scientific reports

OPEN

Lesion to the vmPFC abolishes intentionality (mis)attribution in the Knobe effect

Francesca Starita[1]✉, Luigi A. E. Degni[1,2], Daniela Dalbagno[1] & Elisa Ciaramelli[1]

Understanding others' intentions is crucial for effective social interaction; however, the psychophysiological and neural mechanisms that drive intentionality attribution remain elusive. The Knobe Effect refers to the tendency to attribute more intentionality to actions resulting in negative compared to positive side effects, even if side effects are unintentional in both cases. This bias has been associated with the emotional responses and moral judgments elicited by the negative valence of the side effects, but findings remain contradictory, and few studies have explored their neural bases. This study investigates the causal contribution of the ventromedial prefrontal cortex (vmPFC), a region subserving emotional processing and moral judgment, in the Knobe Effect. Participants with vmPFC lesions and healthy controls rated the intentionality and valence of scenarios involving negative or positive side effects. Additionally, they rated the morality of the side effects, the subjective emotional response experienced during scenario reading and had their electrodermal activity recorded. Results showed that vmPFC patients, unlike controls, attributed similar levels of intentionality to positive and negative side effects, thus lacking the expected Knobe Effect. This occurred despite preserved recognition of the valence of the side effects but was accompanied by blunted emotional responses to the side effects and altered moral judgments. These findings support the view that emotional responses and moral judgment are fundamental for the Knobe Effect in healthy adults, and highlight the causal role of the vmPFC in driving this ubiquitous human bias.

Keywords Emotions, Intentionality attribution, Moral judgement, VmPFC, Ventromedial prefrontal cortex, Skin conductance, Electrodermal activity

Understanding others' intentions is crucial for effective social interaction, and impairments in attributing intentionality—whether due to brain lesions or psychiatric disorders—are associated with disruptions in social functioning[1–6]. Intentionality attribution is also fundamental in legal contexts, where judicial systems often distinguish between intentional and unintentional harm[7]. Thus, understanding the mechanisms that underpin intentionality attribution, and its potential misattribution, is of paramount importance.

Although intentions often align with the outcomes of actions, mismatches can occur when one's actions lead to additional unintended side effects. Let's consider this scenario: the CEO of a company starts a new program to increase profit; however, this program also harms the environment. When presented with such a scenario, most people judge the side effect of the CEO's program on the environment as intentional[8]. Crucially, such misattribution of intentionality does not occur if starting the program results in the side effect of helping, rather than harming, the environment. In fact, people ascribe intentionality more readily to actions leading to morally negative than positive side effects. This bias in intentionality attribution is known as the "Knobe Effect" or "side-effect effect"[8–15] and since its discovery it has been repeatedly observed across various scenarios[16,17], cultures[18], and age groups[19].The discovery of the "Knobe Effect" challenged earlier theories of intentionality attribution, which posited that observers ascribe intentionality based solely on the agent's mental states, such as its actual intentions, beliefs, desires, and awareness regarding the outcome of its action[20]. Contrary to this, the Knobe effect revealed that the ascription of intentionality is significantly shaped by the valence of the outcomes: people are more inclined to judge harmful side effects as intentional than helpful ones, even when the agent's actual intention is neutral or identical in both cases. This finding opened the possibility that the observer's subjectivity could play a central role in intentionality attribution. Specifically, what it means for someone to "intentionally" do something may be influenced by the observers' emotional response to and moral

[1]Department of Psychology, Center for Studies and Research in Cognitive Neuroscience, University of Bologna, Cesena, Italy. [2]International School of Advanced Studies, University of Camerino, Camerino, Italy. ✉email: francesca.starita2@unibo.it

judgement of the consequences[9,15,21–23]. In support of such idea, in healthy individuals, judgments of moral blame were found to mediate the relationship between self-reported emotional response and intentionality attribution to negative side effects[16]. Additionally, higher intentionality attribution for negative side effects was found to correlate with higher amygdala response, a brain region crucial for emotional processing[24–26], and this correlation was mediated by self-reported emotional response[16]. Also, our previous work showed that individuals with high levels of alexithymia, a personality trait linked to difficulties in emotional processing[27–32], exhibit a reduced Knobe Effect, due to decreased intentionality attributed to negative side effects, accompanied by a reduced modulation of physiological emotional response[33]. However, findings on the role of emotions and moral judgement in the Knobe Effect are not univocal. For example, a preserved Knobe Effect has been found in adolescents with psychopathic traits, who are characterized by persistent immoral behavior, deficient moral emotions, and impairments in moral judgment[34]. Thus, whether and how emotions and moral judgement shape the Knobe effect remains an open question.Here, we adopt a neuropsychological approach to advance the understanding of the neural and psychophysiological mechanisms underlying intentionality attribution and the Knobe Effect. Specifically, we examine whether the integrity of the ventromedial prefrontal cortex (vmPFC) is critical for the misattribution of intentionality to negative side effects. The vmPFC is well-established as a key region supporting emotional processing and moral judgment, with extensive evidence demonstrating that damage to the vmPFC results in deficits in both domains[35,36]. For example, patients with vmPFC lesions have a reduced ability to recognize emotional expressions[37,38], spending less time attending to emotionally salient facial features[39]. Additionally, the vmPFC plays a pivotal role in supporting the expression and regulation of emotional responses[40–45] with patients often exhibiting blunted physiological responses[46,47], especially in anticipation of future emotional outcomes[48,49]. In terms of moral judgment, vmPFC lesions are associated with altered evaluations of personal moral violations, with patients more likely to find such behaviors acceptable compared to healthy controls[47,50,51]. Notably, this atypical moral judgment was observed to be accompanied by a diminished physiological emotional response to personal moral violations[51]. Additionally, vmPFC damage impairs the ability to evaluate the intentions behind morally significant behaviors, particularly when those intentions are negative rather than positive[52]. Surprisingly, at odds with this evidence are the results of one study that previously investigated the Knobe Effect in vmPFC patients and found a preserved intentionality bias in patients with vmPFC damage[53]. In that study, however, intentionality judgements were limited to two scenarios only, and were made on a dichotomous measure (intentional or not intentional). Thus, we think that the causal role of the vmPFC in intentionality attribution and the Knobe Effect warrants further investigation.

Here, individuals with vmPFC lesions and healthy controls rated the intentionality of side effects with varying valence (positive or negative) and salience (high or low), as part of scenarios modeled after the original Knobe vignette[8,16,33]. To test the contribution of moral judgment and emotional response in intentionality attribution, participants rated the morality of the side effects, reported their subjective arousal while reading the scenarios, and had their electrodermal activity recorded to assess physiological arousal. Additionally, participants rated the valence of the scenarios as a control measure to ensure they could differentiate between negative and positive outcomes. We hypothesized vmPFC patients would exhibit impaired emotional responses and aberrant moral judgements to (negative) side effects. Crucially, if emotional responses and moral judgments drive the Knobe Effects, patients with vmPFC lesions should also exhibit a reduced Knobe Effect, judging the intentionality of actions with negative and positive side effects more similarly than healthy controls do, possibly due to reduced intentionality associated with negative side effects.

Methods
Participants
Two groups of volunteers participated in the study: a group of 11 participants with lesions affecting vmPFC (8 males, mean age = 53.91 years, SD = 10.74, mean education = 10.00 years, SD = 10.09), and a control group of 24 healthy participants. One healthy participant was excluded because they could not complete the experiment, therefore the final sample used for the analysis included 23 healthy participants (14 males, mean age = 56.65 years; SD = 6.77, mean education = 11.64 years, SD = 11.09). All participants had normal or corrected-to-normal vision. Groups did not differ significantly in terms of age ($t_{32} = 0.911$, $p = 0.369$), years of education ($t_{31} = 1.162$, $p = 0.254$) and distribution of males and females ($\chi^2 = 0.458$, $p = 0.498$). Sample size was based on the effects reported in our previous publication using the same experimental paradigm (Zucchelli et al., 2019).

The study was conducted in accordance with institutional guidelines and the World Medical Association Declaration of Helsinki and was approved by the Bioethics Committee of the University of Bologna. All participants gave their written informed consent to take part in the experiment and their privacy rights were observed.

Lesion analysis
vmPFC patients were selected based on the location of their lesion as reported from their medical history, magnetic resonance imaging (MRI) or computerized tomography (CT) and recruited at the Center for Studies and Research in Cognitive Neuroscience of the University of Bologna, where the study was conducted. Imaging of lesions was available for 10 out of 11 patients. vmPFC damage was due to aneurysm of the anterior communicating artery (10 cases) and aneurysm of the right anterior cerebral artery (1 case). For each vmPFC patient, lesion extent and location were documented using the most recent clinical computed tomography or magnetic resonance imaging. Lesions were manually drawn by a trained neuroscientist directly on each slice of the normalized T1-weighted template MRI scan from the Montreal Neurologic Institute[54]. This template is approximately oriented to match Talairach space and distributed with MRIcro[54]. MRIcro[55] was used to estimate lesion volume (in cm^3 and generate brain lesion overlap images. Figure 1 shows the extent and overlap of brain lesions in vmPFC patients. The mainly affected Brodmann's areas (BAs) were BA 10, BA 11, BA 24, BA 25 and

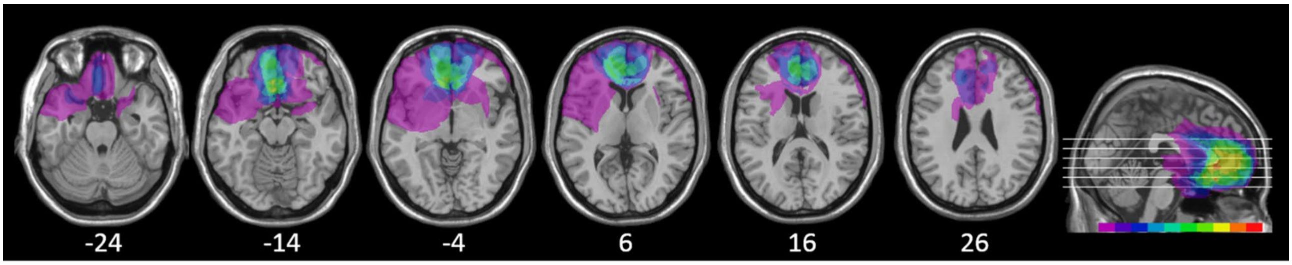

Fig. 1. Location and overlap of brain lesions of vmPFC patients. Representative axial slices and midsagittal view showing cumulative lesions overlap. The white horizontal lines on the sagittal view indicate the positions of the axial slices. The white numbers below the axial slices indicate the z-coordinates of each slice. The color bar indicates the number of overlapping lesions. The left hemisphere is on the left side. Maximal overlap occurs in BAs 10 and 11.

vmPFC patient ID	Sex	Age (years)	Education (years)	Raven progressive matrices	Stroop Test	Digit span	Phonemic Fluency	Semantic Fluency
vmPFC_1	M	52	13	4	4	4	4	4
vmPFC_2	M	57	8	4	4	3	3	3
vmPFC_3	M	55	8	-	4	4	4	4
vmPFC_4	F	61	13	4	2	3	3	4
vmPFC_5	F	69	8	4	1	4	2	3
vmPFC_6	M	48	13	3	0	4	4	3
vmPFC_7	M	57	13	3	-	4	3	3
vmPFC_8	M	60	8	4	3	3	2	4
vmPFC_9	M	54	10	4	4	1	1	2

Table 1. Equivalent scores for each VmPFC patient and neuropsychological test. The equivalent score is a 5-point scale, ranging from 0 to 4, with 0 = pathological performance, 1 = borderline performance, and 2–4 = normal performance. The symbol - indicates that the test's score is missing.

BA 32. The region of maximal lesion overlap occurred in BA 11 (M = 16.54 cc, SD = 10.73), BA 10 (M = 8.20 cc, SD = 7.37).

Neuropsychological assessment

Nine out of the 11 patients with vmPFC lesions completed a neuropsychological assessment that included the following standardized tests: Raven Progressive Matrices[56], Stroop Test[57], Digit Span[58], and Phonemic and Semantic Fluency[59]. For all tests, patients' raw scores were converted into equivalent scores[60], adjusted for age and years of education. The equivalent score is a 5-point scale, ranging from 0 to 4, with 0 = pathologic performance, 1 = borderline performance, and 2–4 = normal performance (see Table 1). As can be seen from the table, vmPFC patients' performance was generally preserved, with the exception of one patient who showed increased interference effects at the Stroop test.

Experimental task

Participants made ratings on 20 scenarios, as in a previously validated paradigm[33]. These scenarios had been selected from a set of 80 scenarios[16], modelled on the original Knobe vignette[8]. In each scenario, a leading character performed an action that resulted in a side effect, which varied in valence (positive, negative) and salience (high, low) across scenarios. Ten scenarios involved a negative side effect (i.e., harming someone/something; e.g., Jenny's pesticide harmed Susie's crops), and ten (corresponding) scenarios involved a positive side effect (i.e., benefiting someone/something; e.g., Stanley's anti-fungal protected Billy's crops; see also[33]). Within both positive and negative scenarios, five scenarios had high salience, that is, the side effect affected people's life or health (e.g., Bill's gadget killed babies), while five scenarios had low salience, that is, the side effect affected the environment, or the quality of the relationships among people (e.g., Curtis release of documents ruined his friend's reputation).

Each scenario was presented for 30 s during which participants read its content. Afterwards, four questions were sequentially presented on the screen, respectively asking to rate (1) the intentionality of the side effect (e.g., Did Jessica intentionally kill her neighbor's crops?), (2) the morality of the side effect (e.g., How morally acceptable is to kill someone's crops?), (3) the subjective arousal experienced in response to the scenario (i.e., How much did the scenario arouse you?), (4) the valence of the scenario (i.e., The scenario was). All questions had to be answered on a scale from 0 (questions 1–3: not at all; question 4: negative) to 10 (questions 1–3:

completely; question 4: positive). Questions remained on the screen until participants provided an answer by pressing a number on the computer keyboard.

Due to technical reasons, two identical versions of the task were used for the experiment, respectively programmed on E-Prime[61] and OpenSEsame3.2[62] software. The first 23 participants (16 healthy participants, 7 vmPFC patients) completed the E-Prime version, while the other 10 participants (7 healthy participants, 3 vmPFC patients) completed the Opensesame version.

Procedure

Participants were comfortably seated in a silent room; experimenters explained to them the procedure and collected the written informed consent. Then, electrodes for skin conductance were attached to the participant and the correct recording of skin conductance was verified. Afterwards, participants' position was centered relative to a computer screen at about 60-cm viewing distance task instructions were provided. They were instructed to carefully read each scenario and answer the questions by pressing one number from 0 to 10 on the computer keyboard. Participants were told to answer as soon as they made a decision, but that there was no time limit. Two practice scenarios (one positive and one negative) were administered to ensure participants understood the instructions. The entire experimental procedure lasted about 45 min.

Dependent measures

Scenarios ratings

For each question, participants' ratings were collected as described above, and then averaged across valence and salience of the scenario, to obtain four average ratings for each participant for each condition, corresponding to negative scenarios with low salience, negative scenarios with high salience, positive scenarios with low salience, positive scenarios with high salience. These ratings enabled us to obtain measures of (1) intentionality attribution, (2) moral judgement, (3) subjective emotional (arousal) response and (4) to control participants' ability to perceive the positive or negative valence of each scenario.

Physiological arousal

Electrodermal activity (EDA) was continuously collected during the task, as a measure of physiological emotional (arousal) response. Note that the psychophysiological registration of one patient was not available for the analyses due to technical problems that occurred during the experimental session. EDA was recorded from pre-gelled snap electrodes (BIOPAC EL501) placed on the hypothenar eminence of the palmar surface of the left hand, using a BIOPAC MP-150 System (Goleta, CA), which fed into AcqKnowledge 3.9 software (Biopac Systems, USA). The signal was recorded at 200 Hz for the E-Prime version and at 1250 Hz for the Opensesame version, but then downsampled at 200 Hz for analysis. The gain switch was set to 5 μS/V and the low-pass to 10 Hz, in order to remove high-frequency noise.

The raw EDA signal was then analyzed using custom-made MATLAB scripts to extract participants' skin conductance level (SCL) and number of skin conductance response (SCR) peaks during the reading of scenarios[63,64]. These two EDA measures were chosen due to the prolonged duration of scenarios presentation. First, the signal was epoched from 0s to 30s after scenario appearance and baseline corrected by subtracting the mean of EDA during 1 s prior to the appearance of the scenario. Then, to examine the modulation of EDA activity while reading the scenarios, SCL and number of peaks were calculated separately for three consecutive 10s intervals (see also[33,51]). For each interval, SCL corresponded to the average EDA response, while the number of SCR peaks to the total number of peaks. A peak was considered valid if it lasted for a minimum of 0.7 and a maximum of 5 s, was greater than 0.02 μS and had a minimum distance of 0.5 s from each other (Green et al., 2014). Then, both SCL and number of peaks resulting for each interval were averaged across trials of the same condition for each participant (i.e. negative high salience, negative low salience, positive high salience, positive low salience).

Statistical analyses

Statistical analyses and figures were performed with Jamovi (Version 2.2)[65]. Normality of data distribution was ensured using Q-Q plots and Skewness parameters. The statistical analyses were conducted following a Fisherian approach. Mixed-design analyses of variance (ANOVA) were used to investigate differences between more than two conditions followed by post hoc tests with Holm's correction, wherever appropriate. A statistical significance threshold of $p < 0.05$ was adopted. Partial eta-squared (η^2_p) were computed as estimates of effect sizes for the ANOVAs' main effects and interactions, while Cohens' d for the post hoc tests[66].

Differences in ratings between groups were analyzed with a series of 2 (valence: negative, positive) x 2 (salience: low, high) x 2 (group: VMPFC, healthy) mixed-design ANOVAs. Differences in physiological response were analyzed with a series of 2 (Valence: negative, positive) x2 (Salience: high, low) x3 (Time: 0to10, 10to20, 20to30) x2 (Group: VMPFC, healthy) mixed-design ANOVAs.

Additionally, Pearson correlations (one-tailed, uncorrected) were computed to further assess the relationship between participants' emotional response during the contemplation of scenarios and their subsequent intentionality and morality ratings. These analyses and the related figures were performed with MATLAB.

Results

Intentionality rating

Results on intentionality ratings (Fig. 2a) showed a significant main effect of *valence* ($F_{1,32}$= 16.73, $p < 0.001$, η^2_p= 0.34), indicating higher intentionality ratings for negative scenarios (Mean (M) = 7.84, standard deviation (SD) = 1.65) compared to positive scenarios (M = 6.22, SD = 2.1).

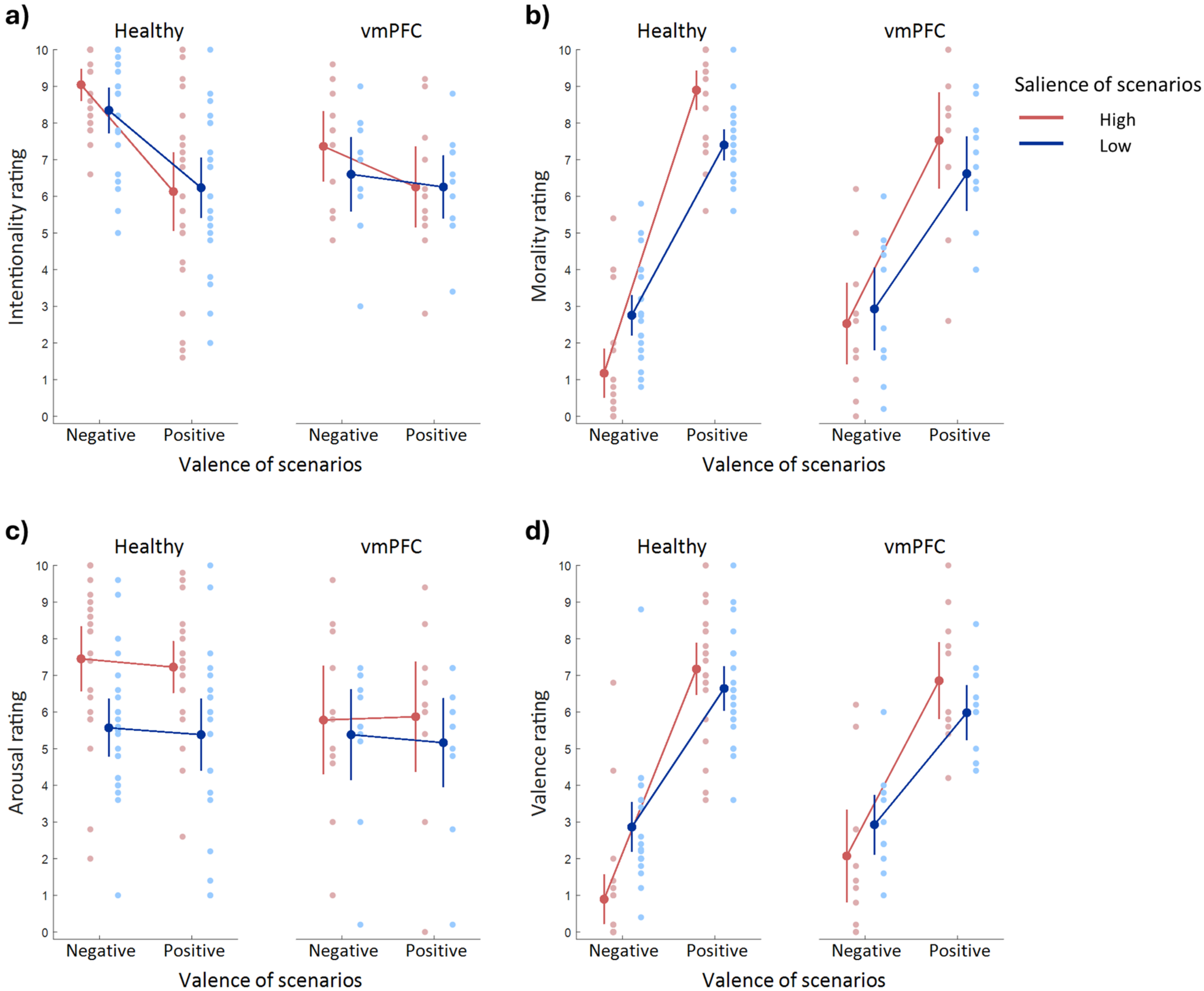

Fig. 2. Ratings of healthy participants and vmPFC patients in high (red) and low (blue) salience scenarios with negative and positive valence. Light red and light blue dots indicate single subjects' values, dark dots represent means, and vertical bars represent their standard error. **(a)** The intentionality rating plot shows the presence of the Knobe Effect; higher intentionality for negative than positive scenarios, only in healthy participants and not in vmPFC patients emerged. **(b)** The morality rating plot shows greater morality attributed to positive than negative scenarios, with a salience-driven polarization, for high salience scenarios, in healthy participants and not in vmPFC patients. **(c)** The arousal rating plot shows higher arousal attributed to high than low salience scenarios only in healthy participants and not in vmPFC participants. **(d)** The valence rating plot shows higher valence attributed to negative than positive scenarios, and polarization of this effect due to the high salience, in both groups.

Crucially, the *valence by group* interaction was significant ($F_{1,32}= 5.08$, $p=0.031$, $\eta^2_p = 0.14$). This interaction shows that only the healthy participants presented the Knobe Effect. Indeed, healthy participants attributed higher intentionality to negative than positive side effects (negative: M = 8.69, SD = 1.34; positive: M = 6.18, SD = 2.31, $p<0.001$, Cohen's d = 1.38), while vmPFC patients did not show any evidence of such difference (negative: M = 6.98, SD = 1.67; positive: M = 6.25, SD = 1.63, $p=0.55$, Cohen's d = 0.40). Additionally, the absence of Knobe Effect in the vmPFC patients appeared to be related to lower intentionality attributed to the negative side effects, relative to healthy controls ($p=0.003$, Cohen's d = 0.94). In contrast, intentionality attribution to positive scenarios did not differ between vmPFC patients and healthy controls ($p=0.92$, Cohen's d = 0.04). There was no other significant effect in the ANOVA (all $p>0.07$). In sum, compared to controls, patients with vmPFC lesions showed no evidence of a Knobe Effect, due to a diminished attribution of intentionality to scenarios with negative side effects.

Morality rating

Results on morality ratings (Fig. 2b) showed a significant main effect of *valence* ($F_{1,32}$= 147.44, $p<0.001$, η^2_p = 0.82), with positive scenarios attaining greater morality ratings than negative scenarios (positive: M = 7.61, SD = 1.67; negative: M = 2.35, SD = 1.76).

There was also a significant *valence by group* interaction ($F_{1,32}$= 4.49, $p=0.042$, η^2_p = 0.123), with similar morality ratings between vmPFC patients and healthy controls for negative scenarios (vmPFC: M = 2.73, SD = 1.86; healthy: M = 1.96, SD = 1.68; $p=0.16$, Cohen's d = 0.49), and a tendency for positive scenarios to be rated as less morally permissible by vmPFC patients compared to the controls (vmPFC: M = 7.07, SD = 1.99; healthy: M = 8.15, SD = 1.39; corrected $p=0.074$, uncorrected $p=0.037$, Cohen's d = 0.69), suggesting vmPFC patients differentiated less than controls between negative and positive scenarios. To further clarify this result, we computed the differences of morality ratings between positive and negative scenarios for each participant and then performed a t-test to compare such difference score between groups. Results confirmed a significantly reduced difference in ratings of positive and negative side effects (i.e. more flattened ratings) in vmPFC patients relative to controls (vmPFC: M = 4.35, SD = 3.36; controls: M = 6.18, SD = 1.74; $t_{1,32}$ = -2.12, $p=0.042$, Cohen's d = 0.78).

There was also a *valence by salience* significant effect ($F_{1,32}$= 31.27, $p<0.001$, η^2_p = 0.494), indicating that salience polarized moral rating. Indeed, high salience negative scenarios attained lower morality ratings than low salience negative scenarios (high salience: M = 1.61, SD = 1.8; low salience: M = 2.81, SD = 1.53; $p=0.001$, Cohen's d = 0.64), whereas high salience positive scenarios attained higher morality ratings than low-salience positive scenarios (high salience: M = 8.45, SD = 1.75; low salience: M = 7.15, SD = 1.32; $p<0.001$, Cohen's d = 0.77).

This finding was qualified by a significant *valence by salience by group* interaction ($F_{1,32}$= 5.08, $p=0.031$, η^2_p = 0.137), showing that this salience-driven polarization of moral rating was observed in controls but not in vmPFC patients. Thus, in healthy participants, for negative scenarios, morality was rated lower when salience was high than low (high salience: M = 1.17, SD = 1.62; low salience: M = 2.75, SD = 1.36; $p<0.001$, Cohen's d = 1.02), while for positive scenarios, morality was rated higher when salience was high than low (high salience: M = 8.9, SD = 1.3; low salience: M = 7.4, SD = 1.03; $p<0.001$, Cohen's d = 0.96). In contrast, salience did not affect morality ratings in vmPFC patients, either between high and low salience negative scenarios (high salience: M = 2.53, SD = 1.89; low salience: M = 2.93, SD = 1.9; $p=1.000$, Cohen's d = 0.26) or between high and low salience positive scenarios (high salience: M = 7.53, SD = 2.22; low salience: M = 6.62, SD = 1.72; $p=0.141$, Cohen's d = 0.59). There was no other significant effect in the ANOVA (all $p>0.341$).

In sum, patients with vmPFC lesions rated negative side effects as less morally permissible than positive side effects, albeit the difference in ratings between negative and positive side effects was flattened compared to controls. This occurred due to a failure of vmPFC patients to adjust their moral judgments based on the salience of the scenarios, as though they were unable to discern the difference between scenarios involving severe versus mild moral violations.

Valence rating

Results on valence ratings (Fig. 2d) showed a significant main effect of valence ($F_{1,32}$= 147.28, $p<0.001$, η^2_p = 0.822), with higher valence ratings (rated from negative to positive) for positive than negative scenarios (positive: M = 6.66, SD = 1.55; negative: M = 2.19, SD = 1.82).

Also, the *valence by salience* interaction was significant ($F_{1,32}$= 19.25, $p<0.001$, η^2_p = 0.376). This interaction shows that salience polarized valence ratings, such that for negative scenarios, valence was rated lower when salience was high than low (high salience: M = 1.28, SD = 1.87; low salience: M = 2.88, SD = 1.55; $p=0.001$, Cohen's d = 0.86), while for positive scenarios, valence was rated higher when salience was high than low (high salience: M = 7.07, SD = 1.73; low salience: M = 6.43, SD = 1.43; $p=0.012$, Cohen's d = 0.43). There was no other significant effect in the ANOVA (all $p>0.09$).

Emotional response

Subjective arousal rating

Results on arousal ratings (Fig. 2c) showed a significant main effect of salience ($F_{1,32}$= 21.95, $p<0.001$, η^2_p = 0.41), qualified by a *salience by group* interaction ($F_{1,32}$= 6.43, $p=0.016$, η^2_p = 0.17). This interaction shows that healthy controls rated high salience scenarios as more arousing than low salience scenarios (high salience: M = 7.34, SD = 1.93; low salience: M = 5.48, SD = 2.16; $p<0.001$, Cohen's d = 0.87), in contrast, vmPFC patients did not show such modulation (high salience: M = 5.83, SD = 2.47; low salience: M = 5.27, SD = 2.02; $p=0.602$, Cohen's d = 0.26). There was no other significant effect in the ANOVA (all $p>0.514$).

Physiological arousal

Results on SCL (Fig. 3a) showed a significant main effect of *Time* ($F_{2,62}$= 7.50, $p=0.001$, η^2_p = 0.20), indicating a higher SCL in the first 10 s of the trials (M = 0.12, SD = 0.03) compared to the following 10 s (M = 0.01, SD = 0.42; $p=0.009$, Cohen's d = 0.27) and the last 10 s of the trials (M = -0.08, SD = 0.46; $p=0.012$, Cohen's d = 0.47), with no differences between the last two time intervals ($p=0.087$, Cohen's d = 0.2). There was no other significant effect (all $p>0.058$).

Results on the number of SCR peaks (Fig. 3b) showed a significant main effect of *Group* ($F_{1,31}$= 9.02, $p=0.005$, η^2_p = 0.225), indicating that vmPFC patients produced fewer peaks than healthy participants (vmPFC: M = 0.33, SD = 0.3; healthy: M = 0.61, SD = 0.36).

Also, the main effect of *Time* emerged as significant ($F_{2,62}$= 5.057, $p=0.009$, η^2_p = 0.14), indicating more peaks in the first 10 s of the trials (M = 0.63, SD = 0.39) compared to the following 10 s (M = 0.48, SD = 0.37, $p=0.037$, Cohen's d = 0.38) and the last 10 s of the trials (M = 0.5, SD = 0.36, $p=0.05$, Cohen's d = 0.38), with no differences between the last two time intervals (all $p>0.161$).

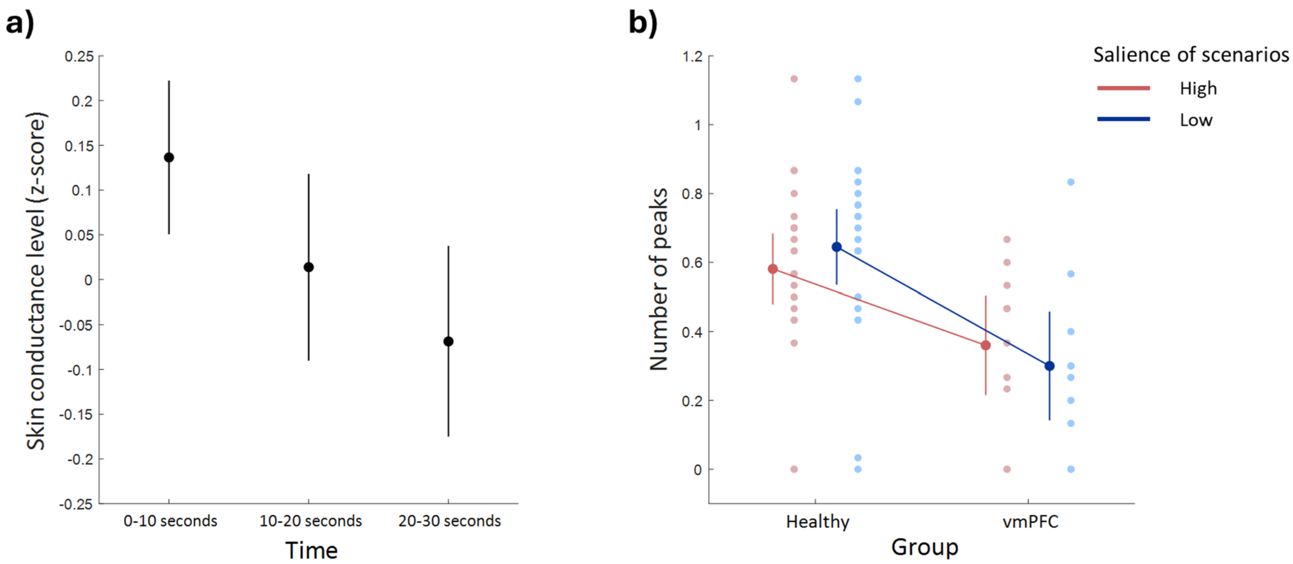

Fig. 3. **(a)** Skin Conductance Level (SCL) of participants showing higher SCL during the first 10 s of the scenarios' presentation compared to the following time intervals. Dots represent means and vertical bars represent their standard error; **(b)** Number of peaks in electrodermal activity showing that healthy participants had a higher number of peaks than vmPFC patients, especially for low salience scenarios. Light red and light blue dots indicate single subjects' values, dark dots represent means, and vertical bars represent their standard error.

Finally, the *salience by group* interaction was significant ($F_{1,31}$ = 8.196, p = 0.007, η^2_p = 0.209). This interaction shows that in the low salience scenarios vmPFC patients produced fewer peaks than healthy participants (vmPFC: M = 0.3, SD = 0.3; healthy: M = 0.66, SD = 0.38; p = 0.01, Cohen's d = 1.10), while in the high salience scenarios vmPFC patients and healthy participants did not differ significantly (vmPFC: M = 0.36, SD = 0.29; healthy: M = 0.58, SD = 0.35, p = 0.14, Cohen's d = 0.64).

Exploratory correlations between scenarios' ratings and emotional response

To further understand the relationship between vmPFC patients' deficits in intentionality and morality ratings and their emotional response, a series of linear correlations was conducted between the intentionality and morality ratings previously found altered in vmPFC patients and the subjective arousal ratings and number of SCR peaks. Greater subjective arousal and more SCR peaks were expected to correlate with greater intentionality ratings in negative scenarios, especially of high-salience. Conversely, the opposite pattern was anticipated for morality ratings, with greater subjective arousal and more SCR peaks associated with lower morality ratings in negative scenarios.

Subjective arousal and intentionality

Results, illustrated in Fig. 4, showed a significant positive correlation between subjective arousal and intentionality ratings for high-salience negative scenarios (r = 0.405, p = 0.009). In contrast, a significant negative correlation was found for high-salience positive scenarios (r = -0.310, p = 0.037). No significant correlations emerged for low-salience scenarios (all $p \geq 0.366$). These results suggest that, in high-salience negative scenarios, participants who experienced greater arousal while reading the scenarios attributed higher intentionality to the side effect. Conversely, in high-salience positive scenarios, those who experienced greater arousal attributed lower intentionality to the side effect.

Subjective arousal and morality

Results, illustrated in Fig. 5, showed a significant negative correlation between subjective arousal and morality ratings for negative scenarios with high salience (r = -0.446, p = 0.004). No other significant correlations emerged (all p > = 0.055). These results suggest that, in high-salience negative scenarios, participants who experienced greater arousal while reading the scenarios judged the side effect as less morally permissible.

Number of peaks and intentionality

Results, illustrated in Fig. 6, showed a significant positive correlation between the number of SCR peaks and intentionality ratings for high-salience negative scenarios (r = 0.293, p = 0.049). Conversely, a significant negative correlation was found for high-salience positive scenarios (r = -0.300, p = 0.045). No significant correlations emerged for low-salience scenarios (all $p \geq 0.133$). These results suggest that, in high-salience negative scenarios, participants who exhibited more SCR peaks while reading the scenarios attributed higher intentionality to the side effect. In contrast, in high-salience positive scenarios, those with more SCR peaks attributed lower intentionality to the side effect.

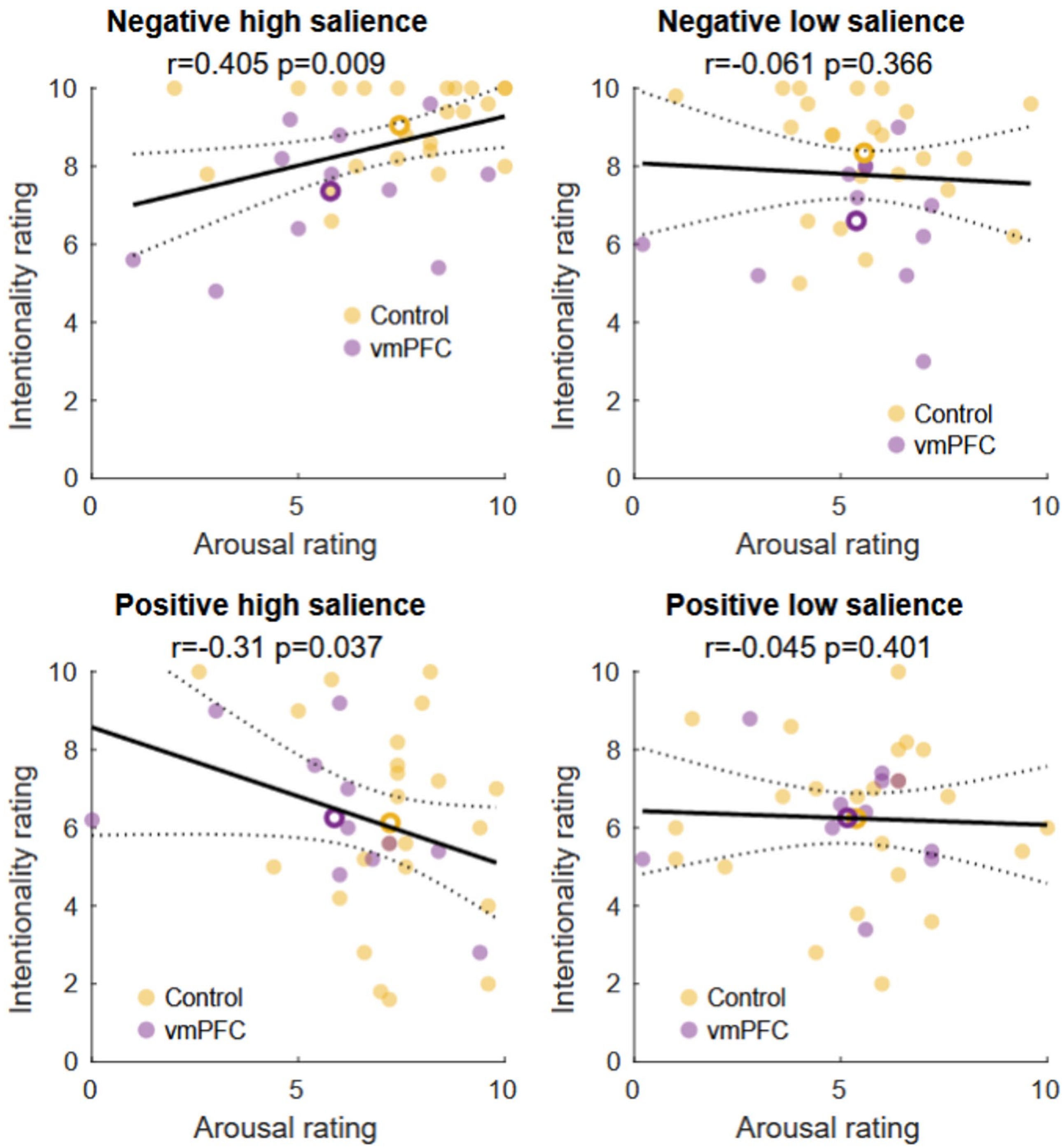

Fig. 4. Correlations between subjective arousal ratings and intentionality ratings in the different types of scenarios. Solid line represents the correlation line, dotted lines represent 95% confidence interval (CI). Filled circles represent individual participants' scores and empty circles represent group means, for the vmPFC and control groups.

Number of peaks and morality
Results, illustrated in Fig. 7, showed a significant negative correlation between the number of SCR peaks and morality ratings for high-salience negative scenarios ($r = -0.364$, $p = 0.019$). No other significant correlations emerged (all $p \geq 0.179$). These results suggest that, in high-salience negative scenarios, participants who exhibited more SCR peaks while reading the scenarios judged the side effect as less morally permissible.

Discussion
This study advances the understanding of the neural bases of intentionality attribution and the Knobe Effect. A group of participants with vmPFC lesions and a group of healthy controls rated the intentionality of a series of

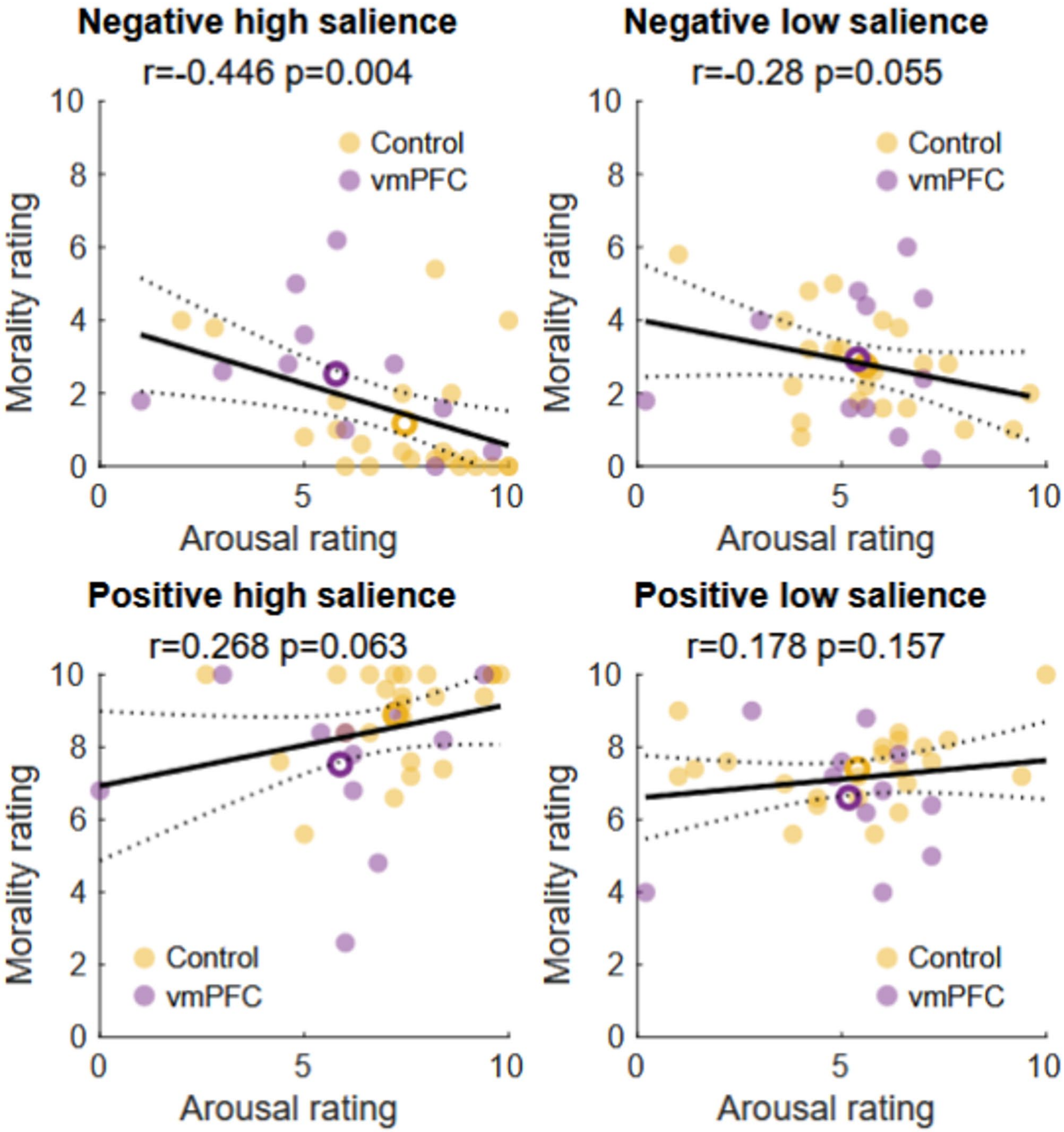

Fig. 5. Correlations between subjective arousal ratings and morality ratings in the different types of scenarios. Solid line represents the correlation line, dotted lines represent 95% confidence interval (CI). Filled circles represent individual participants' scores and empty circles represent group means, for the vmPFC and control groups.

scenarios involving side effects with varying valence (positive or negative) and salience (high or low). To test the contribution of moral judgement and emotional response to intentionality attribution, participants also rated the morality of the side effects, the subjective arousal experienced while reading the scenarios, and electrodermal activity was recorded as a measure of physiological arousal. Finally, participants rated the valence of scenarios as a control measure of their ability to discriminate between negative and positive scenarios. Our findings highlight the role of the vmPFC in linking emotion, morality, and intentionality attribution, revealing that this brain region plays a causal role in the attribution of intentionality.

In line with previous findings[8–15], healthy controls exhibited a consistent Knobe Effect, attributing greater intentionality to actions associated with negative, as opposed to positive, side effects. This result supports the idea that intentionality judgement is not a mere reflection of the actual agent's intentions, but that it is closely

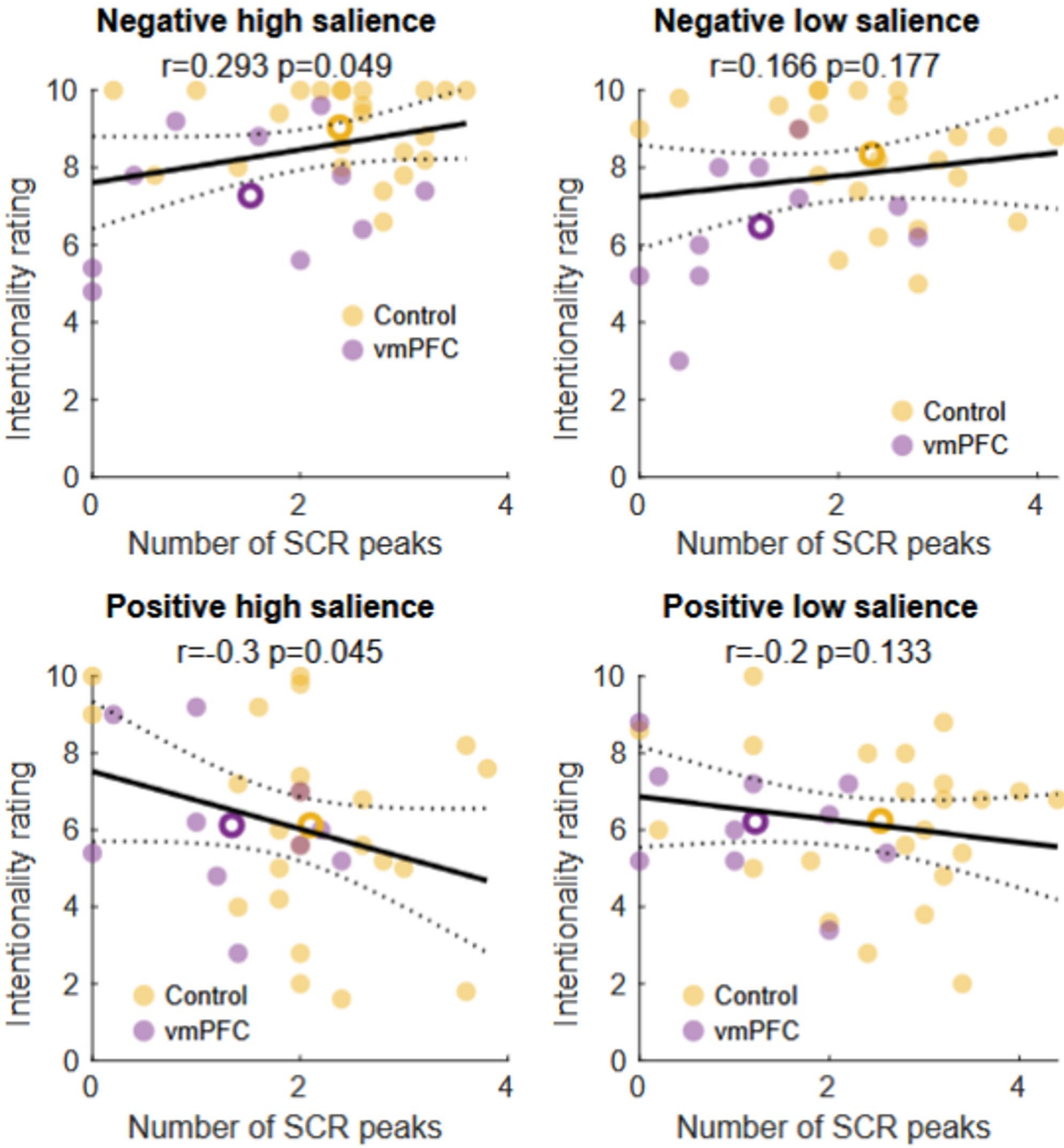

Fig. 6. Correlations between the number of SCR peaks and intentionality ratings in the different types of scenarios. Solid line represents the correlation line, dotted lines represent 95% confidence interval (CI). Filled circles represent individual participants' scores and empty circles represent group means, for the vmPFC and control groups.

tied to the negative valence of (unintentional) outcomes of other's actions, with adverse outcomes being assessed differently and more harshly (here as more intentional) than positive ones[9,15,21–23].

Crucially, patients with vmPFC lesions showed no evidence of a Knobe Effect, attributing similar levels of intentionality to scenarios with positive or negative side effects. Importantly, this result emerged from a reduced attribution of intentionality to scenarios with negative side effects specifically. This finding suggests that vmPFC patients, unlike controls, are less influenced by the valence of the outcomes resulting from others' actions and instead attribute similar levels of intentionality to both positive and negative side effects. The absence of the Knobe Effect in vmPFC patients highlights the vmPFC's causal role in intentionality attribution.

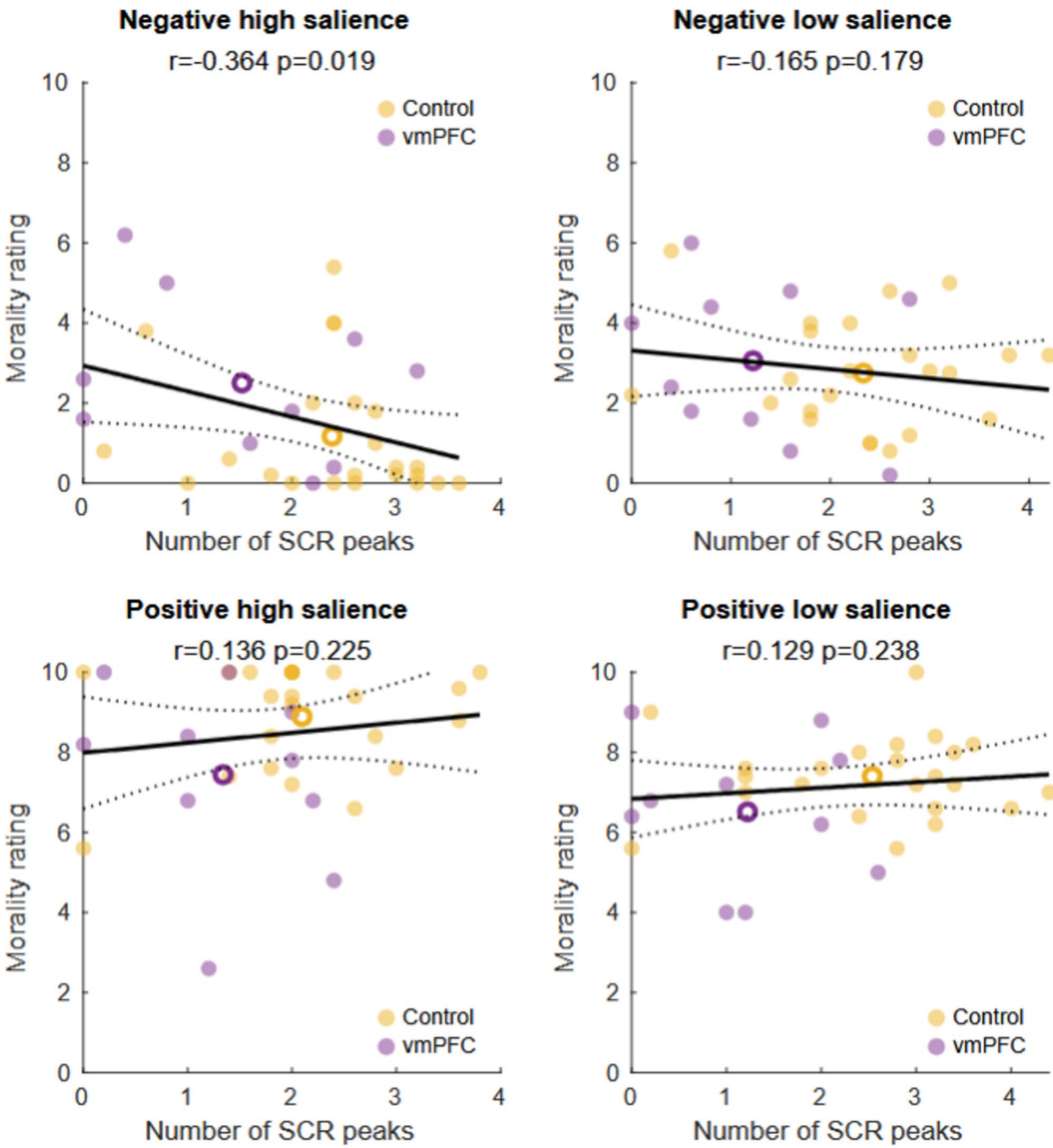

Fig. 7. Correlations between the number of SCR peaks and morality ratings in the different types of scenarios. Solid line represents the correlation line, dotted lines represent 95% confidence interval (CI). Filled circles represent individual participants' scores and empty circles represent group means, for the vmPFC and control groups.

Notably, the vmPFC group's failure to inflate intentionality attribution for negative side effects occurred despite their preserved recognition of scenarios' valence. Similarly to controls, vmPFC patients rated scenarios with positive side effects as more pleasant than those with negative effects and exhibited an even greater polarization in their ratings for high-salience scenarios. These findings suggest that valence recognition alone does not account for the Knobe Effect.

Rather, our results point to emotional response as the key driver of the Knobe Effect. Indeed, vmPFC patients exhibited a blunted emotional response in both its subjective and physiological arousal components. Unlike healthy controls, who self-reported having experienced more arousal for high-salience than low-salience scenarios, vmPFC patients showed no such modulation. Additionally, despite preserved tonic modulation of skin conductance levels, vmPFC patients had a reduced phasic modulation of skin conductance, producing fewer peaks than controls while reading the scenarios. These findings align with robust evidence that vmPFC lesions result in emotion processing deficits, and indeed, the vmPFC plays a pivotal role in emotional perception[37–39], as well as response expression and regulation[40–45], especially when in anticipation of future outcomes[48,49,67].

The critical role of emotional responses in shaping intentionality attribution and moral judgement is further corroborated by the significant correlations between participants' subjective and physiological arousal experienced during scenarios contemplation and intentionality and moral ratings. Overall, participants who experienced greater subjective arousal while reading, salient, negative scenarios attributed more intentionality to the side effect, and judged it as less morally permissible. Conversely, greater subjective arousal was associated with lower intentionality attributions in positive scenarios. A similar pattern emerged when considering physiological arousal: participants who exhibited more physiological arousal attributed higher intentionality to the side effect of high-salience negative scenarios, and judged them as less morally permissible, while greater physiological arousal corresponded to lower intentionality attribution in positive scenarios. These exploratory correlations provide further insight into the mechanisms underlying the absence of the Knobe effect in the vmPFC group. Visual inspection of the correlation plots suggests that vmPFC patients contribute to the correlation between emotional response and judgments of intentionality and morality in high-salience scenarios. However, their blunted subjective and physiological emotional responses are associated with altered intentionality and morality judgments, accounting for the absence of the Knobe effect in this group.

Notably, previous literature has shown that moral judgement of blame mediates the relationship between emotional response and intentionality attribution for negative scenarios[16]. Building on this, the role of vmPFC in the Knobe Effect may reflect its critical contribution to integrating emotional response within moral judgment. Indeed, results showed that moral judgement was also disrupted in vmPFC patients. Specifically, although vmPFC patients rated negative side effects as less morally permissible than positive side effects, the difference in ratings between negative and positive side effects was flattened in vmPFC patients compared to controls. This occurred due to a failure of vmPFC patients to adjust their moral judgments based on the salience of the scenarios, as if they were unable to appraise the difference between scenarios involving severe versus mild moral violations. This finding aligns with previous evidence showing that vmPFC patients judge personal moral violations as more acceptable behaviors compared to healthy individuals[47,50,51]. Notably, vmPFC patients' performance is instead comparable to the controls' with impersonal moral violations, which involve moral emotions to a lesser extent, suggesting that impaired moral judgment in vmPFC patients may indeed originate from a deficit in emotion processing. Consistent with this hypothesis, Moretto et al.[51] showed that whereas in healthy controls the endorsement of personal moral violations was associated with an increase in phasic skin conductance response, this was not observed in vmPFC patients, who failed to exhibit such a response and endorsed more personal moral violations.

Relatedly, in the context of moral judgement, vmPFC patients are also known to exhibit an impairment in appraising the intentions underlying morally relevant behaviors, again particularly when these intentions are negative[52] and would normally elicit a stronger emotional response than when positive. In the present study, vmPFC patients were paradoxically more accurate in attributing (lack of) intentions to the characters in the story. However, the physiological and self-reported results of emotional arousal show that this spared capability actually betrays a deficit in emotional response that typically guides moral judgment and the inflation of intentionality attribution to negative scenarios.

The altered intentionality attribution and moral judgments observed in vmPFC patients, accompanied by blunted subjective and physiological emotional responses, may also reflect a disruption in interoceptive processing. Indeed, recent research highlights that interoceptive awareness shapes moral emotions[68] and decision-making[69–71], with the vmPFC playing a central role in processing interoceptive signals and supporting emotional awareness[72,73], along with the anterior insula (see also the additional analyses in the supplementary material) and precuneus. Thus, our findings also align with the growing evidence linking interoceptive processes to moral cognition and extend this connection to intentionality attribution.

Conceptually similar findings to those reported here in vmPFC patients have been observed in individuals with high levels of alexithymia, a personality trait characterized by difficulties in emotional processing[27–32]. These individuals also exhibit a reduced Knobe Effect, due to decreased attribution of intentionality to negative side effects, which was accompanied by reduced physiological and subjective emotional response[33]. However, while vmPFC patients and individuals with high levels of alexithymia both demonstrated generally reduced physiological responses, the nature of their subjective response deficits differed. In individuals with high alexithymia, the reduction in subjective emotional response was limited to negative scenarios. In contrast, vmPFC patients exhibited a more pervasive deficit, with diminished subjective emotional responses observed regardless of scenarios' valence. Despite these differences, both groups exhibit a reduced Knobe Effect due to decreased intentionality ratings to the negative side effects, pointing to the crucial role that emotional responses play to the attribution of intentionality for side effects of negative valence specifically.

While our findings offer novel insights into the role of the vmPFC in intentionality attribution, certain limitations of the present study should be considered. The relatively small sample size of the vmPFC group and correlational results should be corroborated in future larger-scale studies. These would enable to explore individual differences in emotional and interoceptive processing and more precisely delineate the neurocognitive mechanisms through which vmPFC damage disrupts moral and intentionality judgement. Additionally, future work may employ a broader battery of Theory of Mind and moral judgment tasks to clarify the interplay between emotional blunting and potential impairments in mental state attribution in giving rise to the alterations in the Knobe effect observed in vmPFC patients.

In conclusion, our study shows that lesions to the vmPFC blunt subjective and physiological emotional response, disrupt moral judgment and abolish the misattribution of intentionality to unintentional negative outcomes of others' actions. Notably, such results were evident despite patients' preserved discrimination of side effects' negative or positive valence, suggesting that the misattribution of intentionality core to the Knobe Effect is not solely dependent on emotional valence recognition. Instead, our results support the view that the negativity of the outcomes must elicit an increase in emotional response and moral blame to then inflate intentionality judgement. Overall, our results show that emotional responses and moral judgment are fundamental to the Knobe Effect in healthy adults and highlight the critical role of the vmPFC in driving this ubiquitous human bias.

Data availability

All data, analysis code, and research materials will be made available in a timely manner to other researchers upon request. Interested researchers should contact the corresponding author.

Received: 10 February 2025; Accepted: 20 May 2025
Published online: 01 July 2025

References

1. Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A. & Saxe, R. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc. Natl. Acad. Sci. U S A.* **107**, 6753–6758 (2010).
2. Ciaramelli, E., Braghittoni, D. & Di Pellegrino, G. It is the outcome that counts! Damage to the ventromedial prefrontal cortex disrupts the integration of outcome and belief information for moral judgment. *J. Int. Neuropsychol. Soc.* **18**, 962–971 (2012).
3. Baez, S. et al. Comparing moral judgments of patients with frontotemporal dementia and frontal stroke. *JAMA Neurol.* **71**, 1172–1176 (2014).
4. Sarfati, Y., Hardy-Baylé, M. C., Besche, C. & Widlöcher, D. Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strips. *Schizophr Res.* **25**, 199–209 (1997).
5. Brüne, M. Theory of Mind in schizophrenia: A review of the literature. *Schizophr Bull.* **31**, 21–42 (2005).
6. Fletcher, P. C. & Frith, C. D. Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* **10**, 48–58 (2008).
7. Georges, L. C., Wiener, R. L. & Keller, S. R. The angry juror: Sentencing decisions in first-degree murder. *Appl. Cogn. Psychol.* **27**(2), 156–166 (2013).
8. Knobe, J. Intentional action and side effects in ordinary Language. *Analysis* **63**, 190–194 (2003).
9. Knobe, J. Intentional action in folk psychology: an experimental investigation. *Philos. Psychol.* **16**, 309–324 (2003).
10. Nadelhoffer, T. Blame, badness, and intentional action: A Rseply to Knobe and Mendlow. *J. Theoretical Philosophical Psychol.* **24**, 259–269 (2004).
11. Nadelhoffer, T. Bad acts, blameworthy agents, and intentional actions: some problems for juror impartiality. *Philosophical Explorations*. **9**, 203–219 (2006).
12. Pellizzoni, S., Siegal, M., Surian, L. & Foreknowledge, Caring, and the Side-Effect effect in young children. *Dev. Psychol.* **45**, 289–295 (2009).
13. Knobe, J. Folk psychology and folk morality: response to critics. *J. Theoretical Philosophical Psychol.* **24**, 270–279 (2004).
14. Knobe, J. & Intention, Intentional action and moral considerations. *Analysis* **64**, 181–187 (2004). https://www.jstor.org/stable/3329125
15. Nadelhoffer, T. On praise, side effects, and folk ascriptions of intentionality. *J. Theoretical Philosophical Psychol.* **24**, 196–213 (2004).
16. Ngo, L. et al. Two Distinct moral mechanisms for ascribing and denying intentionality. *Sci. Rep.* **5**, 1–11 (2015).
17. Mele, A. & Cushman, F. Intentional action, folk judgments, and stories: Sorting things out. *Midwest Stud. Philos. (Wiley-Blackwell)* **31** (2007).
18. Knobe, J. & Burra, A. The folk concepts of intention and intentional action: A Cross-Cultural study. *J. Cogn. Cult.* **6**, 113–132 (2006).
19. Leslie, A. M., Knobe, J. & Cohen, A. Acting intentionally and the side-effect effect theory of Mind and moral judgment. *Psychol. Sci.* **17**, 421–427 (2006).
20. Malle, B. F. & Knobe, J. The folk concept of intentionality. *J. Exp. Soc. Psychol.* **33**, 101–121 (1997).
21. Alicke, M. D. & Rose, D. Culpable control and causal deviance. *Soc. Personal Psychol. Compass*. **6**, 723–735 (2012).
22. Malle, B. F. & Nelson, S. E. Judging mens Rea: the tension between folk concepts and legal concepts of intentionality. *Behav. Sci. Law.* **21**, 563–580 (2003).
23. Knobe, J. & Mendlow, G. S. The good, the bad and the blameworthy: Understanding the role of evaluative reasoning in folk psychology. *J. Theoretical Philosophical Psychol.* **24**, 252–258 (2004).
24. Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* **372**, 669–672 (1994).
25. Bertini, C. et al. Fear-specific enhancement of tactile perception is disrupted after amygdala lesion. *J. Neuropsychol.* **14**, 165–182 (2020).
26. LaBar, K. S., LeDoux, J. E., Spencer, D. D. & Phelps, E. A. Impaired fear conditioning following unilateral Temporal lobectomy in humans. *J. Neurosci.* **15**, 6846–6855 (1995).
27. Starita, F. & di Pellegrino, G. Alexithymia and the reduced ability to represent the value of aversively motivated actions. *Front. Psychol.* **9**, 1–11 (2018).
28. Starita, F., Làdavas, E. & di Pellegrino, G. Reduced anticipation of negative emotional events in alexithymia. *Sci. Rep.* **6**, 27664 (2016).
29. Starita, F., Borhani, K., Bertini, C. & Scarpazza, C. Alexithymia is related to the need for more emotional intensity to identify static fearful facial expressions. *Front. Psychol.* **9** (2018).
30. Starita, F., Pietrelli, M., Bertini, C. & di Pellegrino, G. Aberrant reward prediction error during Pavlovian appetitive learning in alexithymia. *Soc. Cogn. Affect. Neurosci.* **14**, 1119–1129 (2019).
31. Sifneos, P. E. The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychother. Psychosom.* **22**, 255–262 (1973).
32. Taylor, G. J., Bagby, M., Parker, J. D. A. & R. & The alexithymia construct: A potential paradigm for psychosomatic medicine. *Psychosomatics* **32**, 153–164 (1991).
33. Zucchelli, M. M., Starita, F., Bertini, C., Giusberti, F. & Ciaramelli, E. Intentionality attribution and emotion: the Knobe effect in alexithymia. *Cognition* **191**, 103978 (2019).
34. Cardinale, E. M. et al. Influences its perceived intentional status in adolescents with psychopathic traits. *Oxf. Stud. Exp. Philos.* **1** (2014).
35. Ciaramelli, E. & Di Pellegrino, G. Ventromedial prefrontal cortex and the future of morality. *Emot. Rev.* **3** (preprint) (2011). https://doi.org/10.1177/1754073911402381

36. McCormick, C., Ciaramelli, E., De Luca, F. & Maguire, E. A. Comparing and contrasting the cognitive effects of hippocampal and ventromedial prefrontal cortex damage: A review of human lesion studies. *Neuroscience* **374**, 295–318 (2018).
37. Heberlein, A. S., Padon, A. A., Gillihan, S. J., Farah, M. J. & Fellows, L. K. Ventromedial frontal lobe plays a critical role in facial emotion recognition. *J Cogn. Neurosci* **20** (2008).
38. Tsuchida, A. & Fellows, L. K. Are you upset? Distinct roles for orbitofrontal and lateral prefrontal cortex in detecting and distinguishing facial expressions of emotion. *Cereb. Cortex*. **22**, 2904–2912 (2012).
39. Wolf, R. C., Philippi, C. L., Motzkin, J. C., Baskaya, M. K. & Koenigs, M. Ventromedial prefrontal cortex mediates visual attention during facial emotion recognition. *Brain* **137** (2014).
40. Anderson, R. J. Industrial firm linkages in a post-Soviet urban economy: implications for development policy and programmes. *Prog. Dev. Stud.* **6**(3), 224–241 (2006).
41. Bechara, A. The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain Cogn.* **55** (2004).
42. Quirk, G. J. & Beer, J. S. Prefrontal involvement in the regulation of emotion: Convergence of rat and human studies. *Curr. Opin. Neurobiol.* **16** (preprint) (2006). https://doi.org/10.1016/j.conb.2006.07.004
43. Berlin, H. A., Rolls, E. T. & Kischka, U. Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain* **127** (preprint) (2004). https://doi.org/10.1093/brain/awh135
44. Krueger, F. et al. The neural bases of key competencies of emotional intelligence. *Proc Natl. Acad. Sci. U S A* **106** (2009).
45. Jonker, F. A., Jonker, C., Scheltens, P. & Scherder, E. J. A. The role of the orbitofrontal cortex in cognition and behavior. *Rev. Neurosci.* **26** (2015).
46. Koenigs, M. & Tranel, D. Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *J. Neurosci.* https://doi.org/10.1523/JNEUROSCI.4606-06.2007 (2007).
47. Koenigs, M. et al. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* **446**, 908–911 (2007).
48. Bechara, A., Tranel, D., Damasio, H. & Damasio, A. R. Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb. Cortex*. **6**, 215–225 (1996).
49. Battaglia, S., Garofalo, S., di Pellegrino, G. & Starita, F. Revaluing the role of VmPFC in the acquisition of Pavlovian threat conditioning in humans. *J. Neurosci.* **40**, 8491–8500 (2020).
50. Ciaramelli, E., Muccioli, M., Làdavas, E. & Di Pellegrino, G. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Soc. Cogn. Affect. Neurosci.* **2** (2007).
51. Moretto, G., Làdavas, E. & Mattioli, F. & Di Pellegrino, G. A Psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *J. Cogn. Neurosci.* **22** (2010).
52. Young, L. et al. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron* **65**, 845–851 (2010).
53. Young, L., Cushman, F., Adolphs, R., Tranel, D. & Hauser, M. Does emotion mediate the relationship between an action's moral status and its intentional status? Neuropsychological evidence. *J. Cogn. Cult.* **6**, 291–304 (2006).
54. Holmes, C. J. et al. Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr*. **22**, 324–333 (1998).
55. Rorden, C. & Brett, M. Stereotaxic display of brain lesions. *Behav. Neurol.* **12**, 191–200 (2000).
56. Spinnler, H. & Tognoni, G. Standardizzazione e Taratura Italiana Di test neuropsicologici. Gruppo Italiano per lo studio neuropsicologico Dell'Invecchiamento. *Ital. J. Neurol. Sci* **1–120** (1987).
57. Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F. & Venneri, A. A short version of the Stroop test: Normative data in an Italian population sample. *Nuova Riv Neurol*. (2002).
58. Orsini, A. et al. Verbal and Spatial immediate memory span: normative data from 1355 adults and 1112 children. *Italian J. Neurol. Sci.* https://doi.org/10.1007/BF02333660 (1987).
59. Novelli, G. et al. Tre test clinici Di memoria verbale a Lungo termine. Taratura Su Soggetti normali. *Arch. Psicol. Neurol. Psichiatr.* **47**, 278–296 (1986).
60. Capitani, E. & Laiacona, M. Aging and psychometric diagnosis of intellectual impairment: some considerations on test scores and their use. *Dev. Neuropsychol.* **4**, 325–330 (1988).
61. Schneider, W., Eschman, A. & Zuccolotto, A. *E-Prime (Version 2.0)*. Preprint (2002).
62. Mathôt, S., Schreij, D., Theeuwes, J. & OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods*. **44**, 314 (2012).
63. Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. Publication recommendations for electrodermal measurements. *Psychophysiology* **49**, 1017–1034 (2012).
64. Dawson, M. E., Schell, A. M. & Filion, D. L. The electrodermal system. *Handb. Psychophysiol. Fourth Ed*. 217–243. https://doi.org/10.1017/9781107415782.010 (2016).
65. Caldwell, A. R. & SimplyAgree: An R package and jamovi module for simplifying agreement and reliability analyses. *J. Open. Source Softw.* **7**, 4148 (2022).
66. Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front. Psychol.* **4**, 863 (2013).
67. Motzkin, J. C., Philippi, C. L., Wolf, R. C., Baskaya, M. K. & Koenigs, M. Ventromedial prefrontal cortex lesions alter neural and physiological correlates of anticipation. *J. Neurosci.* **34** (2014).
68. Tamura, K., Kobayashi, Y. & Ohira, H. Interoceptive accuracy did not affect moral decision-making, but affect regret rating for one's moral choices. *Front. Psychol.* **12** (2022).
69. Vabba, A., Porciello, G., Panasiti, M. S. & Aglioti, S. M. Interoceptive influences on the production of self-serving Lies in reputation risk conditions. *Int. J. Psychophysiol.* **177** (2022).
70. Feruglio, S., Panasiti, M. S., Crescentini, C., Aglioti, S. M. & Ponsi, G. Training the moral self: An 8-week mindfulness meditation program leads to reduced dishonest behavior and increased regulation of interoceptive awareness. *Mindfulness (N Y)* **14** (2023).
71. Cui, S. & Nakano, T. Interoceptive brain processing influences moral decision making. *Hum. Brain Mapp.* **45**, e70108 (2024).
72. Terasawa, Y., Fukushima, H. & Umeda, S. How does interoceptive awareness interact with the subjective experience of emotion? An fMRI study. *Hum. Brain Mapp.* **34** (2013).
73. Roy, M., Shohamy, D. & Wager, T. D. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn. Sci.* **16**, 147–156 (2012).

Acknowledgements
The authors thank Sara Brini for her help with data collection. The authors declare no competing interests. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions
F.S. and E.C. conceptualized the experiment. FS., L.A.E.D., D.D. performed the experiment. FS., L.A.E.D., D.D. analyzed the data. All authors contributed to the writing of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-03457-y.

Correspondence and requests for materials should be addressed to F.S.