Received: 22 April 2017 | Revised: 11 December 2017 | Accepted: 12 December 2017
DOI: 10.1002/hbm.23930

WILEY

RESEARCH ARTICLE

# Affective and cooperative social interactions modulate effective connectivity within and between the mirror and mentalizing systems

Maria Arioli[1,2] | Daniela Perani[2,3,4] | Stefano Cappa[1,5] | Alice Mado Proverbio[6] | Alberto Zani[7] | Andrea Falini[2,4,8] | Nicola Canessa[1]

[1]NEtS Center, Scuola Universitaria Superiore IUSS, Pavia 27100, Italy

[2]Division of Neuroscience, IRCCS San Raffaele, Milan 20132, Italy

[3]Nuclear Medicine Unit, Ospedale San Raffaele, Milan 20132, Italy

[4]Vita-Salute San Raffaele University, Milan 20132, Italy

[5]IRCCS Centro San Giovanni di Dio Fatebenefratelli, Brescia 25125, Italy

[6]Department of Psychology, University of Milano-Bicocca, NeuroMi - Milan Center for Neuroscience, Milan 20126, Italy

[7]Institute of Bioimaging and Molecular Physiology, IBFM-CNR, Milan 20125, Italy

[8]Neuroradiology Unit, IRCCS Ospedale San Raffaele, Milan 20132, Italy

**Correspondence**
Nicola Canessa, Scuola Universitaria Superiore IUSS, Piazza della Vittoria 15, 7100 Pavia, Italy.
Email: nicola.canessa@iusspavia.it

## Abstract

Decoding the meaning of others' actions, a crucial step for social cognition, involves different neural mechanisms. While the "mirror" and "mentalizing" systems have been associated with, respectively, the processing of biological actions versus more abstract information, their respective contribution to intention understanding is debated. Processing social interactions seems to recruit both neural systems, with a different weight depending on cues emphasizing either shared action goals or shared mental states. We have previously shown that observing cooperative and affective social interactions elicits stronger activity in key nodes of, respectively, the mirror (left posterior superior temporal sulcus (pSTS), superior parietal cortex (SPL), and ventral/dorsal premotor cortex (vPMC/dPMC)) and mentalizing (ventromedial prefrontal cortex (vmPFC)) systems. To unveil their causal organization, we investigated the effective connectivity underlying the observation of human social interactions expressing increasing cooperativity (involving left pSTS, SPL, and vPMC) versus affectivity (vmPFC) via dynamic causal modeling in 36 healthy human subjects. We found strong evidence for a model including the pSTS and vPMC as input nodes for the observed interactions. The extrinsic connectivity of this model undergoes oppositely valenced modulations, with cooperativity promoting positive modulations of connectivity between pSTS and both SPL (forward) and vPMC (mainly backward), and affectivity promoting reciprocal positive modulations of connectivity between pSTS and vmPFC (mainly backward). Alongside *f*MRI data, such divergent effective connectivity suggests that different dimensions underlying the processing of social interactions recruit distinct, although strongly interconnected, neural pathways associated with, respectively, the bottom–up visuomotor processing of motor intentions, and the top–down attribution of affective/mental states.

KEYWORDS
dynamic causal modeling, effective connectivity, intention understanding, mentalizing system, mirror neuron system, social cognition, social interaction

## 1 | INTRODUCTION

The neural mechanisms responsible for decoding others' intentions from their actions, a core component of social cognition, have been the focus of intensive investigation. Available evidence suggests that the concepts of action, goal, and intention can be ordered hierarchically according to their level of abstractness or the time required to complete them (Hamilton & Grafton, 2006; Van Overwalle & Baetens, 2009). The "mirror" and "mentalizing" neural systems seem to play complementary roles in processing such different aspects of intentions, based on input from the posterior superior temporal sulcus (pSTS) specialized for the perception of biological motion (Blakemore & Decety, 2001).

The mirror system, including the posterior inferior frontal gyrus (pIFG), dorsal and ventral premotor cortex (dPMC and vPMC), and inferior parietal lobule (IPL), superior parietal lobule (SPL), and intraparietal sulcus (IPS) (Buccino et al., 2004; Caspers, Zilles, Laird, & Eickhoff, 2010; Molenberghs, Cunnington, & Mattingley, 2012), underpins the automatic understanding of actions (e.g., the motor act of reaching/grasping a glass), immediate goals (taking the glass), and final goals (taking it to drink vs clean the table) (Iacoboni et al., 2005). Moreover, a "mirror-response," that is, the activation of the brain areas associated with the observer's direct experience, has been shown in different affective domains (Canessa et al., 2009; Canessa, Motterlini, Alemanno, Perani, & Cappa, 2011; Wicker et al., 2003).

The mentalizing system, including medial precuneus, temporoparietal junction (TPJ), ventromedial (vmPFC) and dorsomedial (dmPFC) prefrontal cortex, allows to extract and understand others' intentions via inferences on their thoughts and beliefs (Amodio et al., 2006), also when processing communicative intentions (Enrici, Adenzato, Cappa, Bara, & Tettamanti, 2011; Tettamanti et al., 2017). Unlike the mirror system, the latter comes into play when intentions cannot be automatically derived from visual cues, for example, in context-based inferences of mental states such as drinking to drown one's sorrows versus to make a toast (Hamilton & Grafton, 2006; Van Overwalle & Baetens, 2009), or when processing false beliefs, for example in the Sally-Ann test (Wimmer & Perner, 1983).

Even when inferences are drawn from action observation, however, the core regions of mirror and mentalizing systems are specifically recruited by identifying *how* (executed movements) versus *why* (beliefs and intentions) an action is performed (Spunt, Falk, & Lieberman, 2010, Spunt, Satpute, & Lieberman, 2011; Spunt & Adolphs, 2014; Spunt & Lieberman, 2012a, 2012b). Decoding others' actions at different levels of abstraction thus involves the relative activity of the mirror and mentalizing systems, processing *what* and *how* another person is doing (i.e., a behavioral state) versus *why* (i.e., a mental state) (Spunt, Kemmerer, & Adolphs, 2016; see also Chiavarino, Apperly, & Humphreys, 2012).

While fitting a basic distinction in social cognition between inferences on transitory behavioral states (e.g., momentary goals) versus enduring and more abstract characteristics (e.g., stable dispositions and intentions) (Hassin, Aarts, & Ferguson, 2005), the segregation between the mirror and mentalizing systems is however only partially supported by neuroscientific evidence.

Only in the case of *individual* actions, indeed, experimental and meta-analytic evidence supports their complementary roles. Namely, the engagement of the mirror and mentalizing systems would be driven by the presence of, respectively, *biological* actions versus *abstract* information (e.g., observing real scenes vs reading stories) or *implicit* versus *explicit* instructions (e.g., to passive observe vs to infer characters' intentions) (van Overwalle & Baetens, 2009), and by identifying *how* versus *why* the character is expressing a feeling (i.e., explicit identification vs attribution; Spunt & Lieberman, 2012a).

Increasing evidence, however, suggests that this segregation might not hold in the case of social interactions. Data based both on point-light displays (Centelles, Assaiante, Nazarian, Anton, & Schmitz, 2011) and ecological stimuli (Iacoboni et al., 2004; Kujala, Carlson, & Hari, 2012) rather show that the mirror and mentalizing systems can be simultaneously engaged in situations eliciting inferences on prospective social intentions. In this context, mirror areas in charge of action recognition might also provide the mentalizing network with sensorimotor information supporting and constraining inferential processes of intention understanding (Catmur, 2015). Their relative contribution, indeed, depends on cues highlighting either shared behavioral intentions (e.g., helping each other in cooperative interactions) or shared mental states (e.g., gazing to each other in affective interactions) (Figure 1), which recruit the mirror (vMPC and SPL) and mentalizing systems (mPFC), respectively (Canessa et al., 2012).

The analysis of brain connectivity may provide additional evidence on the interplay between the mirror and mentalizing systems while processing social interactions. To date, however, only few studies have addressed this issue. Psycho-physiological interaction (PPI) analyses unveiled a robust coupling between mirror and mentalizing regions, suggestive of their complementary roles, during the imitation (compared with passive observation) of hand gestures (Sperduti, Guionnet, Fossati, & Nadel, 2014), and when inferences on intentions are drawn from actions depicted in videos compared with text (Spunt & Lieberman, 2012b). Dynamic causal modeling (DCM) studies on action observation highlighted stronger forward than backward connectivity between motion-sensitive MT/V5 and pSTS while watching objects interacting (animate intentional motion) versus moving mechanically (Hillebrandt, Friston, & Blakemore, 2014), and reduced effective connectivity bidirectionally between parietal and temporal "mirror" nodes when participants observed increasingly familiar actions (Gardner, Goulden, & Cross, 2015). These results support the "predictive coding" framework, in which forward connections reflect the bottom–up propagation of prediction-error signals concerning stimulus-related unexpected sensory information, while upcoming prediction-errors are minimized by top–down backward connections carrying refined predictions based on an internal model (Koster-Hale & Saxe, 2013).

It is thus still unknown whether and how the causal organization within and between the two systems is modulated by specific cues, such as cooperative versus affective goals, during the extraction of intentions from social scenes. To fill this gap, we used DCM to investigate effective connectivity among the regions in which activity associated with observing social interactions also tracked increasing cooperativity (pSTS, SPL, and vPMC) versus affectivity (vmPFC). We first explored the endogenous connections among these regions, and then assessed how their direction and strength is modulated by increasing levels of perceived affectivity or cooperativity, and by the observer's empathic aptitude.

We predicted that a common engagement of the mirror and mentalizing systems alongside pSTS underpins the processing of both interaction types (Centelles et al., 2011; Georgescu et al., 2014; Iacoboni et al., 2005; Van den Stock et al., 2015), while increasing levels of cooperativity and affectivity should reflect in heightened activity of, respectively, mirror and mentalizing areas processing shared behavioral intentions versus shared mental states (Canessa et al., 2012). We expected such a functional distinction to emerge from effective connectivity within a strongly interconnected network with pSTS as driving input, in which increasing cooperativity and affectivity promote, respectively, the visuomotor processing of motor intentions and the

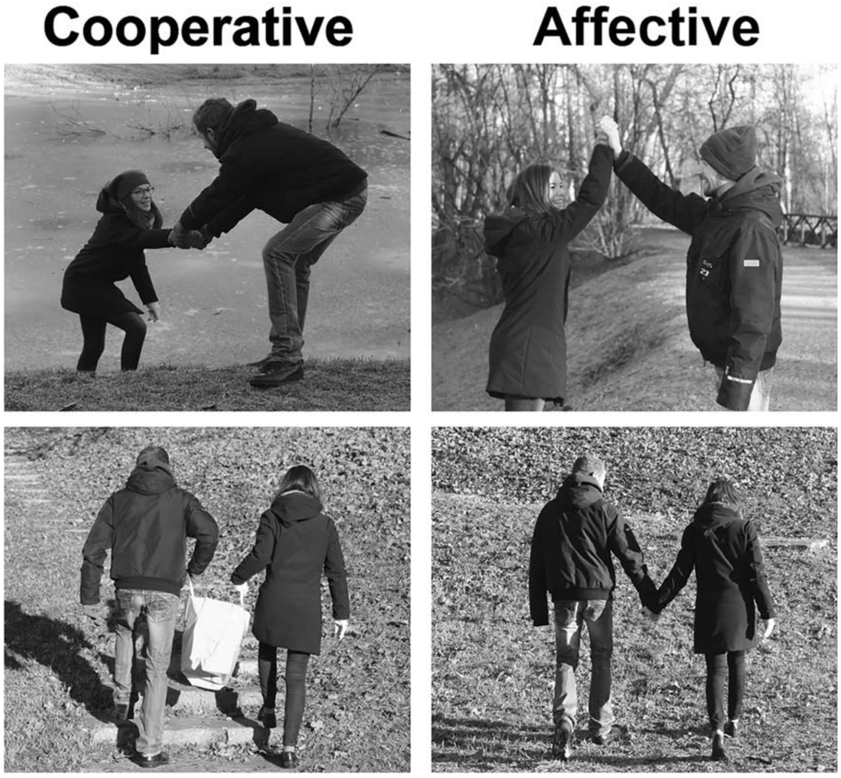

**FIGURE 1** Experimental stimuli. Representative examples of color pictures depicting cooperative (left) and affective (right) social interactions (see also Canessa et al., 2012; Proverbio et al., 2011). Reproduced with permission of the copyright owner

attribution of mental states by key-nodes of the mirror (vPMC and SPL) and mentalizing (vmPFC) systems. Based on recent evidence of stronger modulation of forward than backward connectivity in occipito-temporal circuitry while attending intentional biological motion (Hillebrandt et al., 2014), we predicted that higher cooperativity would reflect in stronger forward, than backward, connectivity conveying sensory information from the expected pSTS input node to the mirror system. We additionally assessed whether higher affectivity would reflect in stronger backward, than forward, connectivity carrying top–down information on the agents' mental states from vmPFC to pSTS.

## 2 | MATERIALS AND METHODS

### 2.1 | Participants

Thirty-six right-handed (Oldfield, 1971) healthy volunteers (20 females and 16 males; females mean age = 24.4 years, standard deviation (*SD*) = 4.75; males mean age = 25.4 years, SD = 4.16) participated in the study. The sample included 27 subjects from our previous study (Canessa et al., 2012). All subjects had normal or corrected-to-normal visual acuity, and none of them reported a history of neuropsychiatric conditions or substance abuse, nor was currently taking any medication interfering with cognitive functioning. They gave their written informed consent to the experimental procedure, which was approved by the local Ethics Committee.

### 2.2 | Stimuli

The stimulus-set comprised 260 color pictures depicting couples of both male and female individuals of various ages actively engaged in goal-directed interactions belonging to the human repertoire and expressing positive emotions (Figure 1). The goal of the action might consist in reaching a common aim (e.g., helping each other climb a tree) (130 cooperative interactions), or in establishing an affective contact (e.g., shaking hands) (130 affective interactions) (Canessa et al., 2012; Proverbio et al., 2011).

The selection of stimuli involved three stages of a rating procedure aimed to (a) prevent possible *confounding effects* such as the presence of an "action state" or an "action goal," emotional salience, gender, age, and number of persons, and body parts (whole-length bodies vs half-length bodies) and objects depicted; (b) identify the pictures fulfilling a *categorical* distinction between affective and cooperative social interactions, by means of 52 raters who evaluated the pictures for their affective or cooperative content; (c) lead back such dimension to a *parametric* variability, to overcome the potentially artificial categorization of pictures as either "affective" or "cooperative," by means of 81 raters reporting how much each picture expressed a sense of affectivity and, separately, cooperativity (see Supporting Information, Text 1 for further information on stimuli selection).

The outer background of all retained pictures was dark grey. Their average luminance was 15.48 Foot-lamberts, with no significant difference across categorical conditions or correlation with continuous affectivity/cooperativity values.

## 2.3 | Task and experimental procedure

To ensure and assess participants' engagement in the observation of pictures, still without emphasizing their affective versus cooperative content, we used a secondary task unrelated with mental state attribution. To this purpose, we also included 44 pictures matched to "social" ones for size and luminance, depicting common natural or urban landscapes without visible persons. Participants were asked to observe all pictures and respond to landscape ones. This task was aimed at preventing the induction of a conscious awareness of two types of interaction goal. Indeed, a postscanning debriefing session confirmed that no subject realized the twofold nature of the interactions displayed.

We assessed subjects' behavioral performance in terms of commission accuracy (% of button presses) and omission accuracy (% of missed responses) in association with landscape and social pictures, respectively.

Pictures were shown at the center of the screen for 1300 ms, and they were temporally separated by a red fixation-cross. The duration of this implicit baseline was varied ("jittered") at every trial to desynchronize the timings of event-types with respect to the acquisition of single slices within functional volumes and thus optimize statistical efficiency (Dale, 1999). We used the OptSeq2 Toolbox (http://surfer.nmr.mgh.harvard.edu/optseq/) to estimate the optimal interstimulus intervals (ISIs; mean ISI = 2.064 s, range = 0.325–9.750 s). Stimuli belonging to the three experimental conditions (cooperative and affective pictures, plus the secondary "landscape" pictures) were equally subdivided in 4 *f*MRI-runs, each comprising 152 pictures randomly intermixed, whose order was counterbalanced across subjects. To prevent any lateralization-effect of the motor response on cerebral activity, participants responded to target pictures with the right hand in two out of the four runs, and with the left hand in the other two. We counterbalanced the order of "left-hand" and "right-hand" runs across both male and female participants. Subjects viewed visual stimuli via a back-projection screen located in front of the scanner and a mirror placed on the head-coil. We used the software Presentation 11.0 (Neurobehavioral systems, Albany, CA, http://www.neurobs.com) both for stimuli presentation and subjects' answers recording. All participants underwent a training session during which they were instructed to gaze at the center of the screen and to avoid eye or body-movements during the scanning session. In a debriefing postscanning session, they were asked to report their personal impressions about the task and to complete an Italian version (Meneghini, Sartori, & Cunico, 2006) of the Balanced-Emotional-Empathy-Scale (BEES; Mehrabian & Epstein, 1972), a 30-items questionnaire measuring the individual tendency to empathize with others' emotional experiences.

## 2.4 | MRI data acquisition

We collected anatomical T1-weighted and functional T2*-weighted MR images with a 3 Tesla Philips Achieva scanner (Philips Medical Systems, Best, NL), using an 8-channels sense head coil (sense reduction factor = 2). Functional images were acquired using a T2*-weighted gradient-echo, echo-planar (EPI) pulse sequence (48 interleaved transverse slices, TR = 2600 ms, TE = 30 ms, flip-angle = 85°, field-of-view (FOV) = 192 mm × 192 mm, slice-thickness = 2.6 mm, interslice gap = 0.2 mm, in-plane resolution = 3 mm × 3 mm). Owing to specific hypotheses on the involvement of the vmPFC in social cognition, we tilted the FOV 30° downward with respect to the bicommissural line to reduce susceptibility artefacts from this region. While resulting in the loss of signal from the occipital cuneus and cerebellum in some subjects (Supporting Information, Figure 1), this procedure enhanced data quality from one of our primary regions of interest close to air/tissue interfaces. Each scanning sequence comprised 187 sequential volumes, preceded by 5 "dummy" functional volumes covering the amount of time needed to allow for T1-equilibration effects. Immediately after the functional scanning a high-resolution T1-weighted anatomical scan (150 slices, TR = 600 ms, TE = 20 ms, slice-thickness = 1 mm, in-plane resolution = 1 mm × mm) was also acquired for each subject.

Participants were positioned comfortably on the scanner bed and fitted with soft ear plugs; foam pads were used to minimize head movements.

## 2.5 | fMRI data preprocessing and statistical analysis

We performed image preprocessing using SPM8 (Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm), implemented in Matlab v7.4 (Mathworks, Inc., Sherborn, MA) (Worsley & Friston, 1995). The first 5 volumes of each functional run were discarded to allow for T1 equilibration effects. All remaining 748 volumes from each subject underwent a standard spatial preprocessing including slice-timing correction with the middle slice in time as a reference, spatial realignment to the first volume and unwarping, spatial normalization into the standard Montreal Neurological Institute (MNI) space and resampling in 2 × 2 × 2 mm$^3$ voxels, as well as spatial smoothing with a 8 mm full-width half-maximum (FWHM) isotropic Gaussian kernel. The resulting time series across each voxel were then high-pass filtered to 1/128 Hz, and serial autocorrelations were modelled as an AR(1) process. To evaluate effective connectivity with DCM (see below), we concatenated volumes from the four functional runs to form one single time series per subject, and added a regressor modeling session effects. In addition, we used the MotionFingerprint toolbox (http://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/) to compute, for each subject, a comprehensive indicator of scan-to-scan head motion.

We used SPM12 to perform an event-related statistical analysis of *f*MRI data aimed to assess a continuous relationship between the strength of BOLD activity and the degree of affectivity or cooperativity expressed by the observed social interactions. This analysis additionally provided driving input and contextual modulators of effective connectivity in subsequent DCM analyses. Statistical maps were based on a random-effect model implemented in a two-levels procedure (Friston, Zarahn, Josephs, Henson, & Dale, 1999). At the first (single-subject) level, we modeled event-related *f*MRI responses as "stick" functions by a design-matrix comprising the onset of all social scenes ("observation of social interaction vs implicit cross-fixation baseline" regressor, with duration equal to zero). Two further parametric regressors reflected a linear modulation of the observation-related activity by the degree of "affectivity" and "cooperativity." The "affectivity" and "cooperativity" values, resulting from the rating procedure previously described, were not significantly correlated ($r = .19$, $p = .761$; Supporting Information, Text 1).

To control for the neural processing of complex background elements, in a secondary analysis, we used a categorical (rather than a parametric) modeling of stimuli to contrast cooperative/affective scenes with landscape pictures (Supporting Information, Text 2 and Figures 2 and 3). To this purpose, at the single-subject level, we separately modeled the onset of cooperative and affective pictures, and of landscape pictures to avoid they could represent an implicit baseline.

Additional regressors modeled scan-to-scan head motion and constant session effects. We then convolved regressors modeling events with a canonical hemodynamic response function (HRF), and obtained parameter estimates for all regressors by maximum-likelihood estimation.

At the second level, we performed a random-effect group analysis across the 36 subjects using a full-factorial design with sphericity correction for repeated measures (Friston et al., 2002). Namely, we used the statistical maps resulting from the parametric analysis to identify (a) the voxels activated by the "observation of social interactions" regardless of the underlying cooperative/affective dimension (Figure 2a and Table 1a), as well as (b) those in which the strength of BOLD activity displayed a positive linear relationship with the degree of affectivity or cooperativity expressed by the observed scene (Figure 2b,c and Table 1b,c). We reported as statistically significant only the voxels surviving a statistical threshold of $p < .05$ corrected for multiple comparisons based on cluster-extent using topological false discovery rate (FDR; Chumbley & Friston, 2009).

We used the SPM Anatomy Toolbox (v.2.2c; Eickhoff et al., 2005) to localize the activated brain regions in the MNI space in terms of (a) microanatomical labels based on the overlap between each cluster and probabilistic cytoarchitectonic maps (when available); (b) macroanatomical labels provided by the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) for clusters located outside these maps.

## 2.6 | DCM

DCM estimates the experimental modulation of (intrinsic) self-connections or (extrinsic) forward and backward connections between brain regions that are active during a particular task in a directional manner (Friston, Harrison, & Penny, 2003). This approach allows inferring whether experimental manipulations affect top–down influences, bottom–up influences, or both, in terms of the strength and direction of coupling between the modeled regions of interest. In this study, we performed dynamic causal modeling with DCM12 (v6685) to test whether and how the degree of affectivity versus cooperativity expressed by observed social interactions modulates effective connectivity within and between the key nodes of the mirror and mentalizing systems. We modeled 4 regions highlighted both by our previous evidence (Canessa et al., 2012) and in preliminary *f*MRI analyses (2.5, 2.6.1 and Supporting Information, Text 2), that is, left posterior superior temporal sulcus (pSTS), superior parietal lobule (SPL), and ventral premotor cortex (vPMC) tracking increasing cooperativity, alongside the ventromedial prefrontal cortex (vmPFC) tracking increasing affectivity. We used a post-hoc model selection routine (i.e., "Optimize") to identify the best fitting model, among all possible ones, at the group level. We pursued such a Bayesian approach via Network discovery (Friston, Li, Daunizeau, & Stephan, 2011), to make inferences both on *model structure* (i.e., to compare alternative DCMs) and *model parameters* (i.e., to unveil the functional architecture of the winning reduced DCM).

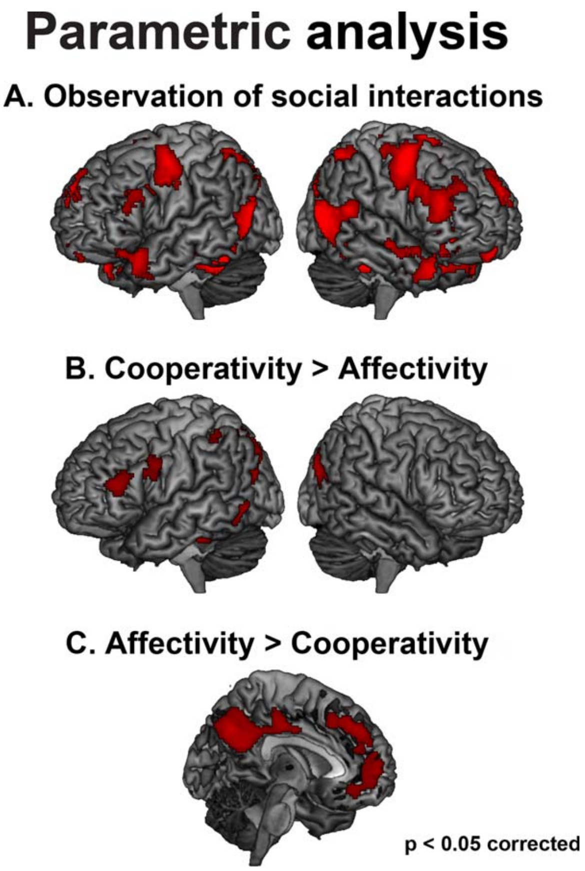
Parametric analysis
A. Observation of social interactions
B. Cooperativity > Affectivity
C. Affectivity > Cooperativity
p < 0.05 corrected

FIGURE 2 Brain activity highlighted by a parametric coding of social interactions. The brain regions activated by the observation of social interactions regardless of their purpose (a), and those in which activity is more strongly related to the degree of cooperativity than affectivity (b) or affectivity than cooperativity (c). All the reported activations survived a statistical threshold of $p < .05$ corrected for multiple comparisons based on cluster extent (Chumbley & Friston, 2009)

### 2.6.1 | Volume of interest selection

DCM aims to explain the activations and deactivations highlighted by standard SPM analyses (Friston et al., 2003). On this assumption, we performed DCM on four Volumes-of-interest (VOIs) selected based on our a priori hypotheses that (a) a common engagement of the superior parietal and ventral premotor nodes of the mirror system, alongside pSTS, underpins the processing of both interaction types, while (b) increasing cooperativity and affectivity specifically reflects in heightened activity of, respectively, mirror and mentalizing areas processing

**TABLE 1** Parametric *f*MRI analysis: neural processing of cooperativity and affectivity in observed social interactions

| H | Anatomical region | AT | *x* | *y* | *z* | *t* score | *K* | Cluster *p* value |
|---|---|---|---|---|---|---|---|---|
| 1a. Observation of social interactions | | | | | | | | |
| L | Precentral gyrus | | −40 | −6 | 46 | **9.44** | 3797 | **<.0001** |
| L | IFG (pars triangularis) | | −32 | 32 | −2 | **6.73** | | |
| L | Insula lobe | | −26 | 28 | 2 | **7.62** | | |
| R | Middle frontal gyrus | | 34 | −2 | 62 | **6.95** | 3325 | **<.0001** |
| R | Precentral gyrus | | 36 | −4 | 50 | **9.58** | | |
| R | IFG (pars opercularis) | | 44 | 14 | 30 | **9.30** | | |
| R | Insula lobe | | 34 | 30 | 0 | **7.57** | | |
| L | Posterior-medial frontal | | −8 | 8 | 52 | **8.41** | 1433 | **<.0001** |
| R | Posterior-medial frontal | | 8 | 10 | 52 | **8.64** | | |
| L | Superior medial gyrus | | 0 | 60 | 32 | **6.46** | 143 | **.016** |
| R | Rectus gyrus | Fp2 | 6 | 56 | −20 | **5.90** | 95 | **.043** |
| L | Fusiform gyrus | FG3 | −28 | −60 | −10 | **19.72** | 21992 | **<.0001** |
| R | Fusiform gyrus | | 36 | −56 | −10 | **16.97** | | |
| L | Amygdala | | −20 | −8 | −18 | **7.14** | | |
| R | Amygdala | LB | 20 | −8 | −20 | **8.76** | | |
| L | Middle temporal gyrus | | −48 | −70 | 12 | **9.77** | | |
| R | Middle temporal gyrus | Hoc4la | 46 | −72 | 10 | **14.27** | | |
| R | Superior temporal gyrus | | 52 | −40 | 16 | **7.00** | | |
| L | Intraparietal sulcus | hIP3 | −28 | −54 | 48 | **6.78** | | |
| R | Intraparietal sulcus | hIP1 | 28 | −54 | 48 | **11.04** | | |
| R | Superior temporal gyrus | | 52 | −10 | −12 | **5.18** | 135 | **.017** |
| R | Temporal pole | | 54 | 14 | −16 | 4.52 | | |
| R | Precuneus | | 4 | −50 | 48 | **5.48** | 103 | **.038** |
| L | Paracentral lobule | 4a | −6 | −36 | 62 | **6.02** | 848 | **<.0001** |
| R | Paracentral lobule | 4a | 10 | −36 | 64 | **8.21** | | |
| 1b. Cooperativity > affectivity | | | | | | | | |
| L | IFG (pars triangularis) | | −48 | 38 | 14 | **4.42** | 641 | **.001** |
| L | Precentral gyrus | | −40 | −6 | 44 | 3.61 | | |
| L | IFG (pars opercularis) | 44 | −50 | 10 | 28 | 3.81 | | |
| L | Fusiform gyrus | FG3 | −28 | −54 | −6 | **11.82** | 1793 | **<.0001** |
| L | Inferior temporal gyrus | | −46 | −60 | −6 | **4.32** | | |
| L | Middle temporal gyrus | | −54 | −64 | −4 | 3.76 | | |
| L | Middle/superior temporal gyrus | | −46 | −62 | 6 | 3.47 | | |
| L | Middle occipital gyrus | | −26 | −66 | 34 | **5.63** | 1831 | **<.0001** |
| L | Superior parietal lobule | 7a | −18 | −70 | 44 | **5.82** | | |
| R | Fusiform gyrus | | 28 | −44 | −10 | **12.29** | 3692 | **<.0001** |
| R | Middle temporal gyrus | | 44 | −74 | 22 | **5.26** | | |
| L | Intraparietal sulcus | hIP1 | −34 | −40 | 38 | 3.78 | 236 | **.050** |
| L | Intraparietal sulcus | hIP3 | −38 | −40 | 44 | 3.17 | | |

(Continues)

**TABLE 1** (Continued)

| H | Anatomical region | AT | *x* | *y* | *z* | *t* score | *K* | Cluster *p* value |
|---|---|---|---|---|---|---|---|---|
| 1c. Affectivity > cooperativity | | | | | | | | |
| L | Superior medial Gyrus | Fp2 | −10 | 56 | 4 | 4.01 | 600 | **.001** |
| R | Mid orbital gyrus | Fp2 | 4 | 60 | −6 | 3.61 | | |
| L | Mid orbital gyrus | s32 | −4 | 42 | −12 | 3.16 | | |
| R | Superior medial gyrus | | 10 | 50 | 36 | 4.62 | 381 | **.008** |
| R | Superior frontal gyrus | | 20 | 36 | 44 | 3.84 | | |
| R | Precuneus | | 6 | −58 | 38 | **5.70** | 2028 | **<.0001** |
| R | Posterior cingulate cortex | | 4 | −50 | 26 | 4.99 | | |
| L | Middle cingulate cortex | | −2 | −20 | 40 | 4.25 | | |

*Note*. Abbreviations: K: cluster extent in number of voxels (2 × 2 × 2 mm$^3$); H: hemisphere; L: left; R: right; IFG: inferior frontal gyrus; Fp2: medial frontopolar area 2; FG: fusiform gyrus; LB: laterobasal amygdala nuclei; 4hOc4la: anterior portion of lateral occipital cortex; hIP: human intraparietal; s32: subgenual portion of anterior cingulate cortex. Anatomical labeling was performed in the MNI space based on the overlap between each cluster and cytoarchitectonic probability maps on the Anatomy Toolbox for SPM (v.2.2c; Eickhoff et al., 2005) when available (AT), or with the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) otherwise (Anatomical region). Bold font denotes effects significant at $p < .05$ FDR corrected for multiple comparisons either at voxel- or cluster-level.

shared behavioral intentions versus shared mental states (Canessa et al., 2012; see Introduction). In particular, we selected the vmPFC among the regions tracking increasing affectivity because our previous study had shown a correlation between its activation strength and individual empathy scores.

On this basis, we used conjunction analyses to identify the coordinates fulfilling the above criteria (Table 2). First, we performed a conjunction analysis among the contrasts for cooperative pictures and affective pictures (both from the categorical analysis) as well as increasing cooperativity (from the parametric analysis). This analysis provided the coordinates for the left pSTS (*xyz* = −46 −62 6), SPL (*xyz* = −18 −70 44), and vPMC (*xyz* = −40 −6 44) associated with both increasing cooperativity and the common effect of cooperative and affective categorically defined pictures. Then, we performed a conjunction analysis between the contrasts for the direct comparison between affective and cooperative pictures (from the categorical analysis) and the contrast for increasing affectivity (from the parametric analysis). This analysis provided the coordinates for the vmPFC (*xyz* = −10 56 4) associated with both increasing affectivity and a stronger response to affective than cooperative categorically defined pictures.

In line with a recent DCM study describing effective connectivity in the mirror system (Gardner et al., 2015), the results of the parametric analysis implicitly constrained the selection of VOIs to the left hemisphere.

We used these coordinates as the center of 8-mm-radius spheres from which we extracted the first eigenvariate of the single-subject time series (threshold $p < .05$ uncorrected), using the subject-level parametric contrasts and adjusting at $p < .05$ for the effects of interest (i.e., only for those regressors that would be used in the DCMs for input or modulation). The eigenvariate of each VOI accounted for a considerable amount of variance of the original time series (mean = 61.9, *SD* = 7.72 for pSTS; mean = 68.3, *SD* = 6.04 for SPL; mean = 51.7, *SD* = 5.36 for vPMC; mean = 67.0, *SD* = 8.22 for vmPFC; mean = 62.2, *SD* = 7.58 across all VOIS).

**TABLE 2** Volume of interest selection

| H | Anatomical region | *x* | *y* | *z* | *K* | FDR-corrected cluster *p* value |
|---|---|---|---|---|---|---|
| L | pSTS | −46 | −62 | 6 | 1467 | <.001 |
| L | SPL | −18 | −70 | 44 | 985 | <.001 |
| L | vPMC | −40 | −6 | 44 | 142 | .063 |
| L | vmPFC | −10 | 56 | 4 | 236 | .1 |

*Note*. MNI coordinates of the DCM nodes (Section 2.6.1).

### 2.6.2 | Specification of dynamic causal models

We then specified bilinear DCMs for subsequent estimation and post-hoc Bayesian selection of a reduced model. The input into bilinear DCM is represented by the VOIs (i.e., BOLD signal in the nodes of the model) and three types of matrices representing: (a) the endogenous connections, that is, “fixed” connectivity between such nodes (A matrix); (b) the exogenous driving input, creating activity into the system (C matrix); (c) the connections which are modulated by contextual factors, that is, in terms of the change in the effective connectivity value under specific conditions (B Matrices). The B and C matrices thus represent the experimentally manipulated conditions.

The DCMs were based on the design-matrix created in the parametric analysis previously described. We specified the “observation of social interaction versus implicit baseline” contrast as exogenous driving input into the model, and the parametric regressors (i.e., degree of affectivity and degree of cooperativity) as contextual modulators of connectivity. Therefore, while the driving input regressor modeled non-specific effects of observing social interactions, the parametric regressors modeled the effects of perceived affectivity or cooperativity expressed by such interactions. Under these assumptions, and following the "Network discovery" approach (Friston et al., 2011), we specified a model characterized by fully connected A, B, and C matrices. All DCMs were deterministic, bilinear one-state models without mean-centered inputs.

In sum, our DCM entailed a full 4 × 4 A matrix (endogenous connectivity), full 4 × 4 B matrices (modulation by "affectivity" and "cooperativity"), and a C matrix coding all the 4 specified regions as possible driving inputs of "observation of social interaction" into the model. This corresponds to modeling the effect of the degree of affectivity and cooperativity as context-sensitive changes in coupling induced by the observation of social interactions.

### 2.6.3 | Post-hoc Bayesian model selection

We then used "Network discovery" (Friston et al., 2011) to make inferences on the parameters (i.e., strength and direction of coupling between the modeled regions) of the best-fitting reduced model. Via Bayesian model reduction (Friston & Penny, 2011), this approach implements an exhaustive search over all possible combinations of connections (and how they are differentially modulated by the degree of affectivity and cooperativity) to identify the best model and thus the underlying functional architecture (Yang et al., 2017). In DCM12, this is done via the "Optimize" routine, which searches over all possible reduced submodels of a fully connected model and uses a post-hoc model selection routine to identify the best fitting one, in terms of the tradeoff between model fit and model complexity, at the group level. This approach first fits the full model with all its free parameters to the given data. The evidence for all reduced models, that is, all possible models nested in the full model, is approximated by effectively removing the parameters (using a "greedy search" routine (Friston & Penny, 2011; Rosa, Friston, & Penny, 2012) when their numerosity is larger than 8). The selection routine results in posterior probabilities for the model (i.e., the probability of that model being the best compared to any other model given the group data) and its connections (i.e., whether a given connections exists or not). In both cases, we reported as statistically significant only posteriors larger than 0.95 (Supporting Information, Figure 4).

### 2.6.4 | Comparison of connection strength

Subsequent inferences involve either the model level (i.e., whether a connection exists or not) or the parameter level (i.e., direction and strength of effective connectivity, assuming that the connection exists) in the winning model. We pursued the latter approach to address different facets of the modulation exerted by the degree of affectivity versus cooperativity on effective connectivity in the reduced model. To this purpose, we investigated

a which connections are retained in the winning model highlighted by the "Optimize" routine;

b which direct inputs and endogenous parameters are significantly different from zero, by means of non-parametric, two-sided, Wilcoxon signed-rank tests of means (Figure 3a–c and Table 3).

c which connections are differentially modulated by the degree of affectivity versus cooperativity expressed by social interactions, by means of nonparametric, two-sided, Wilcoxon signed-rank tests of means (Figure 3d and Table 3);

d the relationship between the strength of such modulatory influences and an individual empathy score (as measured by the BEES questionnaire; Mehrabian & Epstein, 1972) by means of Spearman's rho correlation coefficient.

For (b) to (d), we discussed only the results surviving a correction for multiple comparisons based on FDR (Benjamini & Hochberg, 1995).

As previously discussed, we expected that the cooperative and affective dimensions would exert different modulations on effective connectivity between the predicted pSTS input region and, respectively, the main nodes of the mirror (SPL, vPMC) and mentalizing (vmPFC) systems. We assessed this hypothesis with a 2 × 3 × 2 repeated measures analysis of variance (ANOVA), with the strength of modulation (i.e., Matrix B) as dependent variable and, as independent variables, the "source" of modulation (cooperativity or affectivity, i.e., Matrix B2 or B3), the "target" region connecting with pSTS (vPMC, SPL or vmPFC), and the "direction" of connectivity (forward or backward). We additionally used Fisher LSD post-hoc tests to obtain specific information on which means are significantly different from each other.

We additionally assessed whether the differential connectivity from pSTS to distinct downstream nodes would involve mainly forward or backward connections when modulated by cooperativity versus affectivity. To this purpose, we used two 2 × 2 repeated measures ANOVAs with the modulatory effect (cooperativity or affectivity) and the direction of connectivity (forward or backward) as independent variables, and either the difference between pSTS-SPL and pSTS-vmPFC connection strength, or the difference between pSTS-vPMC and pSTS-vmPFC connection strength, as dependent variable.

# 3 | RESULTS

## 3.1 | Behavioral performance

The behavioral assessment of participants' responses highlighted a high level of performance, with no significant effect of picture-type (cooperative mean percentage of correct responses = 95.3%, *SD* = 0.77%; affective mean = 95.5%, *SD* = 0.94%; landscape mean = 95.5%, *SD* = 0.75%; F(2) = 0.027, *p* = .973) and participants's gender (F (1) = 0.686, *p* = .413), nor of their interaction (F(2) = 1.003, *p* = .372). These results confirmed that all picture-types were carefully observed by both male and female participants.

## 3.2 | Neural processing of cooperativity and affectivity

The observation of social interactions (i.e., regardless of their degree of cooperativity or affectivity) activated a bilateral network previously associated with action observation, including regions with mirror properties (Caspers et al., 2010; Molenberghs et al., 2012) (Figure 2a and Table 1a). This network extended from the fusiform gyrus to posterior middle and superior temporal cortex, as well as to posterior parietal regions (superior parietal lobule and intraparietal area (hIP)) and both frontolateral and frontomedial areas. In particular, frontolateral activations involved a widespread cluster encompassing the inferior frontal gyrus (pars triangularis in the left hemisphere and pars opercularis in the right hemisphere) and the ventral premotor cortex (close to the border with dorsal premotor cortex; Mayka, Corcos, Leurgans, & Vaillancourt, 2006). Finally, the temporal pole and the amygdala, alongside dorsomedial and ventromedial prefrontal cortex, were bilaterally activated when observing social interactions.

Some of these regions were also associated with the parametric effects of observing social interactions expressing variable levels of cooperativity versus affectivity. Increasing cooperativity, however, recruited only the *left hemispheric* sectors of the mirror system (Figure 2b and Table 1b). Compared with affectivity, indeed, increasing

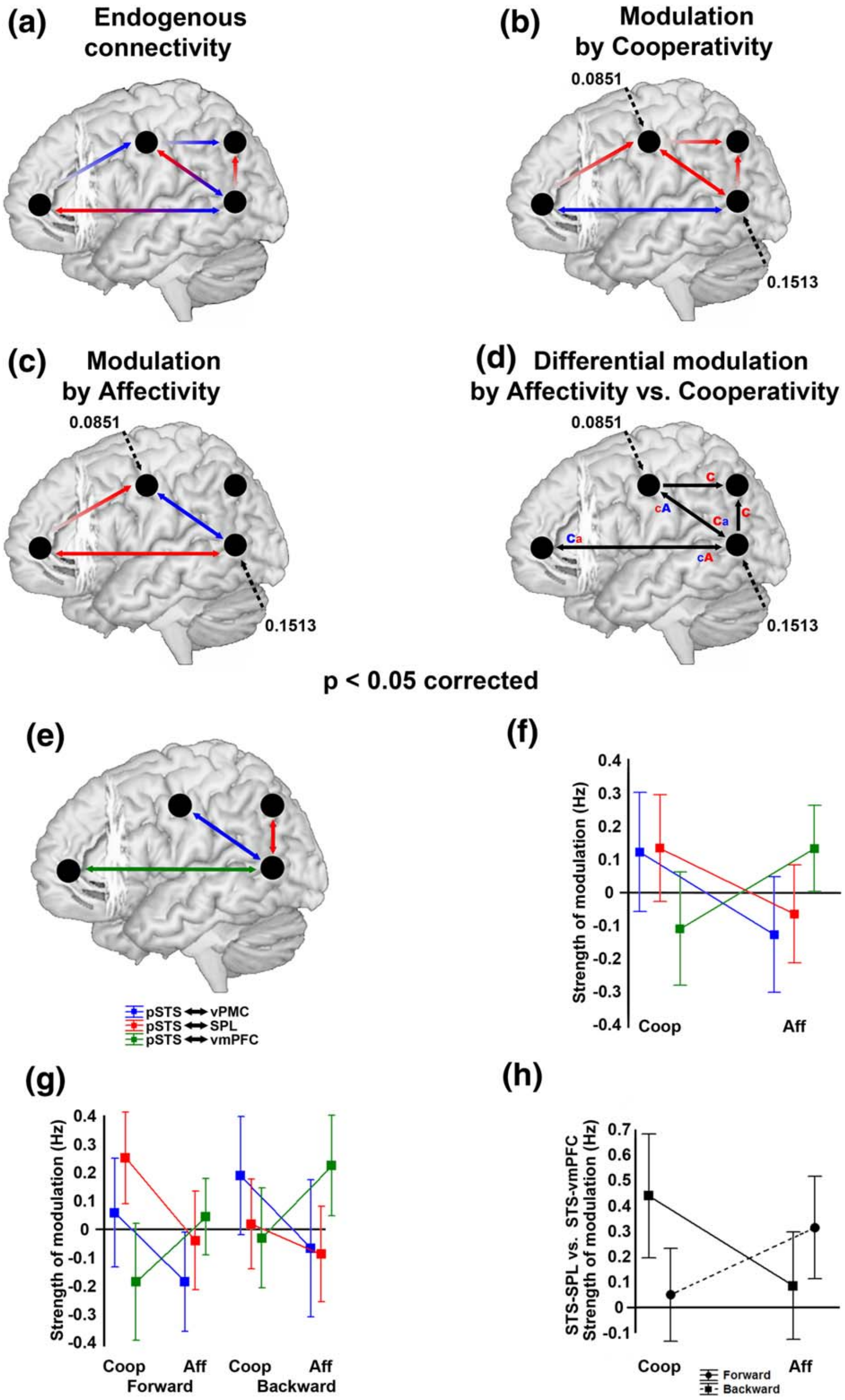

FIGURE 3.

cooperativity reflected in the activation of bilateral extrastriate regions (from fusiform gyrus to superior occipital gyrus), and of the left superior parietal (7a) and intraparietal (hIP3 and hIP1) areas associated with the human mirror system (Caspers et al., 2010; Molenberghs et al., 2012). In the frontal lobe, increasing cooperativity was associated with stronger activity in the left ventral premotor cortex (pars opercularis and pars triangularis). In contrast, higher affectivity specifically reflected in linearly increasing activity in ventromedial and dorsomedial prefrontal regions, and in the posterior cingulate cortex and precuneus (Figure 2c and Table 1c).

### 3.3 | Effective connectivity

We used Network discovery (Friston et al., 2011) to unveil the optimal functional architecture, in terms of tradeoff between model fit and model complexity, within our 4-vois full model. This post-hoc model selection highlighted a "winning" reduced model associated with a posterior probability of 0.9734 (Supporting Information, Figure 4), while the next best model has a very low probability of 0.0142 (ratio = 68.55). This model is thus highly probable both in absolute and relative terms, since a ratio between the "winning" and the next best model (i.e., Bayes factor) between 20 and 150 is considered a strong evidence (Kass & Raftery, 1995).

The driving input, that is, the effect of "observing social interactions" (all events vs implicit baseline, regardless of the degree of cooperativity and affectivity), enters the network into both the pSTS and vPMC (Figure 3b,c). All the endogenous connections (A matrix) and most of the modulatory effects (B matrices) are retained, with the only exception of the modulation exerted by the degree of affectivity on the connections from pSTS and vPMC to the SPL.

After identifying this winning model, we assessed its properties at the parameter level; that is, in terms of differential modulation exerted on effective connectivity by the degree of affectivity versus cooperativity expressed by social interactions. All the parameters were normally distributed (Kolmogorov–Smirnov > .05). In line with the optimized model resulting from Network discovery, both pSTS and vPMC direct input parameters were significantly different from zero (Figure 3b,c and Table 3). With respect to endogenous connectivity, we found significantly different from zero estimates in all the three forward connections originating from the input pSTS node, and in the backward connections reaching the pSTS from the input vPMC node and the vmPFC. The backward connections from vmPFC to vPMC and from the vPMC to the SPL were the only other significantly different from zero endogenous connections. All the forward connections were excitatory, while all the backward connections were inhibitory.

A significantly different modulation by affectivity versus cooperativity involved (a) the reciprocal effective connections between the pSTS and both the vPMC and vmPFC nodes, and (b) the connections from both pSTS and vPMC to the SPL (Figure 3d and Table 3).

In particular, the degree of cooperativity and affectivity exerted an oppositely valenced modulation on the direction and strength of effective connectivity between the pSTS and the frontal nodes of the mirror (vPMC) and mentalizing (vmPFC) systems (Figure 3d). The forward connection from pSTS to vmPFC was more strongly downregulated by increasing cooperativity than upregulated by increasing affectivity, while the backward connection from vmPFC to pSTS was more strongly upregulated by increasing affectivity than downregulated by increasing cooperativity. On the contrary, the forward connection from pSTS to vPMC was more strongly downregulated by increasing affectivity than upregulated by increasing cooperativity, while the backward connection from vPMC to pSTS was more strongly upregulated by increasing cooperativity than downregulated by increasing affectivity. In addition, increasing cooperativity was also associated with a stronger positive modulation of the forward connection from pSTS to SPL and of the backward connection from vPMC to SPL (for both connections, the modulatory effect of affectivity was removed in the optimized model).

Based on our hypotheses (see Introduction), we then further assessed the differential modulatory influence exerted by the degree of cooperativity versus affectivity on the strength of forward versus backward connections between the input pSTS region and all the other 3 modeled nodes (Figure 3e).

First, ANOVA results highlighted a strongly significant two-way interaction between the factors "target" (vPMC/SPL/vmPFC) and

FIGURE 3 Modulation of forward and backward effective connectivity by the degree of perceived affectivity versus cooperativity. (a) The endogenous connectivity architecture of the winning reduced model after random-effect analyses at the parameter level (see Table 3 for the values of connectivity strength). Red and blue arrows depict excitatory and inhibitory endogenous connections, respectively. (b,c) The positive (red arrows) or negative (blue arrows) modulation of endogenous connectivity by the degree of cooperativity (b) or affectivity (c) expressed by observed social interactions. The straight dashed lines depict the driving input (i.e., "Observation of social interactions") entering the system both in pSTS and vPMC. (d) The effective connections showing a significantly different modulation by the degree of cooperativity versus affectivity in the winning model (see Table 3 for the values of modulation strength). The reciprocal effective connections between pSTS and vPMC, and the connections from vPMC and pSTS to SPL, were more strongly upregulated by cooperativity (positive modulation) than affectivity (negative modulation), while the opposite occurs in pSTS-vmPFC reciprocal effective connectivity (see also (f)). Red and blue letters denote, respectively, positive and negative modulations of endogenous connectivity by the degree of cooperativity ("C"; "c") or affectivity ("A"; "a"), with letter-size representing their relative effect (e.g., Ca = larger modulation by cooperativity than affectivity). (e) The forward and backward effective connections between pSTS and the other three network nodes reported in subsequent panels. As shown in (f), perceived affectivity and cooperativity exerted *oppositely valenced* modulations on pSTS-vmPFC (green in (e–g); positively modulated by affectivity), and pSTS-SPL (red) and pSTS-vPMC (blue) (positively modulated by cooperativity) reciprocal effective connectivity. A breakdown of this graph into forward and backward connections (g,h) additionally shows that these connections are also subject to *oppositely directed* modulations by cooperativity and affectivity, that is, stronger modulation by cooperativity on forward excitatory pSTS-SPL and backward inhibitory vPMC-pSTS connectivity, and by affectivity on backward inhibitory vmPFC-pSTS connectivity

**TABLE 3** Differential modulation of effective connectivity by the degree of cooperativity versus affectivity

| Input region | Mean input strength | *p* value |
|---|---|---|
| Left pSTS | 0.15 | <.00001 |
| Left vPMC | 0.08 | <.00001 |

| Connection | Endogenous connectivity strength | | One-sample t-test |
|---|---|---|---|
| | Mean | *SD* | *p* value |
| **pSTS→vPMC** | 0.08 | 0.20 | **.014** |
| **pSTS→SPL** | 0.11 | 0.20 | **.011** |
| **pSTS→vmPFC** | 0.09 | 0.14 | **.002** |
| **vPMC→pSTS** | −0.12 | 0.18 | **.002** |
| **vPMC→SPL** | −0.10 | 0.16 | **.003** |
| vPMC→vmPFC | 0.05 | 0.16 | .079 |
| SPL→pSTS | −0.02 | 0.07 | .092 |
| SPL→vPMC | 0.03 | 0.11 | .103 |
| SPL→vmPFC | 0.01 | 0.18 | .975 |
| **vmPFC→pSTS** | −0.07 | 0.08 | **.001** |
| **vmPFC→vPMC** | −0.04 | 0.10 | **.019** |
| vmPFC→SPL | 0.05 | 0.16 | .073 |

| Connection | Modulation by cooperativity | | Modulation by affectivity | | Cooperativity vs affectivity |
|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | *p* value |
| **pSTS→vPMC** | 0.06 | 0.57 | −0.19 | 0.52 | **.010** |
| **pSTS→SPL** | 0.25 | 0.48 | 0.00 | 0.51 | **.028** |
| **pSTS→vmPFC** | −0.19 | 0.61 | 0.04 | 0.40 | **.014** |
| **vPMC→pSTS** | 0.19 | 0.62 | −0.07 | 0.72 | **.015** |
| **vPMC→SPL** | 0.18 | 0.45 | 0.00 | 0.00 | **.012** |
| vPMC→vmPFC | 0.04 | 0.56 | 0.19 | 0.73 | .441 |
| SPL→pSTS | 0.02 | 0.47 | −0.09 | 0.50 | .109 |
| SPL→vPMC | 0.01 | 0.53 | 0.08 | 0.79 | .838 |
| SPL→vmPFC | −0.03 | 0.60 | 0.06 | 0.64 | .451 |
| **vmPFC→pSTS** | −0.03 | 0.52 | 0.22 | 0.52 | **.013** |
| vmPFC→vPMC | 0.08 | 0.48 | 0.08 | 0.68 | .777 |
| vmPFC→SPL | −0.01 | 0.48 | 0.16 | 0.74 | .258 |

| Connection | Endogenous connectivity + modulation by cooperativity | Endogenous connectivity + modulation by affectivity |
|---|---|---|
| | Mean | Mean |
| **pSTS→vPMC** | 0.14 | −0.11 |
| **pSTS→SPL** | 0.36 | 0.11 |
| **pSTS→vmPFC** | −0.1 | 0.13 |
| **vPMC→pSTS** | 0.07 | −0.19 |
| **vPMC→SPL** | 0.08 | −0.1 |
| **vmPFC→pSTS** | −0.1 | 0.15 |

*Note*. From top to bottom, mean and standard deviation (*SD*) of the strength of (a) the driving inputs into the network; (b) the endogenous connections among the 4 network nodes; and (c) the modulatory influence exerted on the endogenous connections by the degree of cooperativity versus affectivity expressed by observed social interactions. Bold font denotes a statistically significant difference ($p < .05$ corrected for multiple comparisons with false-discovery rate (FDR; Benjamini & Hochberg, 1995). For the connections showing both the strength of endogenous connectivity significantly different from zero and a significantly different modulation by the degree of cooperativity versus affectivity, the last section of the table reports the mean net effect of endogenous connectivity and modulatory influences.

"modulation" (cooperation/affectivity) (F(2,70) = 10.640, $p = .00009$) (Figure 3f). Namely, post-hoc tests confirmed that cooperativity, compared with affectivity, exerted a larger modulation of both pSTS-SPL (mean-cooperativity = 0.135; mean affectivity = −0.064) and pSTS-vPMC (mean-cooperativity = 0.123; mean affectivity = −0.127) reciprocal effective connectivity, while the opposite occurred for the pSTS-vmPFC reciprocal effective connectivity (mean-cooperativity = −0.109; mean affectivity = 0.134).

Then, a 2 × 2 ANOVA confirmed the hypothesis that the differential connectivity between pSTS and SPL versus vmPFC additionally reflects a significant interaction between the modulatory effect (cooperativity/affectivity) and the direction of connectivity (forward/backward) ($F(1,35) = 4.649$, $p = .038$). As shown in Figure 3g,h, the higher positive modulation of the connection between pSTS and SPL (compared with vmPFC) elicited by cooperativity involves more strongly the forward than the backward direction. The opposite is true of the higher positive modulation of the connection between pSTS and vmPFC (compared with SPL) elicited by affectivity, which involves more strongly the backward than the forward direction. No such interactive effect was found for the connectivity between pSTS and vPMC, neither when compared with pSTS-vmPFC nor when compared with pSTS-SPL: as shown in Figure 3d, indeed, the sensitivity of this connection to the modulatory effect of cooperativity involves more strongly the backward, compared with the forward, connection.

Finally, we found no significant relationship between subjects' empathic aptitude (as measured by the BEES questionnaire) and the strength of such goal-specific modulations on effective connectivity.

## 4 | DISCUSSION

Implicit intention understanding while observing cooperative and affective social interactions recruits both shared and specific brain regions in the mirror and mentalizing neural systems, respectively (Canessa et al., 2012). In this study, we investigated whether and how variable degrees of affectivity versus cooperativity modulate the direction and strength of causal influence among four brain regions associated with their processing. To this purpose, we first complemented the categorical modeling of *f*MRI events (entailing a clear-cut distinction between the two interaction goals; Supporting Information, Text 2) with a parametric one (based on the *degree* of affectivity/cooperativity expressed by each picture). The parametric approach provided DCMs with a driving input into the network (observation of social interactions) and contextual modulators of its neural activity (degree of affectivity or cooperativity).

In line with previous results (Iacoboni et al., 2004), we found that the observation of social interactions, regardless of their affective/cooperative dimension, activates key-nodes of *both* the mirror (left pSTS, superior parietal, and ventral premotor cortex; Caspers et al., 2010; Molenberghs et al., 2012; Rizzolatti & Craighero, 2004) and mentalizing (mPFC; Canessa et al., 2012; Enrici et al., 2011) networks (Figure 2a and Table 1a). Direct comparisons between the regions tracking cooperativity and affectivity additionally highlighted a functional distinction between the two systems. Higher cooperativity reflected in stronger activity involving the left-hemispheric superior parietal and ventral premotor nodes of the mirror system, likely underpinning the in-depth visuomotor analyses required to decode shared motor intentions (Figure 2b and Table 1b). Conversely, higher affectivity was tracked by activity in the vmPFC, alongside dmPFC, and medial precuneus; that is, key-nodes of the mentalizing network underlying inferences on others' mental states (Amodio et al., 2006) (Figure 2c and Table 1c). Notably, the involvement of the precuneus, previously undetected by a categorical modeling of stimuli (Canessa et al., 2012), is consistent with EEG evidence based on the same task and stimuli (Proverbio et al., 2011).

These results are suggestive of a hierarchical neural decoding of observed social interactions. The ventral premotor and superior parietal nodes of the mirror system (Caspers et al., 2010; Rizzolatti & Craighero, 2004) underpin a preliminary process of action recognition, that is, "what" and "how" (Spunt & Lieberman, 2012a, 2012b). These regions are also more strongly recruited when the decoding of intentions expressed by shared action goals, such as in cooperative interactions, requires a more fine-grained processing of the observed agents' motor acts. Instead, the mentalizing system is more strongly involved when social intentions (why), rather than being decoded by observed behaviors, must be inferred by the agent' shared affective states.

Previous inconsistent claims on the role of these two neural systems in processing social interactions might result from the lack of evidence about the impact of these, and possibly other, underlying dimensions. For instance, Donne, Enticott, Rinehart, and Fitzgerald (2011) reported that corticospinal excitability following transcranial magnetic stimulation to the primary cortex (a measure of motor resonance) is not differentially modulated by observing individual goal-directed versus social behavior. They concluded that "*the failure to detect a strong association between mirror system and social behavior suggests that this relationship is either not as strong as predicted, or more complex than predicted*." Importantly, however, their social stimuli depicted affective interactions such as shaking hands (p. 58), which indeed we did not find to elicit a *selective* recruitment of the mirror system. Conversely, one of their two videos eliciting a mirror response displayed "*physical contact and cooperation between actors*." Our data indeed suggest that such response is more strongly modulated by shared motor intentions, for example, cooperative interactions, and highlight the need to unveil the neural dynamics driving the recruitment of the mirror versus mentalizing systems when processing different facets of social interactions.

We pursued this goal with DCM, which aims to explain brain activations by assessing how endogenous neural dynamics are modulated by external perturbations resulting from experimentally controlled manipulations (Friston et al., 2003). In particular, we aimed to highlight the causal organization driving the outcome of a preliminary visuomotor processing of both interaction types into two distinct neural pathways underlying in-depth analyses of shared motor intentions versus mental states.

DCM analyses provided estimates of effective connectivity within a left-hemispheric network including the pSTS, superior parietal, and ventral premotor cortex (reflecting increasing cooperativity) alongside vmPFC (reflecting increasing affectivity). While Network Discovery (Friston et al., 2011) initially highlighted an optimum model characterized by full *endogenous* connectivity, subsequent random-effect analyses at the parameter level constrained the neural architecture connecting the network nodes (Figure 3a). The input pSTS node is the one associated with the greatest number of endogenous connections, and the fact that all the forward connections from this region (to all the other nodes) are excitatory confirms its prominent role in the bottom–up propagation of visual information. Conversely, all the backward endogenous connections, including those from vmPFC and vPMC to pSTS, are inhibitory. Even in the absence of perceptual inputs, this neural architecture appears to underpin the notion of functional asymmetries in hierarchical organization (Chen, Henson, Stephan, Kilner, & Friston, 2009), in which the net effect of backward connections, compared with forward ones, is inhibitory (Angelucci & Bressloff, 2006; Friston, Kahan, Biswal, & Razi, 2014). Importantly, such architecture is well suited for providing the decoding of social intentions with both affective and visuomotor information, via backward inhibitory connections to pSTS from the vmPFC and the other input node in the vPMC, respectively. In addition, the vmPFC node of the mentalizing system is connected to the mirror system either directly (via pSTS) or indirectly (via vPMC). In line with *f*MRI data (Figure 2a and Table 1a), this evidence highlights a strongly interconnected network exceeding the classical "action observation" system associated with individual actions (Gardner et al., 2015), and involving the key nodes of both the mirror and mentalizing systems when observing social interactions.

Within such intrinsic architecture, the functional distinction between the two systems appears to emerge, under visual stimulation, from divergent connectivity patterns involving their key nodes (Figure 3d). Namely, these patterns originate from the different modulations exerted by perceived affectivity or cooperativity on the activity elicited by the driving input, that is, observing social interactions.

This input enters the network both in the pSTS and vPMC. The former region was largely expected as an input, as the processing of biological motion by pSTS (Beauchamp, Lee, Haxby, & Martin, 2002) is considered to provide higher order visual information to the mirror system (Rizzolatti & Sinigaglia, 2010) (Figure 3b,d). In addition, the second input into the vPMC is consistent with recent DCM evidence showing that both this region and the pSTS induce activity in the mirror system when observing *individual* actions (Gardner et al., 2015).

The double input into pSTS and vPMC may support the common recruitment of the mentalizing and mirror systems when processing social interactions (Figure 2a). In particular, both the degree of affectivity and cooperativity exert a positive modulation, reducing its endogenous inhibitory influence, on the backward connection from the vmPFC to the input vPMC node (Figure 3b,c). Alongside an endogenous architecture linking the mirror and mentalizing systems (Figure 3a), this modulation might foster a preliminary neural decoding of social intentions via the processing of both action goals and affective features. This is in line with the need to engage the attribution of mental states, along with the visuomotor extraction of motor intentions, in the case of social interactions (Centelles et al., 2011; Iacoboni et al., 2004).

On the other hand, the different modulations exerted by the degree of cooperativity and affectivity on the endogenous model architecture seem to explain the emergence of two functionally distinct neural systems associated with their processing (Figure 3d and Table 3). To test this hypothesis we assessed the role played by such modulations on causal influence within the whole network. DCM results showed that two computational routes separate at the early stages of the neural processing of social interactions, that is, in the effective connection from pSTS to vmPFC versus both SPL and vPMC (Figure 3d,f).

Such distinction results from oppositely valenced modulations of connectivity by the degree of cooperativity and affectivity. First, cooperativity was found to promote a positive modulation of connectivity within the mirror system nodes. This modulation results both in the increase of the endogenous forward excitatory influence from pSTS to both SPL and vPMC, and in the decrease of the endogenous backward inhibitory influence from vPMC to both SPL and pSTS (Figure 3b,d). While this connectivity pattern explains the involvement of the mirror system, the negative modulation exerted on the reciprocal connection between pSTS and vmPFC fits with the lack of involvement of the mentalizing system at higher cooperativity levels.

Compared with the pattern of modulation elicited by cooperativity, the degree of perceived affectivity exerts an opposite influence on effective connectivity among pSTS, vPMC, and vmPFC (while showing no influence on the connections involving the SPL mirror node). Namely, higher affectivity was found to increase the endogenous forward excitatory influence from pSTS to vmPFC, and to decrease the endogenous backward inhibitory influence from vmPFC to pSTS (Figure 3c,d).

For most connections (with the only exception of pSTS→SPL), the overall net effect of endogenous connectivity and modulatory influences confirmed the oppositely valenced effects of observing social interactions expressing increasing affectivity versus cooperativity (Table 3). This connectivity pattern appears to act as a gateway mediating the preferential engagement of either the mirror or mentalizing systems depending on the strength of, respectively, cooperative versus affective cues expressed by the observed social scene. In addition, the differential influence on connectivity by cooperativity and affectivity is also subject to *oppositely directed* modulations, that is, higher modulation by cooperativity on forward pSTS→SPL and backward vPMC→pSTS connectivity, and by affectivity on backward vmPFC→pSTS connectivity (Figure 3g,h). The former result confirms recent DCM evidence of increased forward connectivity from lower sensory area MT/V5+ to pSTS when observing animate-intentional versus inanimate motion (Hillebrandt et al., 2014).

In the "predictive coding" framework (Koster-Hale & Saxe, 2013), forward connections are considered to reflect the bottom–up propagation of prediction-error signals concerning stimulus-related unexpected sensory information (Hillebrandt et al., 2014). Conversely, upcoming prediction errors are minimized by top–down backward connections carrying refined predictions based on an internal model.

It is thus noteworthy that the *forward* connection from pSTS to SPL was specifically modulated by cooperativity, and that only in this direction such modulation was stronger that than exerted by affectivity. This evidence suggests that cues emphasizing shared action goals (compared with shared mental states) specifically promote stronger forward (compared with backward) causal influence from the pSTS to its "mirror" parietal target in charge of the visuomotor decoding of observed actions. In line with models of action perception based on predictive coding (Gardner et al., 2015; Keysers & Perrett, 2004; Kilner, Friston, & Frith, 2007; Kilner & Frith, 2007), this evidence is suggestive of a bottom–up, perceptually driven, development of a neural representation of shared action goals in observed social interactions.

In contrast, the degree of cooperativity and affectivity exerted stronger *backward*, compared with forward, positive modulations of endogenous effective connectivity from, respectively, vPMC to SPL and pSTS, and vmPFC to pSTS. In predictive coding, top–down connections send back to sensory areas refined predictions based on an internal model of the attended stimuli, to minimize their prediction errors. The stronger backward modulation of these connections with increasing affectivity (vmPFC→pSTS) or cooperativity (vPMC→pSTS; vPMC→SPL) thus likely reflects the higher accuracy of top–down predictions when observed social interactions are strongly characterized in terms of these underlying dimensions.

In addition, the positive modulation exerted by both affectivity and cooperativity on the backward vmPFC→vPMC connection may indicate that enriched signals from vmPFC can then re-access the mirror system via vPMC and further improve the decoding of social intentions (Tidoni & Candidi, 2016). In the case of interactions expressing high affectivity levels, indeed, this connectivity pattern may support a deeper visuomotor analysis performed by the mirror system with information concerning its agents' mental states.

By showing that the processing of social interactions is not confined to either the mirror or mentalizing systems, the present DCM evidence provides novel insights into the neural basis of implicit intention understanding. A limitation of this study is the lack of an explicit control for the processing of individual actions, which in principle does not allow interpreting the observed activations as specific to the processing of social interactions. It is worth noting, however, that all our findings result from the modeling of dimensions which are inherent in social interactions, that is, the degree of affectivity or cooperativity expressed by actions necessarily entailing two interacting individuals. Under this caveat, our results show that such dimensions preferentially recruit distinct, although strongly interconnected, neural pathways associated with the bottom–up visuomotor processing of motor intentions and the top–down attribution of affective and mental states. These insights may prove useful in future studies assessing the status of this neural architecture in conditions characterized by impaired social cognition.

## ACKNOWLEDGMENTS

The authors wish to thank Dr Marco Tettamanti and Dr Benedetta Vai, and two anonymous reviewers, for their valuable comments and suggestions. The authors declare no competing financial interests.

## ORCID

*Nicola Canessa* http://orcid.org/0000-0002-0179-6384

## REFERENCES

Amodio, D. M., Kubota, J. T., Harmon-Jones, E., & Devine, P. G. (2006). Alternative mechanisms for regulating racial responses according to internal vs external cues. *Social Cognitive and Affective Neuroscience*, *1*, 26–36.

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research*, *154*, 93–120.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*, 149–159.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*,

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews. Neuroscience*, *2*, 561–567.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., & Rizzolatti, G. (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron*, *42*, 323–334.

Canessa, N., Alemanno, F., Riva, F., Zani, A., Proverbio, A. M., Mannara, N., . . . Cappa, S. F. (2012). The neural bases of social intention understanding: The role of interaction goals. *PLoS One*, *7*, e42347.

Canessa, N., Motterlini, M., Alemanno, F., Perani, D., & Cappa, S. F. (2011). Learning from other people's experience: A neuroimaging study of decisional interactive-learning. *NeuroImage*, *55*, 353–362.

Canessa, N., Motterlini, M., Di Dio, C., Perani, D., Scifo, P., Cappa, S. F., & Rizzolatti, G. (2009). Understanding others' regret: A FMRI study. *PLoS One*, *4*, e7402.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*, 1148–1167.

Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition*, *36*, 426–433.

Centelles, L., Assaiante, C., Nazarian, B., Anton, J. L., & Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study. *PLoS One*, *6*, e15749.

Chen, C. C., Henson, R. N., Stephan, K. E., Kilner, J. M., & Friston, K. J. (2009). Forward and backward connections in the brain: A DCM study of functional asymmetries. *NeuroImage*, *45*, 453–462.

Chiavarino, C., Apperly, I. A., & Humphreys, G. (2012). Understanding intentions: Distinct processes for mirroring, representing, and conceptualizing. *Current Directions in Psychological Science*, *21*, 284–289.

Chumbley, J. R., & Friston, K. J. (2009). False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage*, *44*, 62–70.

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

Donne, C. M., Enticott, P. G., Rinehart, N. J., & Fitzgerald, P. B. (2011). A transcranial magnetic stimulation study of corticospinal excitability during the observation of meaningless, goal-directed, and social behaviour. *Neuroscience Letters*, *489*, 57–61.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*, 1325–1335.

Enrici, I., Adenzato, M., Cappa, S., Bara, B. G., & Tettamanti, M. (2011). Intention processing in communication: A common brain network for language and gestures. *Journal of Cognitive Neuroscience*, *23*, 2415–2431.

Friston, K., & Penny, W. (2011). Post hoc Bayesian model selection. *NeuroImage*, *56*, 2089–2099.

Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Applications. *NeuroImage*, *16*, 484–512.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Friston, K. J., Kahan, J., Biswal, B., & Razi, A. (2014). A DCM for resting state fMRI. *NeuroImage*, *94*, 396–407.

Friston, K. J., Li, B., Daunizeau, J., & Stephan, K. E. (2011). Network discovery with DCM. *NeuroImage*, *56*, 1202–1221.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *NeuroImage*, *10*, 607–619.

Gardner, T., Goulden, N., & Cross, E. S. (2015). Dynamic modulation of the action observation network by movement familiarity. *Journal of Neuroscience*, *35*, 1561–1572.

Georgescu, A. L., Kuzmanovic, B., Santos, N. S., Tepest, R., Bente, G., Tittgemeyer, M., & Vogeley, K. (2014). Perceiving nonverbal behavior: Neural correlates of processing movement fluency and contingency in dyadic interactions. *Human Brain Mapping*, *35*, 1362–1378.

Hamilton, A. F., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *The Journal of Neuroscience*, *26*, 1133–1137.

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*, 129–140.

Hillebrandt, H., Friston, K. J., & Blakemore, S. J. (2014). Effective connectivity during animacy perception–dynamic causal modelling of Human Connectome Project data. *Scientific Reports*, *4*, 6240.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*, *21*, 1167–1173.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*, e79.

Kass, R. E., & Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association*, *90*,

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, *8*, 501–507.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*, 159–166.

Kilner, J. M., & Frith, C. D. (2007). A possible role for primary motor cortex during action observation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 8683–8684.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, *79*, 836–848.

Kujala, M. V., Carlson, S., & Hari, R. (2012). Engagement of amygdala in third-person view of face-to-face interaction. *Human Brain Mapping*, *33*, 1753–1762.

Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: A meta-analysis. *NeuroImage*, *31*, 1453–1474.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, *40*, 525–543.

Meneghini, A. M., Sartori, R., & Cunico, L. (2006). Adattamento e validazione su campione italiano della balanced emotional empathy scale di a. mehrabian.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience and Biobehavioral Reviews*, *36*, 341–349.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Proverbio, A. M., Riva, F., Paganelli, L., Cappa, S. F., Canessa, N., Perani, D., & Zani, A. (2011). Neural coding of cooperative vs. affective human interactions: 150 ms to code the action's purpose. *PLoS One*, *6*, e22026.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews. Neuroscience*, *11*, 264–274.

Rosa, M. J., Friston, K., & Penny, W. (2012). Post-hoc selection of dynamic causal models. *Journal of Neuroscience Methods*, *208*, 66–78.

Sperduti, M., Guionnet, S., Fossati, P., & Nadel, J. (2014). Mirror Neuron System and Mentalizing System connect during online social interaction. *Cognitive Processing*, *15*, 307–316.

Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *NeuroImage*, *99*, 301–311.

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological Science*, *21*, 1593–1598.

Spunt, R. P., Kemmerer, D., & Adolphs, R. (2016). The neural basis of conceptualizing the same action at different levels of abstraction. *Social Cognitive and Affective Neuroscience*, *11*, 1141–1151.

Spunt, R. P., & Lieberman, M. D. (2012). Dissociating modality-specific and supramodal neural systems for action understanding. *Journal of Neuroscience*, *32*, 3575–3583.

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage*, *59*, 3050–3059.

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: An fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience*, *23*, 63–74.

Tettamanti, M., Vaghi, M. M., Bara, B. G., Cappa, S. F., Enrici, I., & Adenzato, M. (2017). Effective connectivity gateways to the Theory of Mind network in processing communicative intention. *NeuroImage*, *155*, 169–176.

Tidoni, E., & Candidi, M. (2016). Commentary: Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Frontiers in Behavioral Neuroscience*, *10*, 13.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289.

Van den Stock, J., Hortensius, R., Sinke, C., Goebel, R., & de Gelder, B. (2015). Personality traits predict brain activation and connectivity when witnessing a violent conflict. *Scientific Reports*, *5*, 13779.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*, 564–584.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited–again. *NeuroImage*, *2*, 173–181.

Yang, Y., Zhong, N., Friston, K., Imamura, K., Lu, S., Li, M., . . . Hu, B. (2017). The functional architectures of addition and subtraction: Network discovery using fMRI and DCM. *Human Brain Mapping*, *38*, 3210–3225.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Arioli M, Perani D, Cappa S, et al. Affective and cooperative social interactions modulate effective connectivity within and between the mirror and mentalizing systems. *Hum Brain Mapp*. 2017;00:1–16. https://doi.org/10.1002/hbm.23930