and Li, S.C. (2021). Human aging alters social inference about others' changing intentions. *Neurobiol. Aging* 103, 98–108. <https://doi.org/10.1016/J.NEUBIOLAGING.2021.01.034>.
71. Tsvetanov, K.A., Henson, R.N.A., Tyler, L.K., Davis, S.W., Shafto, M.A., Taylor, J.R., Williams, N., Rowe, J.B., and Rowe, J.B. (2015). The Effect of Ageing on fMRI: Correction for the Confounding Effects of Vascular Reactivity Evaluated by Joint fMRI and MEG in 335 Adults. *Hum. Brain Mapp.* 36, 2248–2269. <https://doi.org/10.1002/hbm.22768>.
72. Kitayama, S., and Park, J. (2010). Cultural neuroscience of the self: understanding the social grounding of the brain. *Soc. Cogn. Affect. Neurosci.* 5, 111–129. <https://doi.org/10.1093/SCAN/NSQ052>.
73. Mesquita, B., De Leersnyder, J., and Albert, D. (2014). The cultural regulation of emotions. In *The Handbook of Emotion Regulation*, J.J. Gross, ed. (Guilford Press), pp. 284–301.
74. Mumford, J.A., Demidenko, M.I., Bjork, J.M., Chaarani, B., Feczko, E.J., Garavan, H.P., Hagler, D.J., Jr., Nelson, S.M., Wager, T.D., and Poldrack, R.A. (2025). Unintended bias in the pursuit of collinearity solutions in fMRI analysis. *Imaging Neurosci.* 3, IMAG.a.958. <https://doi.org/10.1101/2025.01.14.633053>.
75. Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebr. Cortex* 26, 3508–3526. <https://doi.org/10.1093/cercor/bhw157>.
76. Cai, M.B., Schuck, N.W., Pillow, J.W., and Niv, Y. (2019). Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. *PLoS Comput. Biol.* 15, e1006299. <https://doi.org/10.1371/journal.pcbi.1006299>.
77. McLaren, D.G., Ries, M.L., Xu, G., and Johnson, S.C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* 61, 1277–1286. <https://doi.org/10.1016/J.NEUROIMAGE.2012.03.068>.
78. Gitelman, D.R., Penny, W.D., Ashburner, J., and Friston, K.J. (2003). Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage* 19, 200–207. [https://doi.org/10.1016/S1053-8119\(03\)00058-2](https://doi.org/10.1016/S1053-8119(03)00058-2).

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE            | SOURCE                                                                   | IDENTIFIER                                                                                                                                                                                                                                                                  |
|--------------------------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Deposited data</b>          |                                                                          |                                                                                                                                                                                                                                                                             |
| Analysis code                  | This study                                                               | DOI: <a href="https://doi.org/10.5281/zenodo.20036054">https://doi.org/10.5281/zenodo.20036054</a> ; GitHub repository: <a href="https://github.com/MLP28/shared-emotion-inference-mpfc/tree/v1.0.0">https://github.com/MLP28/shared-emotion-inference-mpfc/tree/v1.0.0</a> |
| <b>Software and algorithms</b> |                                                                          |                                                                                                                                                                                                                                                                             |
| MATLAB 2019a                   | MathWorks                                                                | <a href="https://www.mathworks.com">https://www.mathworks.com</a>                                                                                                                                                                                                           |
| SPM12                          | Wellcome Trust Center for Neuroimaging                                   | RRID:SCR_007037; <a href="https://www.fil.ion.ucl.ac.uk/spm/">https://www.fil.ion.ucl.ac.uk/spm/</a>                                                                                                                                                                        |
| DISTATIS                       | Hervé Abdi                                                               | <a href="https://www.utdallas.edu/~herve/">https://www.utdallas.edu/~herve/</a>                                                                                                                                                                                             |
| Illustrator                    | Adobe                                                                    | RRID: SCR-010279                                                                                                                                                                                                                                                            |
| E-Prime (version 2)            | Psychology Software Tools (PST), Inc.                                    | <a href="https://www.pstnet.com/eprime">https://www.pstnet.com/eprime</a>                                                                                                                                                                                                   |
| RSA-RSM toolbox                | R. Viviani                                                               | Viviani (2021); arXiv preprint arXiv:2102.08931 <a href="https://github.com/roberto-viviani/rsa-rsm">https://github.com/roberto-viviani/rsa-rsm</a>                                                                                                                         |
| <b>Other</b>                   |                                                                          |                                                                                                                                                                                                                                                                             |
| Brainnetome Atlas              | Brainnetome Center, Institute of Automation, Chinese Academy of Sciences | RRID:SCR_014091; <a href="http://atlas.brainnetome.org">http://atlas.brainnetome.org</a>                                                                                                                                                                                    |

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study participants from the fMRI study, first and second pre-experimental studies were all human participants.

The fMRI study population consisted in 20 young subjects aged 20–30 years (10 men, mean age  $24 \pm 2.7$  years, mean years of education  $13.7 \pm 1.6$  years), 20 middle-aged participants aged 30–49 years (10 men, mean age  $40 \pm 5.9$  years, mean years of education  $13.7 \pm 2.0$  years), and 20 older subjects aged 51–75 years (9 men, mean age  $63.3 \pm 6.7$  years, mean years of education  $12.1 \pm 3.0$  years). Mean years of education did not significantly differ between all of the three groups. None of the participants over 50 years of age had pathological global cognitive performances at the Mattis Dementia Rating Scale.

In the first pre-experimental study, 86 control participants aged 18–80 years (31 men, mean age  $28.3 \pm 13.4$  years, mean years of education  $16 \pm 2.3$  years) were included.

In the second pre-experimental study, 56 participants aged 20–70 years (18 men, mean age  $37.3 \pm 17.4$  years, mean years of education  $13.5 \pm 3.7$  years) were included.

Participants of the fMRI study were 31 women and 29 men. Findings of this study apply to women and men. Each age group had an equivalent number of woman and men, controlling for a potential sex effect. Participants from the first pre-experimental study were 55 women and 31 men. Participants of the second pre-experimental study were 38 women and 18 men. Sex was determined based on self-reporting. As the design of the study controlled for a potential sex effect and this effect was not of interest, we did not perform sex-based analyses.

We did not use any socially constructed or socially relevant categorization variable (such as race or ethnicity).

The experiment was approved by the French national ethics committee (Institutional review board CPP Nord-Ouest III), and all participants gave their written informed consent. The study was carried out following the Declaration of Helsinki.

### METHOD DETAILS

#### Participants

20 young subjects aged 20–30 years (10 men, mean age  $24 \pm 2.7$  years, mean years of education  $13.7 \pm 1.6$  years), 20 middle-aged participants aged 30–49 years (10 men, mean age  $40 \pm 5.9$  years, mean years of education  $13.7 \pm 2.0$  years), and 20 older subjects aged 51–75 years (9 men, mean age  $63.3 \pm 6.7$  years, mean years of education  $12.1 \pm 3.0$  years) participated in this study. Participants underwent anatomical and task-based functional MRI. Data from 3 middle-aged participants were excluded from the analysis: one fell asleep during the MRI acquisition, another had too many movements, and the latter had white matter abnormalities revealed by the anatomical MRI. One participant from the old subjects' group was also excluded because of important artifactual signals on the fMRI acquisition. All participants were French native speakers and right-handed, and none of them reported a history of addiction, psychiatric and neurological disorders, or cognitive complaints. The experiment was approved by the regional ethics committee

(CPP Nord-Ouest III), and all participants gave their written informed consent. The study was carried out following the Declaration of Helsinki. Mean years of education did not significantly differ between all of the three groups (ANOVA:  $F(2,57) = 3.098, p = 0.053$ ). None of the participants over 50 years of age had pathological global cognitive performances at the Mattis Dementia Rating Scale (mean score =  $142.1 \pm 1.8 / 144$ , minimum score = 138/144).

## Procedure

### Task

The affective mental state attribution task used in this paper originates from the same experiment as reported in.<sup>46</sup> This task was carried out by participants, while brain activity was recorded using functional MRI.

The affective mental state attribution task was composed of 120 trials consisting in mute, black-and-white videos that featured two protagonists, Pierre and Marie, presented as roommates (see [Figure 1](#) and [Video S1](#) for example of two trials). The scenes occurred in five different places (dining and living rooms, office, entrance, and park). Trials were divided into two parts. The “mental state inference” videos depicted social interactions of everyday life between the two protagonists, which provoked an affective mental state in one of them. This protagonist was the target of the mental state inference and was distinguished by a pink armband. Videos of this inference part lasted 10–11 s, including a full shot (8–9 s) giving an overview of the context, followed by a close-up shot (2 s) providing a deeper understanding of the situation. The observation of these social situations was followed by a fixation cross (semi-random duration of 2–6 s) and then by a “mental state recognition” part presented in the form of 4-s videos framed on the face and chest of the character who was the target of the inference in the first part, showing an emotional facial expression. The protagonist presented was the target of the affective state recognition. Expression conveyed in this recognition part and the mental state elicited by the social situation during the preceding mental state inference video were congruent in half of the videos and incongruent in the other half. Five mental states could be elicited by the videos: 4 distinct emotional states (i.e., surprise, anger, pride, and embarrassment), crossed in terms of valence and complexity, and a neutral state (i.e., absence of emotion). The latter was considered a control condition and was attributed to videos in which no protagonist experienced any emotion.

Trials were presented in a pseudo-random fashion between participants and were divided into 4 fMRI runs of 30 trials each. Participants were instructed to think about the affective mental state that the protagonist wearing the pink armband could feel during the videos. The emotions inferred were not labeled. Participants only had to make a congruency judgment at the end of the mental state recognition videos, between the emotion felt by the protagonist in the social situation and the one conveyed by the facial expression. Specifically, participants viewed the whole video each time and could make their judgment of emotional congruency during the subsequent recognition phase or immediately after (i.e., during the fixation cross). If a participant did not provide a response by the end of the fixation cross, the trial was recorded as a non-response. For more details about stimuli and the design of the task, see.<sup>46</sup> The number of responses by participants was assessed, as well as the correct response rate at the congruency judgment derived by summing all correct responses for each participant divided by the total number of responses, and response times during the congruency judgment.

### Pre-experimental study

A first pre-experimental study (pre-study 1) was conducted via a website and was conveyed to all the students of the University of Caen Normandy, but also shared with individuals beyond the university. In this pre-experimental study, 86 control participants aged 18–80 years (31 men, mean age  $28.3 \pm 13.4$  years, mean years of education  $16 \pm 2.3$  years) rated the intensity of emotions felt by the protagonist wearing the pink armband for each video separately, according to those 4 distinct emotions; between 0 (does not feel at all) and 10 (does feel this emotion very strongly). The neutral state that was not considered an affective mental state was set to 10 if selected, as its intensity could not be assessed. This allowed us to characterize the intensity of the target dominant emotion that was felt by the main characters in the video, as well as intensities of alternative competing emotions. In a second pre-experimental study, 56 participants aged 20–70 years (18 men, mean age  $37.3 \pm 17.4$  years, mean years of education  $13.5 \pm 3.7$  years) rated the precise moment when the attribution of the mental state could be made. For these pre-experimental studies, additional variables were quantified such as the familiarity and the difficulty in understanding the situations. In both pre-experimental studies, participants could stop the task at any time, which is why some did not fully complete the task.

Following a  $2 \times 2$  factorial design, 4 indices were extracted from the first pre-experimental study: accuracy of dominant emotion, accuracy of competing emotions, uncertainty of dominant emotion, and uncertainty of competing emotions ([Figure 2](#)). Those indices characterizing each mental state inference were computed after fitting the distribution of the scale intensity with a log-normal function. Two distributions were fitted: one for the dominant emotion (i.e., the emotion *a priori* considered as the one to be inferred) and one for the average of intensity judgments expressed for the 3 other alternative emotions. “Accuracy” corresponds to the mean of the intensity distributions while “Uncertainty” was operationalized as the standard deviation of the distributions across participants. The time courses of the attributions of emotional state, irrespective of the emotion expressed by the main character, were extracted from the second pre-experimental study.

### Statistical analysis of behavioral performances

In the main experiment, non-parametric Kruskal Wallis ANOVAs were applied to assess the main effect of the group in the number of responses, correct response rate, and response times at the congruency judgment. Post-hoc tests consisting of Mann Whitney tests were further used to characterize differences between each group. To assess the relationship between the computational indices

characterizing the mental state inference (i.e., type of emotion crossed with parameters), and behavioral responses, mixed-effects models were applied. A full linear mixed-effects model for response time, and a mixed-effects model using binomial distribution family and a logit link function for correct response rate, were implemented. Both models included the four computational indices, their interactions with age group, and participant-level random intercepts and slopes (i.e., (accuracy of dominant emotion + accuracy of competing emotion + uncertainty of dominant emotion + uncertainty of competing emotion)  $\times$  group). In these analyses, in order to increase statistical power, and for a more intuitive interpretation of potential interactions, the young group was compared to an 'Older' group including middle-aged and old individuals.

### MRI acquisition parameters

Anatomical and functional MRI scanning sessions were conducted using a Philips Achieva 3.0 T scanner, at the Cyceron center (Caen, France). A fast field echo sequence was used to acquire high resolution T1-weighted anatomical image (3D-T1-FFE sagittal, repetition time (TR) = 20 ms; echo time (TE) = 4.6 ms; flip angle 10°; 180 slices; no gap; slice thickness = 1 mm; field of view (FOV) = 256  $\times$  256 mm<sup>2</sup>; in-plane resolution = 1  $\times$  1 mm<sup>2</sup>). In a second session, five functional runs were acquired. The four first runs recorded brain activity during the affective mental state attribution task and the fifth run consisted in a localizer task that was not exploited in this study. The first six volumes of each run were discarded, owing to saturation effects. Functional task-based MRI scans were thus acquired using an interleaved 2D T2\* EPI sequence (2D-T2\*-FFE-EPI axial, SENSE factor = 2; TR = 2382 ms; TE = 30ms; flip angle = 80°; 42 slices; no gap; slice thickness = 2.8 mm; FOV = 224  $\times$  224 mm<sup>2</sup>; in-plane resolution = 2.8  $\times$  3.0 mm<sup>2</sup>; 275 volumes per run of the affective mental state attribution task and 231 volumes for the localizer task).

### Preprocessing

All functional imaging analyses were performed using Statistical Parametric Mapping 12 software (SPM12, Wellcome Department of Imaging Neuroscience). The MATLAB analysis codes supporting this study are available at: <https://doi.org/10.5281/zenodo.20036054>. Preprocessing consisted firstly in slice timing correction of functional images by temporal interpolation, which were then spatially realigned. Segmentation of anatomical T1-weighted images was operated. This allowed the following skull-stripping of T1-weighted from these segmented images, to facilitate the following coregistration step. The anatomical T1 skull-stripped image was coregistered to the mean functional image and this transformation was applied to all the other functional images. After this step, segmented anatomical images were normalized onto the template of the Montreal Neurological Institute (MNI), and the nonlinear deformations were used to normalize the realigned functional data. Normalized functional images were smoothed with a 10-mm FWHM kernel. For all participants, a data quality check was performed for each preprocessing step. In the case of representational similarity analysis, normalization and smoothing were not applied, to carry out analyses in the subject's native space.

### First-level univariate analysis

First-level analyses were conducted using general linear models (GLMs). Pre-experimental ratings were used to model brain activity in response to emotional inference during the social context presentation. We first designed a first-level fMRI model including regressors encoding the temporal distribution of inference (from pre-study 2) for both emotional and neutral video contexts, without further modulation by accuracy or uncertainty. The onsets of response time from each participant of the pre-experimental study were encoded as a stick function to encode the moment when inference of emotional state is attributed during the context. This stick function encoding the timing of inference was smoothed with a Gaussian kernel of 3000ms before being convolved with a canonical Hemodynamic Response Function (HRF). This procedure ensures that the neural function of the GLM reflected the moment of the context video when the inference of emotional state was maximal in the pre-experimental study population. In a second first-level GLM, the stick function encoding the timing of inference (pre-study 2) was further parametrically modulated by the mean-centered estimates of accuracy and uncertainty (pre-study 1), for both dominant and competing emotions, before smoothing and convolution with the canonical HRF. This GLM also included non-modulated emotional and neutral regressors capturing the moment of inference for these conditions. For both GLM, additional nuisance regressors encoding the whole context and expression videos were included to account for the confounding effects of watching videos. Trials in which the final congruency judgment was incorrect were classified as errors and not used to construct the regressors of interest. For each run, the 6 movement parameters extracted from the realignment procedure, were also added as nuisance regressors in the design matrices. A high pass filter (1/128 Hz) was applied and autocorrelation between fMRI time series was corrected using a first-order autoregressive AR(1) model of noise temporal autocorrelation estimated using restricted maximum likelihood (ReML).

In the first GLM, first-level contrast corresponded to contrasting emotional against neutral regressor to assess brain substrates of emotional inference. In the second GLM, first-level contrasts, corresponding to main effect of accuracy (including accuracy of dominant emotion and accuracy of competing emotions regressors) and main effect of uncertainty (including uncertainty of dominant emotion and uncertainty of competing emotions regressors) conditions, were estimated against implicit baseline. First-level contrasts also included the effect of emotion (competing vs. dominant emotions and dominant vs. competing emotions) for both accuracy and uncertainty. We conducted two collinearity diagnostics: for each contrast, we computed the contrast-level Variance Inflation Factor (cVIF) recently proposed by Mumford and colleagues.<sup>74</sup> This is a useful tool measuring how the correlation between predictor variables inflates the variance of the contrast performed on the estimated regression coefficients. cVIF values for the

contrasts indexing main effect of accuracy, main effect of uncertainty and emotional competition for accuracy and uncertainty were all  $< 8$ , below the commonly accepted threshold of 10. Condition Index metrics, which is another measure of collinearity assessing variability across design matrix dimension (<https://montilab.psych.ucla.edu/designdiagnostics/>), were inferior to 6, below the recommended upper limit ( $< 15$ ), confirming that the design matrix has sufficient dimensional variability. Taken together, these diagnostics indicate that collinearity does not compromise inference validity in our GLM.

### Univariate brain analysis

In the first GLM, effect of emotional inference was tested using a one-sample  $t$  test in SPM with a threshold of  $p < 0.05$ , FWE corrected for multiple comparisons across the mPFC or whole-brain, with a cluster size set to  $k = 10$ . In the second GLM, among the first-level contrasts, main effect of accuracy and uncertainty were tested using one-sample  $t$ -tests, with a threshold of  $p < 0.05$ , FWE corrected for multiple comparisons across the mPFC or whole-brain, with a cluster size set to  $k = 10$ . Effects of type of emotions (dominant and competing) were tested using one-sample  $t$ -tests in SPM, with a threshold of  $p < 0.025$ , FWE corrected for multiple comparisons across the mPFC or whole-brain, with a cluster size set to  $k = 10$ . Given the absence of *a priori* assumptions regarding the neural mechanisms underlying emotion differentiation, bilateral comparisons were performed comparing brain activities associated with dominant and competing emotions. For both GLMs, the effect of age was investigated using a two-sample  $t$  test for each first-level contrast of interest, contrasting young with older groups. To increase statistical power, and for a more intuitive interpretation of potential interactions, the young group was compared to an 'Older' group including middle-aged and old individuals. This grouping was based on the observation of homogeneous brain activations during the task between those two groups. The SPM maps were thresholded at  $p < 0.05$ , FWE corrected for multiple comparisons across the mPFC or whole-brain, with a cluster size set to  $k = 10$ . The analysis mask corresponded to the ventral and dorsal mPFC subregions. The center coordinates of these subregions were determined through a meta-analysis of brain regions involved in various mindreading tasks.<sup>49</sup> The regions in the Brainnetome atlas<sup>75</sup> overlapping with the ventral (coordinates in the Montreal Neurological Institute (MNI) space [3, 51, -7]) and dorsal ([-1, 54, 24]) mPFC center coordinates were used (see Table S2 for more information about the ROIs).

### Representational similarity analysis

Representational Similarity Analysis (RSA,<sup>47</sup>) was employed to investigate how different aspects of emotional competition are encoded during mental state inference in the brain (Figure 5). This method operates under the premise that representational competition is spatially encoded and distributed through the activity of contiguous voxels within a given brain region. This representational competition is captured by measuring the distances between the spatial distribution of activity associated with each trial, which are then encapsulated within Representational Dissimilarity Matrices (RDMs). These RDMs are subsequently compared to a model RDMs encoding predicted similarity relationships between stimuli along specific representational dimensions. In this study, we focused on two dimensions of emotional competition: 1) the emotional content associated with competing emotions, 2) the magnitude of emotional interference caused by those competing emotions.

At the first level, the design matrix included a regressor per trial. Each regressor was modeled with the onsets of response times from all participants of the pre-experimental study. These 120 regressors were encoded as a stick function, without any parametric modulation. They were convolved with a Gaussian kernel of 3000ms and further convolved with a canonical HRF. Additionally, the duration of contexts and expression videos were included as nuisance regressors to remove the confounding effects of watching videos. For each run, the 6 movement parameters derived from the realignment procedure were incorporated as nuisance regressors in the design matrix. A high pass filter (1/128 Hz) was applied and intrinsic autocorrelation was corrected to consider autocorrelation between the GLM residuals.

The RSA-RSM toolbox developed by Viviani (2021) was used, allowing carrying out RSA, and addressing the issue of non-orthogonality of the design matrix. Specifically, it acknowledges that regressors of temporally close trials may exhibit correlation, introducing a potential bias in the analysis as noted by previous studies.<sup>48,76</sup> To take this bias into account, Viviani (2021) proposed to apply a method of partial correlations for computing neural patterns (for more information, see<sup>48</sup>). More precisely, this bias was accounted for by including covariance structure of the beta estimates as confound in the partial correlation between neural and model RDMs. It removed variance attributable to temporal correlation bias or regressor non-orthogonality; ensuring RSA reflects neural representational competition rather than artifacts of the design.

Two model RDMs were constructed, termed emotional content RDM and interference RDM, each representing pairwise relationships between trials along one of the representational dimensions described above. These models incorporated intensities associated with competing emotions, as derived from pre-experimental study 1. Regarding emotional content, for each participant in the pre-experimental study, an individual model RDM was generated by computing the cosine distance between every trial pair using the vectors of intensities associated with competing (i.e., non dominant) emotions. When comparing two trials sharing the same dominant emotion, the competing emotions encompassed the three remaining alternative emotions. Conversely, when contrasting trials with distinct dominant emotions, the competing emotions comprised the 2 competing emotions common to both trials, excluding their respective dominant emotions. Regarding interference, for each participant in the pre-experimental study, an individual model RDM was generated by computing the Euclidean distance between every trial pair based on the summed intensities of competing emotions, which operationalizes the overall magnitude of emotional competition. This RDM, regardless of the content of the competition, reflects similarities in terms of degree of interference caused by the intensities of the competing emotions. Then, for both

dimensions, the DISTATIS method was used to determine the common structure across individual model RDMs, resulting in two group-level RDMs<sup>50</sup> (i.e., emotional content and interference RDMs). DISTATIS captures the shared similarity structure across individuals, by seeking optimal agreement (or compromise) across multiple distance matrices. The implementation in MATLAB of DISTATIS can be found at <https://www.utdallas.edu/~herve/>. In brief: (1) each of the individual RDMs was first transformed to a cross-product matrix after double-centering and normalization to its first eigenvalue; (2) the cosine/Euclidean similarity structure of all pairwise comparisons of the normalized cross-product matrices was computed using the RV coefficient; (3) the RV coefficient matrix describing the relationships between RDMs was submitted to eigen-decomposition, and the compromise matrix corresponded to the sum of the normalized cross-product matrices weighted by their first eigenvalue; (4) the eigen-decomposition of the compromise produced factor scores, which described the position of each of the context video in an N-dimensional compromise space; and (5) the distance correlation matrix of these multidimensional compromise spaces then corresponded to the best agreement across all individual RDMs. The resulting distance correlation matrix therefore can be viewed as a shared space reflecting the common competition proximity across individuals. In this space, the proximity of two situations reflects similarity either in the content of the emotional competition or in the degree of interference caused by competing emotions, depending on the dimension considered. Brain regions whose neural similarity structure matches these model RDMs can therefore be interpreted as representing either the content of competing emotions or the magnitude of emotional interference. Because the toolbox operates on similarity maps, the model RDMs were transformed into representational similarity matrices (RSMs) by applying the following formula:  $RSM = 1 - RDM / \max(RDM)$ .

Two additional matrices were also computed with the goal to include them as nuisance covariates in the subsequent searchlight analysis, controlling for the effect of the dimensions they represent. The inference timing RDM controlled for a potential bias introduced by the similarity between trial pairs in terms of the moment of the inference during the context videos. For each participant in the pre-experimental study, an individual timing of inference RDM was generated by computing the Euclidean distance between every trial pairs in terms of the timing of inference during the mental state inference phase. Participants with more than 50% missing responses were excluded to ensure stability, resulting in 40 participants out of the 56 participants of the pre-experimental study. The second control matrix, the dominant intensity RDM, reflected a potential bias introduced by the similarity between trial pairs in terms of the intensity of the dominant emotion. Individual dominant intensity RDMs were generated by computing the Euclidean distance between every trial pairs in terms of the intensity of the dominant emotion, irrespective of emotion label. For both confound dimension, a group-level RDM was computed, using the DISTATIS method to determine the commonalities across the individual control RDMs; and RDMs were transformed into RSMs.

The toolbox developed by Viviani (2021) allows performing individual searchlight analyses. More precisely, a spherical ROI with a radius of 8mm was moved through voxels of the individual beta maps. For each voxel, a brain RSM was computed by comparing neural activity across voxels encompassed within the searchlight ROI for every trial pair. This brain RSM was compared with emotional content and interference RSMs respectively. This was done using partial correlation, allowing to include nuisance covariates in the correlation when computing the brain RSM. Two covariates, called BCov and SCov allowed to estimate the temporal correlation bias, as described above (for more details see<sup>48</sup>). The two control matrices (i.e., timing of inference and dominant intensity) were also included as nuisance covariate in these searchlight analyses. Hence, four nuisance covariates were controlled for in the partial correlation analysis between model and brain RSMs.

The resulting correlation values between the model similarity structure (emotional competition or interference RSMs) and the similarity of brain activity (brain RSM) in the area of the sphere were stored in individual *r*-maps. To test the robustness of the results from these analyses, a complementary sensitivity analysis was conducted, following the same methodology as described previously, but with searchlight radii of 6mm and 10mm.

The individual *r*-maps were normalized onto the MNI template and smoothed with an 8-mm FWHM kernel. Subsequently, one-sample *t*-tests were conducted to identify brain regions representing 1) competition content and 2) emotional interference due to competition. Given the strong hypothesis on the mPFC, a small volume correction was performed, using the *a priori* defined mask corresponding to the ventral and dorsal mPFC. Significance thresholds were set at  $p < 0.05$ , small volume FWE corrected. Additionally, age group effects were highlighted via two-sample *t*-tests. Significance thresholds for the group effect were set at  $p < 0.05$ , small volume FWE corrected.

Finally, to dig deeper into the mechanisms subtended by the mPFC, we examined whether individual differences in executive functions could explain the activity of the dmPFC related to the encoding of degree of interference. Executive function measures were available as part of a broader experimental protocol, including inhibitory control (Stroop interference time), cognitive flexibility and set-shifting (Trail Making Test B–A time), verbal fluency (mean of animal, P, and V), and working memory (forward and backward visuospatial span). We conducted Spearman's correlations between these executive functions variables and the RSA rho estimates capturing the encoding of competing-emotion interference in dmPFC, across all participants.

### Psychophysiological interaction analysis

We then examined task-dependent functional connectivity of the mPFC using a generalized psychophysiological interaction (gPPI) framework.<sup>77</sup> Seed regions corresponding to the vmPFC and dmPFC were defined from the Brainnetome atlas, in the same way as in the univariate analysis. Subject-specific time series were extracted within MNI-defined mask voxels. gPPI models were constructed using the same first-level design matrix as in the univariate analysis.

For each seed, voxelwise BOLD signals were extracted and averaged across all voxels within the ROI. To isolate variance specifically related to the psychological context of interest, the physiological timeseries was residualized with respect to all nuisance regressors included in the first-level GLM. The timeseries adjusted for confound were then deconvolved to estimate the neural activity. A full-rank cosine basis set convolved with the HRF, as well as the filtered and whitened matrix of confounds, was used as the design matrix of a hierarchical linear model to estimate the underlying neuronal activity under a parametric empirical Bayes scheme.<sup>78</sup>

Six task conditions were modeled (i.e., Psychological regressors): accuracy–dominant, accuracy–competing, uncertainty–dominant, uncertainty–competing, as well as unmodulated emotion and neutral conditions. Event onsets were converted into high-resolution (1 ms) stick functions, smoothed using a Gaussian kernel, down-sampled to the repetition time (TR = 2.382 s), detrended, and convolved with the canonical HRF.

For each condition, the neural functions of the psychological regressor was multiplied with the deconvolved seed activity to generate the interaction term (i.e., PPI regressor), which was then HRF-convolved and resampled to the TR. All PPI regressors were detrended before model estimation. For each participant and seed, a subject-level gPPI GLM was specified containing (1) the interaction PPI regressors, (2) the psychological regressors, (3) the physiological seed timeseries, and (4) a constant term. The model was estimated voxelwise across the whole brain, and beta coefficients for each PPI regressor were written to image space and smoothed with an 8-mm FWHM Gaussian kernel.

Six contrasts probed condition-dependent modulation of mPFC connectivity, allowing us to assess effects of accuracy, uncertainty, and dominant versus competing emotional content. Group-level analyses were performed using one-sample t-tests to characterize condition-specific connectivity at the population level, and two-sample t-tests to evaluate between-group differences in mPFC functional coupling.

### Relationships between behavioral performances and brain estimates

Finally, to gain an understanding of the involvement of the activity of the mPFC in social behavior, the relationships between the behavioral performances during the task, and brain estimates obtained from both univariate analysis and RSA were investigated, all groups combined. Behavioral performances were defined as the correct response rate at the congruency judgment. We computed Spearman's correlations between behavioral performances and: 1) estimates of accuracy and uncertainty crossed with type of emotions in the univariate analyses, and 2) correlation coefficients extracted from interference model in RSA; within the predefined ROI corresponding to the dorsal and ventral mPFC (see Table S2). The expected proportion of type I error across multiple ROIs was controlled using the FDR correction.

## QUANTIFICATION AND STATISTICAL ANALYSIS

N in this study represents the number of human participants.

Kruskal-Wallis ANOVAs were used to assess group differences in behavior (number of response, correct response rate, and response times at the congruency judgment). Post-hoc tests consisted of Mann Whitney tests. Groups were composed of 20 young, 17 middle-aged and 19 older adults.

To assess the relationship between the computational indices characterizing the mental state inference (i.e., type of emotion crossed with parameters), and behavioral responses, mixed-effects models were applied. A full linear mixed-effects model for response time, and a mixed-effects model using binomial distribution family and a logit link function for correct response rate, were implemented. Groups were composed of 20 young and 36 old individuals (middle-aged and older adults were merged into a single group).

In the univariate analysis, the correlates of statistical regularities in the mPFC were assessed with one-sample t-tests in SPM ( $N = 56$ ). For main effects of accuracy and uncertainty, significance threshold of  $p < 0.05$ , corrected for family-wise error (FWE) accounting for multiple comparisons across the mPFC, with a cluster size  $k = 10$  was employed. For effects of type of emotions (dominant vs. competing), given the absence of *a priori* assumptions, bilateral comparisons were performed. Thus, a threshold of  $p < 0.025$ , FWE corrected for multiple comparisons across the mPFC, with a cluster size  $k = 10$  was applied. Two-sample t-tests were applied for assessing age effect ( $N = 20$  young and 36 old adults). The SPM maps were thresholded at  $p < 0.025$ , FWE corrected for multiple comparisons across the mPFC, with a cluster size  $k = 10$ .

In the RSA, partial correlations were applied in searchlight analysis, using the toolbox from Viviani (2021). Then, one-sample t-tests were conducted to identify main effects of emotional content and interference ( $N = 56$ ). A small volume correction was applied on the mPFC. Significance thresholds were set at  $p < 0.05$ , small volume FWE corrected. Additionally, age group effects were highlighted via two-sample t-tests ( $N = 20$  young and 36 old adults). Significance thresholds for the group effect were set at  $p < 0.05$ , small volume FWE corrected. The relationships between executive functions and the RSA rho estimates capturing the encoding of competing-emotion interference in dmPFC were explored with Spearman's correlations ( $N = 56$ ).

For the PPI, one-sample t-tests characterized condition-specific connectivity at the population level ( $N = 56$ ), and two-sample t-tests assessed between-group differences in mPFC functional coupling ( $N = 20$  young and 36 old adults).

Relationships between behavior and brain estimates were performed with Spearman's correlation ( $N = 56$ ). The expected proportion of type I error across multiple ROIs was controlled using the FDR correction. For all behavioral analyses, significance was considered when  $p < 0.05$ .

MATLAB 2019a was used for statistical analyses of behavioral data. MATLAB2019a with SPM12 was used for univariate statistical analyses, RSA and PPI analyses. For the RSAs the DISTATIS method (Abdi et al. 2012) and the RSA-RSM toolbox from Viviani (2021) were used. The behavioral task was presented to participants in the scanner using E-Prime software (<https://pstnet.com>).

No statistical methods were used to pre-determine the sample size. However, the sample size used in this study is consistent with previous studies investigating neural markers of social cognition or the brain valuation system.

All participants were French native speakers and right-handed, and none of them reported a history of addiction, psychiatric and neurological disorders, or cognitive complaints.

In the fMRI study, data from 3 middle-aged participants were excluded from the analysis: one fell asleep during the MRI acquisition, another had too many movements, and the latter had white matter abnormalities revealed by the anatomical MRI. One participant from the old subjects' group was also excluded because of important artifactual signals on the fMRI acquisition.

All the statistical details of experiments can be found in the Method section, and in each corresponding result section. Significance thresholds were indicated in figures representing brain analyses. Behavioral results are presented in mean. Error bars indicate confidence intervals.