


## RESEARCH ARTICLE

# Somatic engagement alters subsequent neurobehavioral correlates of affective mentalizing

Ofir Shany<sup>1,2</sup>  | Ayam Greental<sup>1,2,3</sup> | Gadi Gilam<sup>4</sup> | Daniella Perry<sup>2</sup> |  
Maya Bleich-Cohen<sup>2</sup> | Moran Ovadia<sup>2,5</sup> | Avihay Cohen<sup>2</sup> | Gal Raz<sup>2,3,5</sup>

<sup>1</sup>School of Psychological Sciences, Tel-Aviv University, Tel-Aviv, Israel

<sup>2</sup>Sagol Brain Institute, Tel-Aviv Sourasky Medical Center, Tel-Aviv, Israel

<sup>3</sup>Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel

<sup>4</sup>Division of Pain Medicine, Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University, Stanford, California

<sup>5</sup>Steve Tisch School of Film and Television, Tel Aviv University, Tel-Aviv, Israel

## Correspondence

Ofir Shany and Gal Raz, Sagol Brain Institute, Tel Aviv Sourasky Medical Center, 6 Weizmann Street, Tel Aviv 6423906, Israel. Email: ofirshany@gmail.com (O. S.) and galraz@tauxex.tau.ac.il (G. R.)

## Funding information

BrainTrain Consortium under the EU PF7 Health Cooperation Program, Grant/Award Number: 602186; Israel Science Foundation, Grant/Award Number: 1333/19

## Abstract

Socio-emotional encounters involve a resonance of others' affective states, known as affect sharing (AS); and attribution of mental states to others, known as theory-of-mind (ToM). Empathy necessitates the integration of both processes, yet their interaction during emotional episodes and subsequent generation of inferences on others' affective states has rarely been tested. To address this, we developed a novel experimental design, wherein we manipulated AS by presenting nonverbal emotionally negative movies twice—each time accompanied by one of two soundtracks that accentuated either somatic cues or externally generated sounds. Movies were followed by questions addressing affective-ToM (emotional inferences), cognitive-ToM (inferences on beliefs and knowledge), and non-ToM aspects. Results revealed a neural differentiation between AS, affective-ToM, and cognitive-ToM. AS movies activated regions that have been implicated in emotional (e.g., amygdala) and somato-sensory processing, and synchronized brain activity between participants in the latter. Affective-ToM activated the middle insula, limbic regions, and both ventral and dorsal portions of the medial prefrontal cortex (ventral medial prefrontal cortex [VMPFC] and dorsal medial prefrontal cortex [DMPFC], respectively), whereas cognitive-ToM activated posteromedial and lateral-prefrontal and temporal cortices. Critically, AS movies specifically altered neural activation in AS and ToM-related regions during subsequent affective-ToM inferences, most notably in the DMPFC. Moreover, DMPFC-VMPFC connectivity correlated with affective-ToM accuracy, when such questions followed AS movies. Our results associate empathic processes with designated neural activations and shed light on how neuro-behavioral indices of affective ToM are shaped by preceding somatic engagement.

## KEYWORDS

emotion, empathy, functional MRI, theory of mind

Ofir Shany and Ayam Greental contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

## 1 | INTRODUCTION

As we navigate through the social realm, we often encounter others going through adversities. During such encounters, processes of two distinct types may dominate our empathic reaction: we somatically experience distress with the other and infer on her state. This dual notion of social engagement, which has been recently supported by a line of behavioral (Kanske, Böckler, Trautwein, Parianen Lesemann, & Singer, 2016), evolutionary (de Waal, 2012), neuroimaging (Cox et al., 2012; Kanske et al., 2016; Raz et al., 2014; Valk et al., 2016), clinical (Bird et al., 2010; Bird & Viding, 2014; Blair, 2008; Hendlar et al., 2018), and brain lesion (Shamay-Tsoory & Aharon-Peretz, 2007) evidence, has raised intriguing questions regarding the dynamic interaction between these two processes: does bodily resonance of another's distress reinforce mentalization of her state? And vice versa: does enhanced mentalization effort toward another person necessarily boost empathic susceptibility to her bodily distress cues? Are there conditions under which these processes may compete or suppress each other? Insights on these issues could be relevant for understanding individual differences in empathic processing and various sources of their disturbances in psychopathological cases such as autism, personality disorder, and schizophrenia. The current study extends the scrutiny of the interaction between these processes by elaborating on its temporal dimension. Specifically, we examine whether the magnitude of somatic empathy cues provided when one witnesses another person's distress would affect *subsequent* inferences and reflection on distressful events.

To date, vast neuroscientific research on social cognition and empathy has verified that constructing others' minds and emotions relies on at least two major neuropsychological components (Zaki & Ochsner, 2012) - Affect sharing (AS) and theory of mind (ToM). AS entails the vicarious sharing of another person's emotional state via somatosensory and affective processes (Singer, 2006; Weisz & Zaki, 2018; Zaki & Ochsner, 2012). It is prompted by neural systems supporting emotion perception, such as the sensory cortices and amygdala (Dricu & Frühholz, 2016); salience detection, which involves anterior portions of the cingulate cortex (ACC) and the insula; and action simulation processes, supported by the inferior parietal lobule (IPL) and inferior frontal gyrus (IFG; Shamay-Tsoory, 2011). ToM, on the other hand, refers to the attribution of mental states to others via inferential processes and reasoning (de Waal & Preston, 2017). It engages a system involved in self-reflective and mnemonic processes, including the medial prefrontal cortex (MPFC), superior temporal sulcus (STS), temporo-parietal junction (TPJ), precuneus, posterior cingulate cortex (PCC), and the temporal poles (TP; Buckner & Carroll, 2007; Shamay-Tsoory, 2011; Spreng, Mar, & Kim, 2009; Zaki & Ochsner, 2012). Furthermore, it has been suggested that ToM can be dissected into two subtypes—cognitive and affective. Cognitive ToM denotes the understanding of another's knowledge and beliefs, whereas affective ToM refers to comprehending another's emotional state and desires (Abu-Akel & Shamay-Tsoory, 2011; Molenberghs, Johnson, Henry, & Mattingley, 2016). While both components share major nodes of the classical ToM network, affective

ToM is assumed to involve cortical and limbic regions that associate with AS as well (e.g., ventral portions of MPFC, amygdala, and IFG; Abu-Akel & Shamay-Tsoory, 2011; Leigh et al., 2013; Schmitgen, Walter, Drost, Rückl, & Schnell, 2016).

Given the considerable dissociation between AS and ToM, a key question about the interaction between their underlying networks arises: do these two distinct modes of social engagement substantially correlate so that bottom-up vicarious AS affects top-down mentalization? While this idea is plausible since empathic experiences typically necessitate the integration of both processes, their interaction during emotional episodes and their subsequent cognitive processing has rarely been tested.

Evidence from several studies that addressed AS–ToM interactions indicate that empathy-eliciting situations may have opposing effects on AS- and ToM-related brain function as well as on ToM performance. On the one hand, several studies reported that social cues such as emotional facial expressions (Schmitgen et al., 2016), video-recorded personal monologs (Zaki, Weber, Bolger, & Ochsner, 2009), and verbal emotional movie scenes (Pehrs et al., 2015; Raz et al., 2014, 2016; Regenbogen et al., 2012) enhance activity and connectivity within both AS and ToM-related networks, and increase AS–ToM connectivity and affective ToM performance. On the other hand, other evidence suggests that AS induction during the viewing of emotional movies may interfere with ToM-related activation (Kanske et al., 2016), elicit inverse dynamic connectivity patterns within AS- and ToM-related networks (Raz et al., 2014), and hamper affective ToM performance (Kanske et al., 2016).

One limitation of these studies is their reliance on experimental manipulations that involve mixtures of bottom-up bodily cues and top-down cognitions. That is, several studies induced AS using spoken emotional narratives (Kanske et al., 2016; Mackes et al., 2018; Zaki et al., 2009) or by explicitly instructing participants to “feel with” others (Dziobek et al., 2011; Oliver, Vieira, Neufeld, Dziobek, & Mitchell, 2018; Spunt & Lieberman, 2012). In these cases, AS induction is inevitably mediated by “top-down” verbal representations of emotions. Similarly, emotional facial expressions (Schmitgen et al., 2016) may prompt both affective resonance and interpersonal prediction and inferences (Knutson, 1996; North, Todorov, & Osherson, 2010). Under this limitation, it is possible to observe paralleling or opposing patterns of AS and ToM behavioral and neural correlates. However, these designs hardly allow for the examination of the effect of each type of social information on the correlates of the other. Moreover, previous efforts to delineate both AS and ToM within a single empathy-evoking paradigm did not distinguish between affective and cognitive ToM inferences (Dziobek et al., 2011; Kanske, Böckler, Trautwein, & Singer, 2015). As AS may differentially affect these two types of ToM functions, we believe that an experimental design that entails sensitivity to this distinction would be valuable.

Our study disentangles AS and ToM stimulation by means of temporal separation between these two types of social cues. We test the effect of enhanced AS-related auditory bodily cues on *subsequent* ToM processing in the absence of explicit bodily cues. Specifically, we

examine how AS priming affects ToM correlates during cognitive and affective mentalization separately.

For that aim, we developed a novel naturalistic fMRI paradigm with a designated collection of short videos in which characters go through adversity (Figure 1). Importantly, as mentioned above we intended to induce AS by altering the magnitude of somatosensory cues, rather than manipulating the emotional content of a narrative. This objective was achieved by manipulating the movies' soundtracks, so that the same movie clip was accompanied by a soundtrack that accentuated either somatic sounds (e.g., puking, crying, and scratching), hereinafter termed the Bod-ST (bodily soundtrack) condition; or environmental sounds (e.g., car honking and kettle whistling), hereinafter termed the Env-ST (environmental soundtrack) condition. The alternative soundtracks were matched by loudness and pitch. Each video was followed by a set of questions that required either cognitive ToM, affective ToM, or non-ToM factual recollection (control questions) regarding the preceding video, hereinafter termed Cog-ToM, Aff-ToM, and Control, respectively.

Our study included two behavioral experiments in which we validated the clarity of the ToM questions and the empathic impact of the video stimuli (see Supporting Information Methods). Then, we conducted an fMRI experiment with 30 healthy participants who completed the task in the scanner. This study had two main objectives. First, we aimed to validate the effectiveness of the task in inducing AS and ToM subtypes as indicated by neurobehavioral measures. We hypothesized that Bod-ST movies will exert stronger emotional and empathic impact, relative to Env-ST movies (Hypothesis 1a). Neurally, we expected that Bod-ST movies would activate AS-related regions more strongly than Env-ST movies (Hypothesis 1b). In addition, we predicted that ToM questions will activate the typical ToM-related brain network, relative to Control questions (Hypothesis 1c), and that Aff-ToM questions will activate regions that have been implicated in emotional processing to a larger extent than Cog-ToM questions (Abu-Akel & Shamay-Tsoory, 2011; Leigh et al., 2013; Hypothesis 1d).

Our second objective was to examine whether processing of visceral cues during a socio-affective event would interact with neural function and ToM performance when one reflects on affective aspects of that event following its termination. Given mixed findings regarding the influence of AS on affective ToM, we hypothesized that the presence of salient visceral cues in a socio-affective event (i.e., Bod-ST movies) would either specifically enhance or deteriorate Aff-ToM performance (Hypothesis 2a). At the neural level, we expected that Bod-ST would specifically modify activity/connectivity patterns in brain networks implicated in AS and ToM in a subsequent phase of reflection on this event. In specific, we expected that this effect will be most evident during Aff-ToM, since we assumed that the type of inferences it involves rely on internally simulating emotional processes that are shared by both AS and affective ToM (Abu-Akel & Shamay-Tsoory, 2011; de Waal & Preston, 2017; Schmitgen et al., 2016; Hypothesis 2b). We further suspected that participants' performance in answering Aff-ToM questions will correlate with Bod-ST-induced alterations in AS and ToM neural functionality (Mackes

et al., 2018; Schmitgen et al., 2016; Zaki et al., 2009; Hypothesis 2c). Finally, we also explored if self-reported trait empathy indices associated differentially with the performance and neural correlates of Aff-ToM depending on the preceding soundtrack.

## 2 | METHODS

### 2.1 | Participants

Thirty healthy volunteers ( $M_{\text{age}} = 29.5 \pm 6.3$ , 15 women) participated in the fMRI experiment. The participants had no known history of neurological or psychiatric disorders and provided written informed consent according to the Tel-Aviv Sourasky Medical Center institutional review board (IRB) committee guidelines prior to the experiment. Participants were recruited via online advertisement and received monetary compensation for their time (50 NIS/hr).

### 2.2 | Empathy task

#### 2.2.1 | Affect sharing induction

Fifteen short movies were scripted, filmed, and edited in our lab. These movies depicted scenarios in which characters undergo unpleasant experiences, such as discomfort, pain, or sadness. AS intensity was manipulated by showing the same movie twice, each time adjusting the ratio between the loudness of two sound channels: one channel included auditory cues from a bodily source (e.g., coughs and groans), whereas the other incorporated nonbodily sounds stemming from environmental sources (e.g., car engines, and boiling pot; Figure 1a). Thus, in Bod-ST and the Env-ST versions, the loudness ratios between the channels were set to 90%:10% and 10%:90% in favor of the bodily sound channel, respectively. In addition, we controlled for differences between soundtracks in basic acoustical features as follows. To control for loudness, we equalized the mean RMS between the Bod-ST and Env-ST versions per movie. To control for pitch content, soundtracks were edited so that the same group of pitches appeared within consecutive time windows lasting 3 s each in both versions. For example, in 12–14 s of a certain movie, the same pitches appeared in both its Bod-ST and Env-ST version. This was accomplished by a professional musician (25 years of musical experience) with absolute pitch. Exemplar movies are presented in Supporting Information Videos. Note that the repeated presentation of the same movie in the paradigm may be problematic: this can generally suppress activation and create expectancy violations during the second movie viewing. Moreover, this could potentially alter the affective response to Bod-ST movies as a function of their presentation order. However, this was essential for controlling the narrative and visual features across conditions, and to deal with this issue the presentation order of movies was counterbalanced across participants (see task procedure below).

## 2.2.2 | Theory-of-mind probing

Each movie was followed by yes/no questions regarding its content. Questions were divided into two categories: Control (non-ToM) and ToM questions. Control questions dealt with characters' actions and appearance (Spunt & Adolphs, 2014; e.g., "Is the woman crying?"). ToM questions were divided into affective versus cognitive ToM (i.e., Aff-ToM and Cog-ToM, respectively). While Aff-ToM focused on characters' affective states (e.g., "Is the woman sad?"), Cog-ToM addressed the characters' knowledge or beliefs (e.g., "Does the man know that the woman is crying?"; Figure 1b). The clarity of questions was validated in a preliminary pilot study, and questions that reached an accuracy level below 70% were removed from the final version of the paradigm (see Supporting Information Methods).

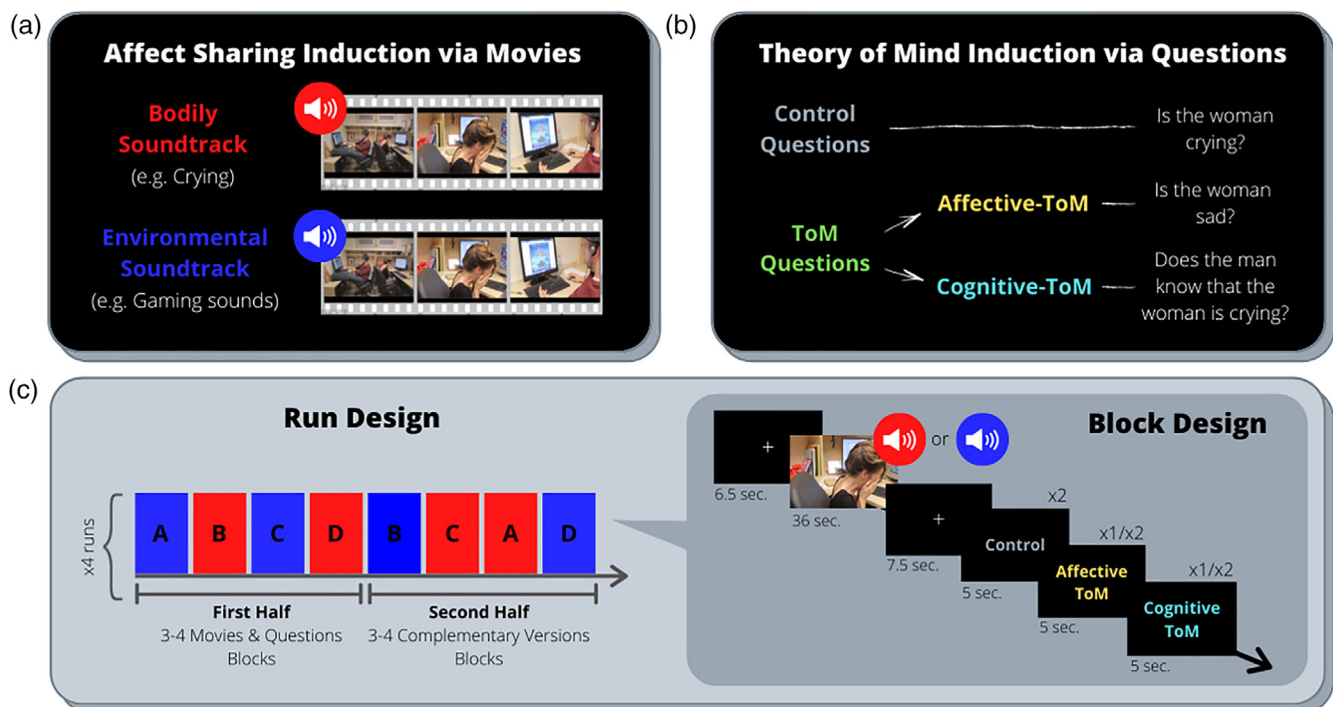
## 2.2.3 | Task procedure

The task was split into four experimental runs for fMRI scanning convenience. Three runs included four movies, and one run included three movies. The movies (36 s) in each run were presented in both versions, Bod-ST and Env-ST. During the first half of each run, 3 and 4 movies appeared in a randomized order (either Bod-ST or Env-ST).

Throughout the second half, their complimentary version was presented, randomized as well. For instance, if a movie was presented in its Env-ST version in the first half, it was presented again in the second half in its Bod-ST version (Figure 1c). Each movie was followed by five questions (5 s each), which comprised of two Control questions and three ToM questions (Aff-ToM and Cog-ToM). The first and fourth questions were always Control questions, and the remaining questions were ToM questions, whose content and presentation order was randomized across the Bod-ST and Env-ST conditions. Participants used an MRI-compatible response box to provide yes/no answers while each question appeared. Fixations appeared before (6.5 s) and after each movie (7.5 s), as well as between each questions, where their duration was jittered in order to last between 2.5–7.5 s (mean = 5 s). The runs including eight movies (4 × 2 soundtrack versions) lasted 13:15 min, and the run including six movies (3 × 2 soundtrack versions) lasted 10:00 min. In total, the fMRI task lasted 49.75 min.

## 2.3 | Trait-empathy questionnaire

To assess trait empathy, we administered the Interpersonal Reactivity Index (Davis, 1983). The IRI is a widely used 28 item self-report



**FIGURE 1** Novel empathy task for AS & ToM induction. (a) Affect sharing (AS) was examined by comparing responses to clips with two alternative soundtracks: Bodily soundtrack (Bod-ST) or environmental soundtrack (Env-ST). (b) Different theory of mind (ToM) subtypes were probed by asking the participant two types of yes/no questions—Control and ToM. ToM questions were divided into two categories: Affective (Aff-ToM) and Cognitive (Cog-ToM). (c) Left panel: Run design. In each run of the task (4 runs in total) 3 and 4 movies and their corresponding ToM and Control questions were presented in the first half of the run in a randomized order (either Bod-ST or Env-ST). Throughout the second half, the movies' complimentary versions were presented, randomized as well. Right panel: Block design. Each movie was followed by five questions (5 s each), which composed of two Control questions and three ToM questions (1x Aff-ToM and 2x Cog-ToM and vice versa in the different session halves)

multidimensional measure of trait-empathy with confirmed validity and reliability (Davis, 1994). It consists of four subscales indexing personal tendencies to adopt another's point-of-view on given situations ("perspective taking"), transposing oneself into the state of fictitious characters ("fantasy"), generally experiencing "other-oriented" feelings of sympathy and concern ("empathic concern"), and feelings of distress in reaction to the aversive emotions of others ("personal distress").

## 2.4 | Experimental procedure

Upon their arrival to the experiment, all participants signed an informed consent and subsequently answered personality questionnaires (~30 min; these included the IRI and additional questionnaires that are not included in this study). Next, participants underwent MRI scanning, during which we acquired anatomical scans and fMRI of the empathy task. Presentation of the four task runs was randomized across participants. During the fMRI scan, participants viewed the movies and listened to their soundtrack through MR compatible headphones (50–15,000 Hz frequency response) with active noise cancellation (Optoacoustics, Israel). For further attenuation of gradient noise, participants used earplugs. The loudness of the sound was fixed for all subjects, and set by previous tests to verify that the participants hear the soundtrack clearly. All visual stimuli were generated on a PC using the E-Prime software (Psychology Software Tools, Pittsburgh).

## 2.5 | Behavioral data analysis

### 2.5.1 | Subjective ratings of movies

To test Hypothesis 1a, according to which the Bod-ST movies will induce stronger empathic and affective responses than Env-ST movies, we conducted an independent behavioral study (see Supporting Information Methods). In brief, in this study participants viewed the movies in both versions and rated them in terms of the experienced level of emotion and empathic engagement.

### 2.5.2 | Accuracy in ToM and Non-ToM questions

Due to the possible ambiguity in ToM judgment, correctness of response to questions was analyzed in relation to the group consensus, rather than by an a-priori definition of "true" or "false" (Schmitgen et al., 2016). Thus, 100% consensus for a question was reached if all participants reported the same answer. A one-sided proportion test fitting our sample size showed that answer consensus could be claimed if it had a 67% probability or more in our sample (binomial test for one proportion;  $z = 1.86$ ,  $p = .031$ ). Thus, questions with less than 67% consensus were excluded from the analyses (12/150 questions). Then, for each participant we calculated accuracy indices for the six question conditions (Control/Aff-ToM/Cog-ToM

questions, following Bod-ST/Env-ST). To test our Hypothesis 2a, which stated that ToM performance will be affected by the somato-affective manipulation, we compared between ToM accuracy after Bod-ST versus Env-ST movies for both ToM types using pairwise Wilcoxon signed ranks tests. In addition, we used the accuracy indices in order to test their relation with neural measurements (Hypothesis 2c), as well as with IRI scores.

## 2.6 | Functional magnetic resonance imaging

### 2.6.1 | MRI acquisition & pre-processing

All scans were performed using a Siemens 3T Prisma Magnetom VD13 echo speed scanner with a 20-channel head coil located at the Wohl Institute for Advanced Imaging at the Tel-Aviv Sourasky Medical Center. Structural scans included a T1-weighted 3D axial spoiled gradient echo (SPGR) pulse sequence (time repetition [TR]/TE = 7.92/2.98 ms, flip angle = 15°, pixel size = 1 mm, FOV = 256 × 256 mm, slice thickness = 1 mm). Functional whole-brain scans were performed in interleaved order with a T2\*-weighted gradient echo planar imaging pulse sequence (TR/TE = 2500/35 ms, flip angle = 90°, pixel size = 1.56 mm, FOV = 200 × 200 mm, slice thickness = 3.1 mm, 38 slices per volume). Anatomical SPGR data were standardized to 1 × 1 × 1 mm and transformed into Talairach space. The preprocessing of the functional data was performed using BrainVoyager QX version 2.1.4. Head motions were detected and corrected using trilinear and sinc interpolations, respectively, applying rigid body transformations with three translation and three rotation parameters. SPGR images were then manually coregistered with the corresponding functional maps. Several datasets were excluded due to exaggerated head movements (>2.5 mm) from the analysis of each run. Data of six participants were excluded from run-1, of five participants from run-2, of five participants from run-3 and of five participants from run-4.

### 2.6.2 | fMRI data analysis—Neural activation

*Construction of general linear models for testing effects of movies' soundtrack (AS) and question types (ToM)*

In order to prevent overfitting of the general linear model, we constructed two separate random-effects general linear models (RFX-GLM) on the fMRI signal. One GLM was used for assessing the neural responses induced by the movies, and the second GLM was created for estimating the neural responses induced by the questions. In the movies GLM, a total of 8/10 regressors were defined, depending on the number of movies in the run: one regressor for each movie (25 s; six/eight regressors in total); and two regressors of no interest—one for the overall questions' phase (45 s), and one for the first 11 s of each movie. The beginning of each movie was excluded from the analysis, since it provided a narrative exposition that did not consist of emotional content. Specifically, the exact onset of emotional content (i.e., the appearance of a noticeable difference between soundtracks)



varied slightly between movies, but generally ranged from 9 to 11 s relative to the movie onset. Thus, to maintain consistency in the analysis we excluded the first 11 s from all movies. Fixations before (6.5 s) and after (7.5 s) each movie served as implicit baseline for the GLM. All predictors were convolved with a canonical hemodynamic response function. To address the neural correlates of the question type, as well as the interaction between the movies' soundtrack and the following questions' type, the questions' GLM was defined using seven regressors—six regressors for question type (Control/Aff-ToM/Cog-ToM; by movie type: following-Bod-ST/following-Env-ST); and one regressor for the overall movies phase. Fixations between each question (randomized jitter, 2.5–7.5 s) and before the movie (6.5 s) served as implicit baseline for the GLM. The length of each question event was 5 s, while the duration of the movies phase was 43.5 s (36 s long movie and the following 7.5 s fixation phase). Six estimates of the motion correction parameters were added as confound regressors to both GLMs as well. Furthermore, to account for brain activation related to response time during questions, we also added a parametric modulator capturing the response time to questions as a confound regressor (i.e., in the questions' GLM).

#### Gray matter mask for whole-brain analysis

We confined all whole-brain analyses to 42,309 voxels included in a gray matter mask. The mask was created by thresholding ICBM 452 map (<http://www.loni.usc.edu/atlases>) to exclude voxels with probability lower than 75% of being classified as gray matter (thus encompassing both cortical and brain stem regions).

#### Whole-brain analysis of neural activation in response to movies and questions

To detect which brain areas showed sensitivity to the AS and ToM induction (Hypotheses 1b–d), we ran two RFX-GLM—one for movies and one for questions (see Supporting Information Methods). In the movies contrast, the effect might vary between the different movies. Thus, we performed a first- and second-level analysis that increased sensitivity to such differences. At the first-level analysis, we created contrast maps for both versions of each of the 15 movies (i.e., a Beta map wherein a specific movie is contrasted against implicit baseline per run). This yielded 30 Beta maps per subject—15 for Bod-ST and 15 for the complimentary Env-ST movies. Next, in order to contrast between movies at the subject-level, we performed a paired samples *t*-test between the Beta values of the 15 Bod-ST movies and the Beta values of the complimentary 15 Env-ST movies in each voxel. This yielded a *t*-value in each voxel, representing the contrast between both movie versions for a specific subject. To calculate group-level statistics, we submitted the *t*-values in each voxel to a one-sample *t*-test. To identify which brain areas showed sensitivity to ToM relative to Control questions (Hypothesis 1c) and to Aff-ToM versus Cog-ToM (Hypothesis 1d), we computed the corresponding contrasts based on the questions' RFX-GLM; (1) Aff-ToM + Cog-ToM > Control questions; (2) Aff-ToM > Cog-ToM, and submitted all contrasts to a one-sample *t*-test at the group-level. The statistical threshold of all whole-brain analyses was set at voxel-level  $q(\text{FDR}) < .05$ .

#### Inter-subject correlation during movies

A common model-free approach to examine fMRI data acquired during movie viewing is the intersubject correlation (ISC) analysis (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004). This method investigates cross-subject similarity between time-courses of BOLD responses per voxel. We used it as a complementary analysis to our GLM comparison of the responses to the movies with the two alternative soundtracks. ISC analysis may unravel cases where the neural activity was synchronized, but not necessarily enhanced, by the movies.

To examine the difference between the movie versions in terms of ISC, we first computed Pearson correlation for every subject *s*, item *i*, and soundtrack version *t*, and voxel *v*. We extracted the time course *TC* of BOLD signal recorded during the clip (after removal of the first four volumes during which the sound difference was minor as described above). For every soundtrack condition, voxel, and item, we computed all sets of correlation between the individual subject's *TC* and the average *TC* of all other subjects:  $ISC_{s,i,t,v} = r(TC_{s,i,t,v}, \overline{TC_{all-s,i,t,v}})$ . The ISC coefficients were subjected to Fisher's transformation, resulting in approximately normally distributed values.

Next, we averaged the transformed ISC values across items per subject and soundtrack condition:  $MISC_{s,t,v} = \overline{ISC_{s,i,t,v}}$ . We then computed Student's *t*-statistics for comparison between these averaged ISC values observed in the two different soundtrack conditions:  $t(MISC_{bodily\ soundtrack,v}, MISC_{environmental\ soundtrack,v})$ . To assess the significance of the difference between the conditions, we generated voxel-wise null distributions of corresponding *t*-statistics. For each subject and item, we randomized the assignment of the soundtrack label and computed the *t*-statistics on the shuffled data. We repeated this procedure to generate null distribution of 10,000 *t*-values per voxel. For each voxel, the absolute value of *t*-statistics obtained in the original contrast was then compared with the absolute value of the *t*-statistics in the null distribution to yield a two-sided *p*-value. The resulting *p*-values were corrected for multiple comparisons using FDR correction at  $q(\text{FDR}) = .05$ . Comparisons were limited to gray matter voxels using the mask mentioned above.

#### Movie by question interaction—ROI analysis

In order to test Hypothesis 2b, stating that the neural activation during Aff-ToM questions will be modulated as a function of the preceding soundtrack, we tested the movie type by question type interaction in key regions of the ToM and AS networks.

**ROI definition.** The theoretical definition of key representative regions of the ToM network was based on an influential empathy review (Zaki & Ochsner, 2012), and included seven regions: the DMPFC, VMPFC, precuneus, and the bilateral TPJ and TP. Coordinates for the ROIs were extracted from a recent meta-analysis on ToM (Molenberghs et al., 2016). Definition of key representative regions of the AS network was based on Shamay-Tsoory (2011) and included the anterior insula/IFG and the ACC. Coordinates for the AS ROIs were extracted from a recent meta-analysis on empathy for pain (Table S2; in Timmers et al., 2018). Coordinates of all ROIs were transformed from MNI space into Tailarach space using the `icbm_spm2tal`

function (<http://brainmap.org/icbm2tal/>), and are presented in Table S2. All ROIs were finally generated by defining a 6 mm sphere around the chosen coordinates. Note that all ROIs also resided within the abovementioned gray matter mask.

*Movie type by question type interaction.* We performed a  $3 \times 2$  repeated-measures ANOVA with question type (Aff-ToM/Cog-ToM/Control) and movie type (following-Bod-ST/following-Env-ST) as within-subject factors in each ROI. Mean Beta values from all six question conditions were extracted from each ROI ( $7 \times$  ToM,  $3 \times$  AS), and entered into 10 separate ANOVAs. Multiple comparisons were corrected by applying a  $q(\text{FDR}) < .05$  threshold on  $p$ -values of all interactions. Post hoc tests for detecting significant pairwise differences between question types were carried only in ROIs in which a significant interaction effect was observed. Significance threshold was set to  $p < .05$  Tukey-HSD corrected within each ROI. This analysis was implemented in Statistica 10 (StatSoft, Tulsa, OK).

#### *Association of brain activation during questions with ToM performance and trait empathy*

Hypothesis 2c stated that neural alterations induced by the Bod-ST movies will correlate with performance on ToM questions, most specifically with Aff-ToM. Thus, we tested if neural activity during questions that followed the Bod-ST condition correlated positively with Aff-ToM accuracy. Specifically, we tested the rank correlation (Spearman's Rho) between DMPFC (see Table S2 for coordinates) activation during Aff-ToM and Aff-ToM performance. We focused on this ROI since meta-analytic evidence recurrently associate it with mentalizing about others (Denny, Kober, Wager, & Ochsner, 2012; Lieberman, Straccia, Meyer, Du, & Tan, 2019). Moreover, among the seven tested ToM ROIs, activity in this ROI during Aff-ToM questions was most selectively affected by Bod-ST movies (see Section 3). We used Spearman's rank correlation, given the non-normal distribution of questions' accuracy scores and frequency of ties in these indices. For each subject, we extracted the mean Beta value for each relevant condition in the relevant ROIs, and correlated the values with individual differences in Aff-ToM accuracy. In addition, we also explored whether IRI scores associated differentially with DMPFC activity during Aff-ToM questions following Bod-ST versus Env-ST.

### 2.6.3 | FMRI data analysis—psychophysiological interaction (PPI) analysis

#### *Whole-brain analysis of DMPFC functional connectivity during questions*

Hypothesis 2b also suggested that functional connectivity patterns during Aff-ToM questions would alter as a function of the preceding movie soundtrack. At the first-level of the PPI analysis, the DMPFC mentioned in the ROI analysis above was defined as a seed region for the abovementioned reasons. The PPI's GLM design matrix included the six original regressors of all question conditions (i.e., Control/Aff-ToM/Cog-ToM questions, following either a Bod-ST or a Env-ST

movie), a regressor of the physiological variable (i.e., the time course of activity in the seed ROI) and six regressors representing the interaction of the time series of the ROI with each of the experimental conditions. The six movement regressors were added to the GLM as well. All PPI analyses were carried out using an in-house MATLAB-based software (Gilam et al., 2015). We examined the effect of movie type by question type interaction on DMPFC connectivity by conducting a whole-brain random-effects ANOVA with question type (Aff-ToM/Cog-ToM/Control) and movie type (following-Bod-ST/following-Env-ST) as within-subject factors. The resulting statistical map was thresholded at voxel-level  $q(\text{FDR}) < .05$ .

#### *Association of functional connectivity during questions with ToM performance*

Our set of hypotheses further suggested that functional connectivity of AS and ToM regions during questions (Hypothesis 2c) will differentially associate with ToM performance depending on the preceding movie type. We specifically examined if functional connectivity of ToM regions during Aff-ToM questions that followed Bod-ST movies, correlated with the corresponding Aff-ToM performance. To test this, we extracted four DMPFC PPI contrasts of interest for each participant from the above-defined first-level GLM of the DMPFC's PPI—one per ToM question condition (vs. baseline)—that were next submitted to a two-staged second-level analysis.

In the first stage of the second-level analysis, we computed the correlation (Spearman's Rho) between connectivity indices during each of the four question conditions and the corresponding accuracy level. This correlation was computed in each voxel, thus resulting in four group-level  $\rho$ -value maps—one for each question condition. In each map, the  $\rho$ -value in a certain voxel represents the correlation strength between DMPFC connectivity during a specific question type and individual differences in accuracy levels. For example, in the question condition “Aff-ToM following-Bod-ST,” the  $\rho$ -value in an exemplar voxel reflects the degree to which connectivity of the DMPFC with that voxel during Aff-ToM questions that followed Bod-ST correlated with the accuracy of Aff-ToM answers. For the upcoming analysis, the  $\rho$ -values in each voxel in each of the four question condition maps were Fisher-Z transformed.

In the second stage of the second-level analysis, our intention was to detect DMPFC connectivity patterns whose contribution specifically to the Aff-ToM accuracy following-Bod-ST movies was significantly stronger than their contribution to other question types. We examined this by comparing the correlation coefficients of the Aff-ToM following-Bod-ST condition to coefficients derived for two other PPI-accuracy correlations: (a) Aff-ToM following-Env-ST (this comparison highlights the effect of soundtrack within the Aff-ToM condition); (b) Cog-ToM following-Bod-ST (this comparison emphasizes the specificity of the Bod-ST effect on Aff-ToM questions, as opposed to Cog-ToM). We tested the correlation coefficients' differences using a statistical test for comparing correlated but nonoverlapping correlations (Raghunathan, Rosenthal, & Rubin, 1996). This test yielded a Z-value in each voxel, which indicated the strength of difference between a certain pair of PPI-accuracy correlations. A voxel-level FDR-corrected  $p < .05$  threshold was then imposed on these Z-values.

### 3 | RESULTS

#### 3.1 | Induction of AS and ToM in the novel paradigm via movies and questions

##### 3.1.1 | Neurobehavioral indices of AS

Hypotheses 1a and b suggested that Bod-ST movies will exert stronger emotional and empathic impact as indicated by subjective ratings and neural activations in AS-related regions, relative to Env-ST movie. In accordance with Hypothesis 1a, results from an independent behavioral experiment (see Supporting Information Methods) showed that Bod-ST movies were rated higher than Env-ST movies in terms of AS ( $t[23] = 2.29, p = .031$ ), arousal ( $t[23] = 3.22, p = .004$ ), unpleasantness ( $t[23] = 4.38, p < .001$ ), prosocial concern ( $t[23] = 3.28, p = .003$ ) and also mentalizing ( $t[23] = 2.35, p = .027$ ) (Figure 2a).

Consistent with Hypothesis 1b, a whole-brain analysis of the Bod-ST > Env-ST movies contrast revealed that Bod-ST movies induced stronger activations in some of the hypothesized regions, such as dorsal amygdala, left IFG/anterior insula, STG, and the dorsal-posterior and ventral-posterior/middle portions of the insula. Additional activations were evident in thalamic nuclei (medial dorsal nucleus, pulvinar, ventrolateral nucleus, and mammillary), occipital lobes, fusiform/lingual gyrus, and the cerebellum ( $q(\text{FDR}) < .05$  corrected; upper panel in Figure 2b and Table S3). Some of these results were further corroborated by a comparison of the inter-subject correlation (ISC) values between the two soundtrack conditions. This analysis revealed that Bod-ST movies induced stronger ISC than Env-ST movies in clusters in the bilateral STG, and also in a small cluster in the right postcentral gyrus ( $q(\text{FDR}) < .05$  corrected; lower panel in Figure 2b, and Table S6). Note that the Env-ST > Bod-ST contrast did not yield and significant effects in terms of brain activity or ISC.

##### 3.1.2 | Neural effects of ToM and segregation of affective versus cognitive ToM

Hypothesis 1c stated that ToM questions will activate a typical ToM-related brain network, relative to Control questions. Indeed, a whole-brain analysis of the ToM > Control Questions contrast revealed widespread activations in a set of brain regions that corresponds with meta-analytic definitions of the ToM network (Molenberghs et al., 2016; Spreng et al., 2009). This included the bilateral TPJ, posterior STS, TP, IFG, and midline areas such as the DMPFC, VMPFC, and PCC-precuneus ( $q(\text{FDR}) < .05$  corrected; Figure 2c and Table S4).

In accordance with Hypothesis 1d, the whole-brain analysis of the Aff-ToM > Cog-ToM questions contrast revealed a clear neural differentiation between the two question types. Aff-ToM associated with enhanced activations in limbic and action-simulation-related regions relative to Cog-ToM, including the middle insula, IFG, caudate and left ventral striatum, and the IPL. Additional activations were evident in medial and lateral prefrontal regions, such as the dorsal and ventral MPFC, ACC, and middle frontal gyrus. In contrast, Cog-ToM

questions engaged more posterior and dorsal portions of the ToM network, including regions such as the TPJ, posterior and middle STS, IFG, middle frontal gyrus, and lingual gyrus; the PCC-precuneus; and also thalamic, occipital, hippocampal, and parahippocampal gyri ( $q(\text{FDR}) < .05$  corrected; Figure 2d and Table S5).

#### 3.2 | Effects of AS induction on subsequent neurobehavioral ToM indices

##### 3.2.1 | Effect of AS induction on ToM accuracy

To examine whether the AS induction influenced ToM accuracy (hypothesis 2a), we examined if the accuracy within each question type differed when questions appeared after Bod-ST versus Env-ST movies. At odds with our hypothesis, accuracy in Aff-ToM questions following Bod-ST movies (median = 94.4%) was not different from accuracy following Env-ST movies (median = 95.7%; Wilcoxon signed ranks test:  $z = -1.14, p = .25$ ). Accuracy also did not differ between Cog-ToM questions that followed the Bod-ST (median = 91.4%) and those that followed Env-ST (median = 94.7%;  $z = -.75, p = .45$ ). In addition, an exploration of correlations between Aff-ToM performance after each soundtrack type and self-reported trait empathy (indexed by the IRI) did not reveal any significant results (Table S9).

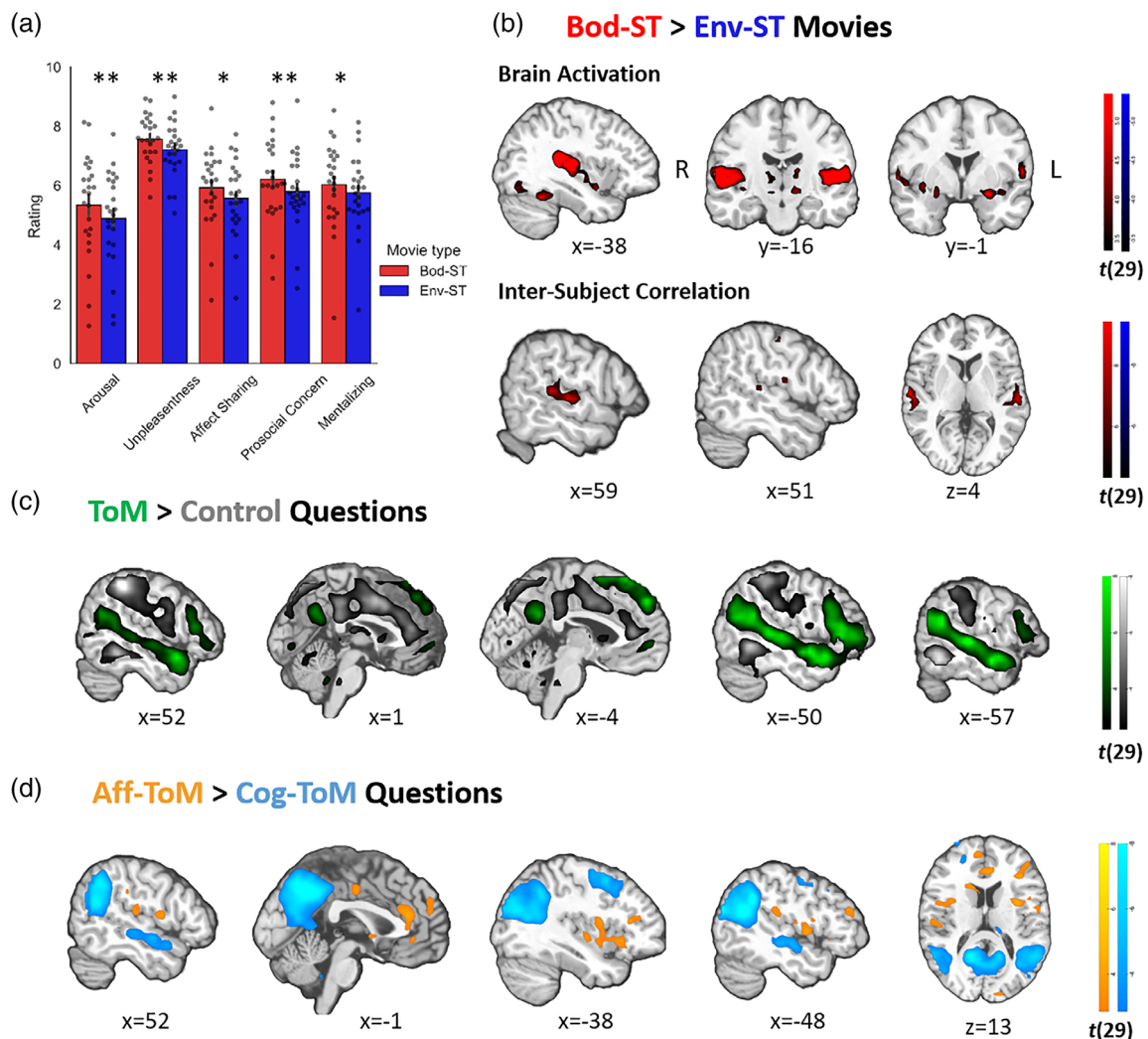
##### 3.2.2 | Neural activity during affective ToM inferences alters as a function of preceding AS induction

Hypothesis 2b posited that AS induction would modify activity and/or connectivity patterns in brain networks implicated in AS and ToM during a subsequent question phase, most specifically during the Aff-ToM condition. To test this hypothesis, we examined the interaction between question type (Aff-ToM/Cog-ToM/Control) and movie type (following-Bod-ST/Env-ST movie) in ToM- and AS-related ROIs (Figure 3a, ROIs coordinates are detailed in Table S2).

The ROI analyses revealed significant movie by question type interaction effects in both ToM and AS regions (Figure 3b). In the ToM-related network, a significant interaction was evident in the DMPFC [ $F(2,58) = 5.46, p = .007, \eta^2 = 0.16, q(\text{FDR}) < .05$ ]. In the AS-related network, we observed a significant interaction in the right IFG-anterior insula [ $F(2,58) = 6.51, p = .003, \eta^2 = 0.18, q(\text{FDR}) < .05$ ]. No significant interactions were found in the remaining AS and ToM ROIs (full statistical details are reported in Table S7).

In accordance with our hypothesis, post hoc pairwise comparisons revealed that the interaction pattern was mainly driven by an effect of the Bod-ST manipulation on Aff-ToM questions. In both the DMPFC and right IFG-anterior insula, activation during Aff-ToM questions was higher during their appearance after Bod-ST movies, relative to Env-ST (DMPFC:  $p = .006$ ; right IFG-anterior insula:  $p = .04$ ). In addition, in the DMPFC, a stronger activation during Aff-ToM compared to the other two question conditions was evident





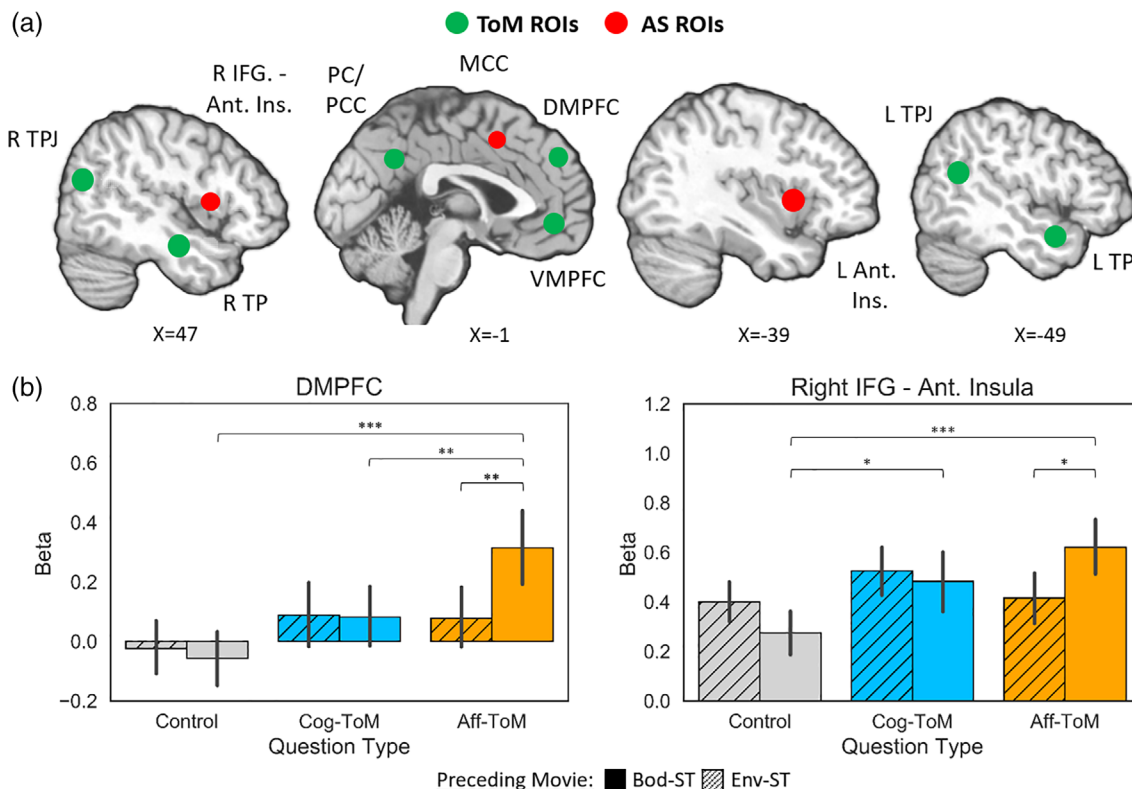
**FIGURE 2** Neurobehavioral effects of AS & ToM induction. (a) Bod-ST movies associated with higher subjective ratings of empathic and negative emotional engagement, relative to Env-ST movies. Note that these ratings were acquired in an independent behavioral study (see Supporting Information Methods). Significance of pairwise comparisons is indicated by asterisks as follows: \*  $p < .05$ , \*\*  $p \leq .005$ , FDR corrected. (b) Whole-brain analysis of the Bod-ST > Env-ST movies contrast is presented. We examined differences between movie types in terms of both brain activity (upper panel) and inter-subject correlation (ISC; lower panel). All regions in which a significant difference in brain activity/ISC was found are listed in Tables S3/S6, respectively. (c) A whole-brain map for the ToM > Control questions contrast is presented. Regions showing a greater response to ToM are colored in green, whereas regions showing a greater response to Control are colored in gray. All regions in which a significant difference was found are listed in Table S4. (d) A whole-brain map for the Aff-ToM > Cog-ToM questions contrast is presented. Regions showing a greater response to Aff-ToM are colored in orange, whereas regions showing a greater response to Cog-ToM are colored in light blue. All regions in which a significant difference was found are listed in Table S5. AS, affect sharing; L, left; R, right; ToM, theory of mind; . All statistical thresholds are set at a voxel-level  $q(\text{FDR}) < .05$

after Bod-ST movies (Aff-ToM vs. Cog-ToM,  $p = .007$ ; Aff-ToM vs. Control,  $p < .001$ ), but not after Env-ST movies. All pairwise comparisons are Tukey-HSD  $p < .05$  corrected. Significant interactions and pairwise comparisons are depicted in Figure 3b. We next examined if functional connectivity during questions was modulated by the preceding soundtrack (this was also part of Hypothesis 2b). We focused this analysis on the DMPFC, given the central role of this region in mentalizing about others' minds (Denny et al., 2012; Lieberman et al., 2019), and since the interaction analysis above showed that Aff-ToM was most selectively affected by the Bod-ST in this region (Figure 3b). Contrary to our hypothesis, a whole-brain

analysis of the effect of the question type by movie type interaction on DMPFC connectivity did not reveal any significant effects.

### 3.2.3 | DMPFC functional connectivity following AS induction correlates with affective ToM performance

In our next analysis, we examined the hypothesis that performance in Aff-ToM questions will correlate with Bod-ST-induced alterations in AS and ToM neural functionality (Hypothesis 2c). We focused this



**FIGURE 3** Activity in ToM- and AS-related regions during questions is modulated by prior AS induction. (a) All ROIs in which the interaction of question type with movie type was examined. AS-related regions are colored in red and ToM-related regions are colored in green. (b) A significant interaction of question type (Aff-ToM/Cog-ToM/Control; in orange, light blue and gray, respectively) with movie type (preceded by Bod-ST vs. Env-ST movie; color-filled, and stripe-filled, respectively) was evident in the DMPFC and right IFG-anterior insula. Ant., anterior; AS, affect sharing; DMPFC, dorsal medial prefrontal cortex; IFG, inferior frontal gyrus; L, left; MCC, middle cingulate cortex; PC, precuneus; PCC, posterior cingulate cortex; R, right; ROI, region of interest; TP, temporal pole; TPJ, temporoparietal junction; ToM, theory of mind. Significance of pairwise comparisons is indicated by asterisks as follows: \*  $p < .05$ , \*\*  $p \leq .01$ , \*\*\*  $p < .001$ , Tukey-HSD corrected

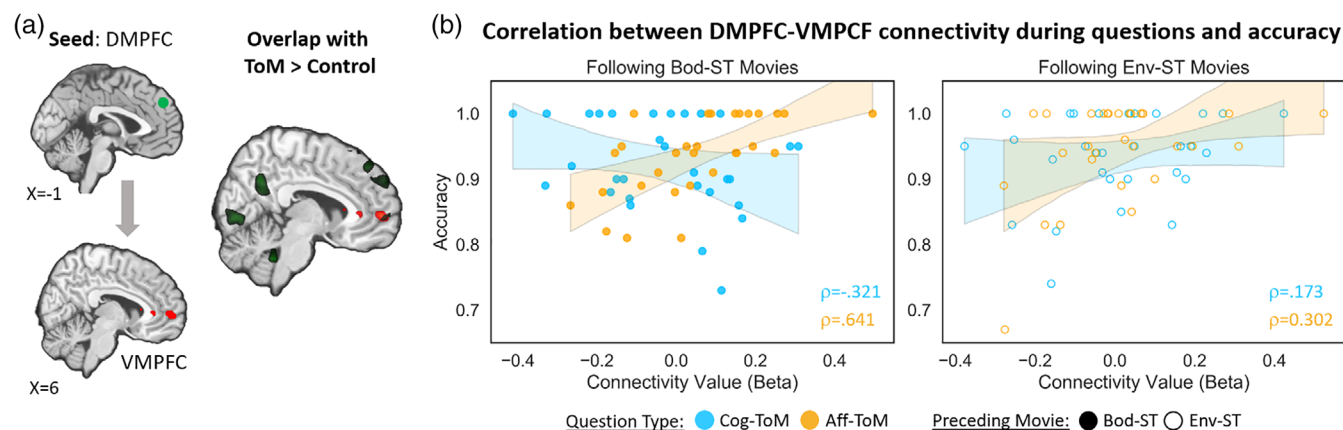
analysis on the DMPFC as well, due to the same reasons mentioned in the previous paragraph.

We first tested the correlation between DMPFC mean activation (Beta value) during Aff-ToM Questions following-Bod-ST movies and the accuracy of Aff-ToM answers, which was non-significant ( $r[28] = -.17$ ,  $p = .38$ ). There were also no significant correlations between DMPFC activity during Aff-ToM and IRI scores (Table S10). Further, we tested if indices of functional connectivity (PPI) of the DMPFC during Aff-ToM questions following-Bod-ST movies associated with accurate performance in these questions. We found that better performance in Aff-ToM following-Bod-ST movies associated with enhanced functional connectivity of the DMPFC seed region to a cluster in the VMPFC, relative to the PPI-accuracy correlation obtained for Cog-ToM following-Bod-ST ( $q(\text{FDR}) < .05$ ; see Figure 4a and Table S8). Notably, this cluster overlapped with activity elicited by the ToM > Control contrast in the VMPFC, thereby corroborating the association of this region with ToM processing (Figure 4a). To explore the direction of the PPI-accuracy correlations underlying this difference, in a post hoc analysis we plotted the correlation of connectivity values in the VMPFC cluster with corresponding accuracy under each of the four ToM conditions (Figure 4b). This depiction

confirmed that the PPI-accuracy correlation was indeed led by a robust positive correlation for the Aff-ToM following-Bod-ST condition, which was also descriptively higher than the correlations observed for ToM questions of both types that followed Env-ST movies (Figure 4b, left panel). The whole-brain comparison of the Aff-ToM following-Bod-ST condition against Aff-ToM following-Env-ST did not reveal significant differences.

## 4 | DISCUSSION

In the current study, we presented a novel naturalistic paradigm designated for examining interactions between distinct empathic processes; namely, AS and affective versus cognitive ToM. By inducing spontaneous AS via emotional movies prior to an explicit ToM inferential phase, we also found neural representations of affective ToM that were affected by the presence of AS cues in the encoded event. Below, we discuss how our results shed light on the nature of interaction between AS and ToM processes, and especially on the formation of affective ToM inferences following stronger empathic experiences.



**FIGURE 4** DMPFC connectivity during questions following Bod-ST movies associates with Aff-ToM performance. (a) The DMPFC ROI wherein we found a question by movie interaction (see Figure 3) was defined as a seed region for testing the association between brain connectivity (PPI, psychophysiological interactions) during questions and ToM performance. The whole-brain map showed that connectivity of the DMPFC to a cluster in the VMPFC during Aff-ToM questions that followed Bod-ST movies correlated with better performance in these questions, relative to the correlation found for Cog-ToM questions presented after Bod-ST movies. Statistical threshold is set at voxel-level  $q$  (FDR)  $< .05$  (presented at  $p < .001$  uncorrected, for display purposes). Notably, the VMPFC cluster found in the PPI analysis overlapped with an area of the VMPFC that was activated by the ToM  $>$  Control contrast (presented in green). (b) Scatterplots depicting the correlations between DMPFC-VMPFC connectivity during Aff-ToM (orange) and Cog-ToM (light blue) questions with accuracy level, when questions followed either Bod-ST (middle panel; full dots) or Env-ST (right panel; empty dots) movies. The mean connectivity values were derived from the VMPFC cluster for each condition, and are presented solely for illustrative purposes along with the 95% confidence interval of the regression line (note that the analysis was based on contrasting between Spearman correlations) and with their corresponding Spearman's Rho value. AS, affect sharing; DMPFC, dorsal medial prefrontal cortex; ToM, theory of mind; VMPFC, ventral medial prefrontal cortex

#### 4.1 | The neural correlates of AS, cognitive ToM, and affective ToM—Corroboration and integration

In accordance with Hypotheses 1a and b, we found that the Bod-ST movies elicited stronger socio-affective engagement (Figure 2a) and neural activation in some of the hypothesized AS-related regions (Figure 2b). The latter was evident in a collection of cortical and sub-cortical regions supporting sensory processing (STG, fusiform, and thalamic nuclei), salience detection (anterior insula/operculum), emotional processing (amygdala) and generation of motivationally directed behavior (basal ganglia and cerebellum; de Waal & Preston, 2017; Zaki, 2013). A complementary ISC analysis corroborated these results to some extent by showing that Bod-ST movies enhanced inter-subject synchronization in the bilateral STG, relative to Env-ST. The ISC analysis also revealed enhanced ISC in the postcentral gyrus, a region which is associated with motor-related functions. The functional roles of the regions specified above are important for tuning to others' distress and forming internal representations of their affective states, and enhanced activation in most of these regions was indeed reported by previous meta-analyses on AS (Fan, Duncan, de Greck, & Northoff, 2011; Lamm, Decety, & Singer, 2011; Timmers et al., 2018). Intriguingly, a robust activation was evident in posterior-middle divisions of the insula, rather than in its hypothesized salience-related anterior part. However, evidence from meta-analyses do associate these insular portions with somatosensory and affective processes, respectively (Chang, Yarkoni, Khaw, & Sanfey, 2013; Kurth, Zilles, Fox, Laird, & Eickhoff, 2010). Thus, the results allude to the

involvement of somato-affective circuits in participants' response to characters' distress during Bod-ST movies, relative to Env-ST.

The greater involvement of auditory and visual processing regions such as the STG and fusiform gyri in response to the Bod-ST movies is noteworthy, as we controlled for acoustic features and maintained the visual properties identical across AS conditions. The STG, and in particular the primary and secondary auditory cortices located in Brodmann areas 41–42 and 22, gives rise to emotionally meaningful auditory percepts throughout unfolding of sounds by operating in concert with limbic and prefrontal regions (Frühholz, Trost, & Kotz, 2016; Salimpoor et al., 2013). Visual processing related activation has been previously associated with AS, but this may have stemmed from comparing negative and salient visual stimuli against neutral stimuli (Timmers et al., 2018). In contrast, here we found greater activation in the fusiform gyrus for identical visual stimuli. Together, these sensory activations may reflect greater allocation of multisensory resources in response to salient socio-emotional cues. Interestingly, despite the recurrent association of sensory regions with empathy and emotional processing, these are not taken into account in influential neural models of empathy (Shamay-Tsoory, 2011; Zaki & Ochsner, 2012). Our results suggest that such regions could be ascribed a more active role in future neural models of empathy.

Consistent with Hypothesis 1c, brain activation during ToM questions associated with a typical ToM-related network relative to control questions (Figure 2c) (Bzdok et al., 2012; Molenberghs et al., 2016; Spreng et al., 2009; Van Overwalle, 2009), thereby validating the tasks capability of probing ToM activity. Amid the

notorious replication crisis, it is not trivial that the finding on the neural correlates of the notion of mentalization is consistently replicated across studies, nationalities, and operationalization. Interestingly, relative to ToM questions, the Control questions activated a posterior cingulo-parietal network (see Supporting Information for further discussion). In accordance with Hypothesis 1d, we found a neural differentiation between Aff-ToM and Cog-ToM questions (Figure 2d). Specifically, these results point to a segregation within the ToM network, wherein lateral-temporal and frontoparietal regions are associated with cognitive processing, whereas MPFC and IFG are linked to affective mentalizing.

An overview on results emerging from both the movies contrast and the Aff-ToM versus Cog-ToM contrast reveals an intriguing activation pattern along the insula: While Bod-ST movies activated the posterior insula, Aff-ToM activated the dorsal parts of the middle and anterior insula (Figure 2d). This corresponds with a theoretical hypothesis stating that the insular cortex represents bodily information in a hierarchical manner along its anterior–posterior axis (Craig, 2009). Whereas the posterior insular end processes internal physiology (i.e., interoceptive information such as pain), the middle, and anterior insula assemble such signals into subjective representations of bodily states. Our results are congruent with this theory, as they associate somatic cue processing (Bod-ST movies) with the posterior insula; and the “top-down” conceptualization of these cues during Aff-ToM with middle-anterior insula activations. Future studies could test this idea more precisely, for instance, by implementing brain stimulation protocols and testing their effects on AS versus affective ToM tasks (Sellaro, Nitsche, & Colzato, 2016).

## 4.2 | Traces of somatic cues in subsequent ToM inferences

Contrary to our Hypothesis 2a, we did not find evidence that AS level influenced Aff-ToM accuracy. This stands in contrast to previous studies, which demonstrated that stronger empathic experiences and availability of emotional cues can both facilitate (Regenbogen et al., 2012; Schmitgen et al., 2016) and deteriorate (Kanske et al., 2016) the accuracy of affective ToM under different circumstances. Overall, the movies in our paradigm presented clear emotional narratives and Aff-ToM questions addressed relatively simple aspects of emotion identification (e.g., determining whether a person was sad or not). Accordingly, performance level reached a ceiling effect. Moreover, the availability of visual emotional cues in both soundtrack conditions perhaps provided sufficient information for participants to reach accurate inferences given the simplicity of questions. Thus, future studies could incorporate more emotionally ambiguous stimuli and difficult affective ToM questions in order to elucidate how different levels of AS impact ToM performance.

In accordance with Hypothesis 2b, we found that the Bod-ST movies affected activation in regions related to both ToM (DMPFC) and AS (IFG–anterior insula; Figure 3b) during subsequent Aff-ToM questions. Within the neuroscientific empathy framework, the

intersection of the IFG and the anterior insula is considered vital for affective “mirroring”—that is, representing others' affective states by covertly imitating the expressive gestures that correlate with these affective states (Oliver et al., 2018; Wicker et al., 2003). Notably, it was previously proposed that the IFG may be activated not only via bottom-up but also through top-down processes during empathic experiences (Shamay-Tsoory, 2011). Meta-analyses have respectively associated the IFG–anterior insula to a larger extent with both sharing others' emotions and with reasoning about those emotions (i.e., affective ToM; Lamm et al., 2011; Molenberghs et al., 2016; Timmers et al., 2018), relative to cognitive ToM (Schurz et al., 2021). Likewise, a recent meta-analysis on emotion perception found that the IFG is frequently activated by tasks requiring both implicit and explicit emotion identification (Dracu & Frühholz, 2016). These notions regarding the IFG are in line with our finding that right IFG–anterior insula activity enhanced within Aff-ToM questions after Bod-ST versus Env-ST movies, as well as during the Bod-ST movies. Thus, this region may link bodily resonance of affective states (i.e., AS) to top-down activation of these representations when generating affective ToM inferences. However, this hypothesis needs to be tested more directly, for example by testing an action-simulated related neural pattern that is present during both AS and affective-ToM (Oliver et al., 2018). Moreover, the IFG is also implicated in emotion labeling processes, which are probably activated during both AS and affective ToM tasks (Schurz et al., 2021). Thus, the exact functions that the IFG–anterior insula area (and perhaps, its specific sub-regions) mediate during AS and affective ToM, demands further clarification.

We further found that the effect of Bod-ST movies on brain activity accompanying Aff-ToM questions was most pronounced in the DMPFC. The DMPFC is well recognized as a core hub for mentalizing about others (Lieberman et al., 2019; Saxe & Powell, 2006), and is consistently implicated in ToM (Molenberghs et al., 2016; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). Moreover, the MPFC is recurrently associated with emotional processing (Brunet-Gouet & Decety, 2006; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). Our results revealed that DMPFC activation was enhanced after an AS accentuation, thereby suggesting that somatic cues modulate the retrospective processing of the socio-affective event in this node of the ToM network. The finding that somato-affective cues modulate subsequent activity in the brain network implicated with ToM, and not only in a network associated with AS, may have a theoretical significance. Current accounts of the relations between ToM- and AS-related brain networks point to their conjoint operation in specific social cognition tasks (Schurz et al., 2021). On the other hand, dynamic causal modeling analysis indicated that during intensive emotional moments, the anterior insula from the AS network inhibited the TPJ from the ToM network (Kanske et al., 2016). Further evidence for inhibitory relations comes from studies showing that the ventral salience and the default-mode networks, which overlap with the AS and ToM networks, respectively, show oppositional patterns of activity during tasks that require externally focused attention (Goulden et al., 2014; Wen, Liu, Yao, & Ding, 2013). Our findings suggest that an additive impact of salient

somatic cues on components of the ToM network may not be limited to the event time.

If this is the case, unconscious bodily memory of the original somato-affective context probably mediates this lasting effect. Our findings about the involvement of the VMPFC may be of relevance here. The follow-up functional connectivity analysis showed that stronger connectivity between the DMPFC and the VMPFC (Figure 4a) corresponded with better Aff-ToM performance relative to Cog-ToM performance. Note that the VMPFC cluster found in this PPI analysis overlapped with activation elicited by the ToM > Control contrast, thereby corroborating the association of this region with neural processing of ToM (Figure 4a). The VMPFC has been implicated in affective ToM (Abu-Akel & Shamay-Tsoory, 2011; Shamay-Tsoory & Aharon-Peretz, 2007) and empathic care (Ashar, Andrews-Hanna, Dimidjian, & Wager, 2017). More specifically, the VMPFC plays a major role in assigning value to self-relevant stimuli (D'Argebeau et al., 2012; Delgado et al., 2016), and several meta-analyses show that it is more tuned to information about oneself rather than others (Lieberman et al., 2007; Murray, Schaer, & Debbané, 2012; Qin & Northoff, 2011). The VMPFC was previously associated also with enhanced activity when people pondered about close or similar others (Courtney & Meyer, 2020; Krienen, Tu, & Buckner, 2010; Mitchell, Macrae, & Banaji, 2006), and with computations of values when one needs to make decisions on behalf of others (Nicolle et al., 2012). In terms of the link between empathy to bodily memory, it is worth mentioning VMPFC's key role in Antonio Damasio's *Somatic Marker Theory* (Bechara, Damasio, Tranel, & Damasio, 1997; Damasio, 1994). This influential neurobiological theory postulates that somatic cues affect subsequent emotional conceptualization and decision making through the VMPFC. Damasio argues that this afferent input, which may be experienced as intuition during recollection of previous experiences, attaches value to the alternative options at stake. Thus, VMPFC may represent both personal value and others' feelings, and our results suggest that its recruitment alongside the DMPFC for accurate affective mentalizing might be contingent on presence of somatic cues. It is possible that the ventromedial-dorsomedial connection functions as a "gate" through which prior affective-somatic traces influence subsequent affective ToM. Yet, the involvement of self-value processes in forming accurate affective ToM necessitates further investigation, and our correlational results also require replication in larger samples. Note also that we did not find a correlation between DMPFC activity and Aff-ToM accuracy, nor did we collect ratings of empathic response to the movie characters from participants in the fMRI sample. Thus, it is difficult to determine whether DMPFC activity during Aff-ToM was related to better understanding or greater empathy toward characters after Bod-ST movies; or perhaps reflected greater mentalizing effort.

### 4.3 | LIMITATIONS

A number of additional limitations of this study should be taken into account. First, our task did not include stimuli depicting positive

emotions. Previous studies found that while the affective sharing of others' negative and positive emotions and mentalizing about them engages the IFG-anterior insula and DMPFC, these regions are activated more strongly and selectively by negative events (Lamm, Silani, & Singer, 2015; Morelli, Rameson, & Lieberman, 2014; Perry, Hendlar, & Shamay-Tsoory, 2012). Moreover, empathic processing of positive events seems to involve distinct regions, such as the orbitofrontal cortex (Lamm et al., 2015). Thus, it is yet unclear if the encoding of affective information in the IFG-anterior insula and DMPFC areas is indeed generalizable to positive empathic scenarios. Another major limitation of this study is that since Bod-ST movies were generally more arousing and negative than Env-ST movies (Figure 2a), it is difficult to determine if the effects we observed during questions are attributable to the AS induction per se, or perhaps to momentary induction of negative mood. To account for this, we reanalyzed the brain response to questions while controlling for participants' post-movie momentary emotional state, and found that the movie by question interactions in the DMPFC and right IFG-anterior insula remained significant (see Supporting Information Results). However, note that for this control analysis we relied on the movies' ratings that we collected in the independent behavioral study (Figure 2a). Subjective emotional ratings and physiological measurements could provide a more accurate depiction of participants' emotional state.

## 5 | CONCLUSION

In the current study, we presented a novel naturalistic empathy paradigm that is designated for distinguishing AS and affective versus cognitive ToM processes, as well as for testing their interactions. Our results extend previous findings in several ways. First, we show that the involvement of insula sub-parts in empathy may be more nuanced than what is currently assumed, and suggest that a posterior/anterior segregation along the insula may support AS and affective ToM, respectively. Second, we provide evidence for traces of somatic information in ToM- and AS-related regions when the individual is required to mentalize on the affective aspects of the experience. Lastly, we show that connectivity between the DMPFC and VMPFC—ToM regions that were previously separately identified with cognitive and affective ToM, respectively (Abu-Akel & Shamay-Tsoory, 2011)—may act in concert to support accurate affective mentalization after stronger empathic experiences. These findings underscore regions that may bridge between the immediate affective experiences of others' distress and its high-order conceptualization. These results also contribute to the accumulating literature on the neurobiological bases of AS and ToM interactions. In a broader sense, our findings shed light on how we construct mental representation of emotions based on somatic cues.

### ACKNOWLEDGMENTS

The authors deeply thank Prof. Talma Hendler for her crucial support of the study. This work was supported by the Israel Science



Foundation (grant no. 1333/19) and by the BrainTrain Consortium under the EU PF7 Health Cooperation Program (602186).

## CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

## AUTHOR CONTRIBUTIONS

Ayam Greental, Gadi Gilam, Daniella Perry, Maya Bleich-Cohen, Moran Ovadia, Avihay Cohen, and Gal Raz designed the research; Ofir Shany and Ayam Greental conducted the study; Ofir Shany, Gal Raz, and Ayam Greental analyzed data; and Ayam Greental, Ofir Shany, and Gal Raz wrote the article.

## DATA AVAILABILITY STATEMENT

Code and raw data that support the findings of this study are available from the corresponding authors upon reasonable request.

## ETHICS STATEMENT

This study received ethical approval from the Tel-Aviv Sourasky Medical Center institutional review board (IRB).

## PATIENT CONSENT STATEMENT

All participants provided written informed consent according to the Tel-Aviv Sourasky Medical Center institutional review board (IRB) committee guidelines prior to the experiment.

## ORCID

Ofir Shany  <https://orcid.org/0000-0002-6160-9034>

## REFERENCES

- Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, 49(11), 2971–2984. <https://doi.org/10.1016/j.neuropsychologia.2011.07.012>
- Ashar, Y. K., Andrews-Hanna, J. R., Dimidjian, S., & Wager, T. D. (2017). Empathic care and distress: Predictive brain markers and dissociable brain systems. *Neuron*, 94(6), 1263–1273e4. <https://doi.org/10.1016/j.neuron.2017.05.014>
- Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304), 1293–1295. <https://doi.org/10.1126/science.275.5304.1293>
- Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain*, 133(5), 1515–1525. <https://doi.org/10.1093/brain/awq060>
- Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, 47, 520–532. <https://doi.org/10.1016/j.neubiorev.2014.09.021>
- Blair, R. J. R. (2008). Fine cuts of empathy and the amygdala: Dissociable deficits in psychopathy and autism. *Quarterly Journal of Experimental Psychology*, 61(1), 157–170. <https://doi.org/10.1080/17470210701508855>
- Brunet-Gouet, E., & Decety, J. (2006). Social brain dysfunctions in schizophrenia: A review of neuroimaging studies. *Psychiatry Research: Neuroimaging*, 148(2–3), 75–92. <https://doi.org/10.1016/j.psychres.2006.05.001>
- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49–57. <https://doi.org/10.1016/j.tics.2006.11.004>
- Bzdok, D., Schillbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, 217(4), 783–796. <https://doi.org/10.1007/s00429-012-0380-y>
- Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cerebral Cortex*, 23(3), 739–749. <https://doi.org/10.1093/cercor/bhs065>
- Courtney, A. L., & Meyer, M. L. (2020). Self-other representation in the social brain reflects social connection. *The Journal of Neuroscience*, 40(29), 5616–5627. <https://doi.org/10.1523/JNEUROSCI.2826-19.2020>
- Cox, C. L., Uddin, L. Q., Di Martino, A., Castellanos, F. X., Milham, M. P., & Kelly, C. (2012). The balance between feeling and knowing: Affective and cognitive empathy are reflected in the brain's intrinsic functional dynamics. *Social Cognitive and Affective Neuroscience*, 7(6), 727–737. <https://doi.org/10.1093/scan/nsr051>
- Craig, A. D. (2009). How do you feel—Now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, 10(1), 59–70. <https://doi.org/10.1038/nrn2555>
- D'Argembeau, A., Jedidi, H., Baetou, E., Bahri, M., Phillips, C., & Salmon, E. (2012). Valuing one's self: Medial prefrontal involvement in epistemic and emotive investments in self-views. *Cerebral Cortex*, 22(3), 659–667.
- Damasio, A. R. (1994). *Descartes' error*. New York, NY: Harper Collins.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113.
- Davis, M. H. (1994). *Empathy: A social psychological approach*. Boulder, CO: Westview Press.
- de Waal, F. B. M. (2012). The antiquity of empathy. *Science*, 336(6083), 874–876. <https://doi.org/10.1126/science.1220999>
- de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, 18(8), 498–509. <https://doi.org/10.1038/nrn.2017.72>
- Delgado, M. R., Beer, J. S., Fellows, L. K., Huettel, S. A., Platt, M. L., Quirk, G. J., & Schiller, D. (2016). Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. *Nature Neuroscience*, 19(12), 1545–1552. <https://doi.org/10.1038/nn.4438>
- Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, 24(8), 1742–1752. [https://doi.org/10.1162/jocn\\_a\\_00233](https://doi.org/10.1162/jocn_a_00233)
- Dricu, M., & Frühholz, S. (2016). Perceiving emotional expressions in others: Activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. *Neuroscience & Biobehavioral Reviews*, 71, 810–828. <https://doi.org/10.1016/j.neubiorev.2016.10.020>
- Dziobek, I., Preißler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *NeuroImage*, 57(2), 539–548. <https://doi.org/10.1016/j.neuroimage.2011.05.005>
- Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, 35(3), 903–911. <https://doi.org/10.1016/j.neubiorev.2010.10.009>
- Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, 68, 96–110. <https://doi.org/10.1016/j.neubiorev.2016.05.002>

- Gilam, G., Lin, T., Raz, G., Azrielant, S., Fruchter, E., Arieli, D., & Hendler, T. (2015). Neural substrates underlying the tendency to accept anger-infused ultimatum offers during dynamic social interactions. *NeuroImage*, *120*, 400–411.
- Goulden, N., Khusnulina, A., Davis, N. J., Bracewell, R. M., Bokde, A. L., McNulty, J. P., & Mullins, P. G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: Replication from DCM. *NeuroImage*, *99*, 180–190. <https://doi.org/10.1016/j.neuroimage.2014.05.052>
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640. <https://doi.org/10.1126/science.1089506>
- Hendler, T., Raz, G., Shimrit, S., Jacob, Y., Lin, T., Roseman, L., ... Bleich-Cohen, M. (2018). Social affective context reveals altered network dynamics in schizophrenia patients. *Translational Psychiatry*, *8*(1), 29. <https://doi.org/10.1038/s41398-017-0055-9>
- Kanske, P., Böckler, A., Trautwein, F.-M., Parianen Lesemann, F. H., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective Neuroscience*, *11*(9), 1383–1392. <https://doi.org/10.1093/scan/nsw052>
- Kanske, P., Böckler, A., Trautwein, F.-M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and theory of mind. *NeuroImage*, *122*, 6–19. <https://doi.org/10.1016/j.neuroimage.2015.07.082>
- Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*(3), 165–182. <https://doi.org/10.1007/BF02281954>
- Krienen, F. M., Tu, P.-C., & Buckner, R. L. (2010). Clan mentality: Evidence that the medial prefrontal cortex responds to close others. *Journal of Neuroscience*, *30*(41), 13906–13915. <https://doi.org/10.1523/JNEUROSCI.2180-10.2010>
- Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: Functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure and Function*, *214*(5–6), 519–534. <https://doi.org/10.1007/s00429-010-0255-z>
- Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage*, *54*(3), 2492–2502. <https://doi.org/10.1016/j.neuroimage.2010.10.014>
- Lamm, C., Silani, G., & Singer, T. (2015). Distinct neural networks underlying empathy for pleasant and unpleasant touch. *Cortex*, *70*, 79–89. <https://doi.org/10.1016/j.cortex.2015.01.021>
- Leigh, R., Oishi, K., Hsu, J., Lindquist, M., Gottesman, R. F., Jarso, S., ... Hillis, A. E. (2013). Acute lesions that impair affective empathy. *Brain*, *136*(8), 2539–2549. <https://doi.org/10.1093/brain/awt177>
- Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science*, *18*(5), 421–428.
- Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews*, *99*, 311–328. <https://doi.org/10.1016/j.neubiorev.2018.12.021>
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(03), 121–143.
- Mackes, N. K., Golm, D., O'Daly, O. G., Sarkar, S., Sonuga-Barke, E. J. S., Fairchild, G., & Mehta, M. A. (2018). Tracking emotions in the brain—Revisiting the empathic accuracy task. *NeuroImage*, *178*, 677–686. <https://doi.org/10.1016/j.neuroimage.2018.05.080>
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*(4), 655–663.
- Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291.
- Morelli, S. A., Rameson, L. T., & Lieberman, M. D. (2014). The neural components of empathy: Predicting daily prosocial behavior. *Social Cognitive and Affective Neuroscience*, *9*(1), 39–47. <https://doi.org/10.1093/scan/nss088>
- Murray, R. J., Schaefer, M., & Debbané, M. (2012). Degrees of separation: A quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neuroscience & Biobehavioral Reviews*, *36*(3), 1043–1059. <https://doi.org/10.1016/j.neubiorev.2011.12.013>
- Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron*, *75*(6), 1114–1121. <https://doi.org/10.1016/j.neuron.2012.07.023>
- North, M. S., Todorov, A., & Osherson, D. N. (2010). Inferring the preferences of others from spontaneous, low-emotional facial expressions. *Journal of Experimental Social Psychology*, *46*(6), 1109–1113. <https://doi.org/10.1016/j.jesp.2010.05.021>
- Oliver, L. D., Vieira, J. B., Neufeld, R. W. J., Dziobek, I., & Mitchell, D. G. V. (2018). Greater involvement of action simulation mechanisms in emotional vs cognitive empathy. *Social Cognitive and Affective Neuroscience*, *13*(4), 367–380. <https://doi.org/10.1093/scan/nsy013>
- Pehrs, C., Zaki, J., Schlottermeier, L. H., Jacobs, A. M., Kuchinke, L., & Koelsch, S. (2015). The temporal pole top-down modulates the ventral visual stream during social cognition. *Cerebral Cortex*, *27*(1), 777–792. <https://doi.org/10.1093/cercor/bhv226>
- Perry, D., Hendler, T., & Shamay-Tsoory, S. G. (2012). Can we share the joy of others? Empathic neural responses to distress vs joy. *Social Cognitive and Affective Neuroscience*, *7*(8), 909–916. <https://doi.org/10.1093/scan/nsr073>
- Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, *57*(3), 1221–1233. <https://doi.org/10.1016/j.neuroimage.2011.05.028>
- Raghuathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods*, *1*(2), 178–183. <https://doi.org/10.1037/1082-989X.1.2.178>
- Raz, G., Jacob, Y., Gonen, T., Winetraub, Y., Flash, T., Soreq, E., & Hendler, T. (2014). Cry for her or cry with her: Context-dependent dissociation of two modes of cinematic empathy reflected in network cohesion dynamics. *Social Cognitive and Affective Neuroscience*, *9*(1), 30–38.
- Raz, G., Shpigelman, L., Jacob, Y., Gonen, T., Benjamini, Y., & Hendler, T. (2016). Psychophysiological whole-brain network clustering based on connectivity dynamics analysis in naturalistic conditions: Psychophysiological whole-brain network clustering. *Human Brain Mapping*, *37*(12), 4654–4672. <https://doi.org/10.1002/hbm.23335>
- Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., & Kellermann, T. (2012). Multimodal human communication—targeting facial expressions, speech content and prosody. *NeuroImage*, *60*(4), 2346.
- Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science*, *340*(6129), 216–219.
- Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*(8), 692–699. <https://doi.org/10.1111/j.1467-9280.2006.01768.x>
- Schmitgen, M. M., Walter, H., Drost, S., Rückl, S., & Schnell, K. (2016). Stimulus-dependent amygdala involvement in affective theory of mind generation. *NeuroImage*, *129*, 450–459. <https://doi.org/10.1016/j.neuroimage.2016.01.029>
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging

- studies. *Neuroscience & Biobehavioral Reviews*, 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>
- Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., ... Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, 147(3), 293–327. <https://doi.org/10.1037/bul0000303>
- Sellaro, R., Nitsche, M. A., & Colzato, L. S. (2016). The stimulated social brain: Effects of transcranial direct current stimulation on social cognition: tDCS and social cognition. *Annals of the New York Academy of Sciences*, 1369(1), 218–239. <https://doi.org/10.1111/nyas.13098>
- Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, 17(1), 18–24. <https://doi.org/10.1177/1073858410379268>
- Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, 45(13), 3054–3067. <https://doi.org/10.1016/j.neuropsychologia.2007.05.021>
- Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, 30(6), 855–863.
- Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, 21(3), 489–510. <https://doi.org/10.1162/jocn.2008.21029>
- Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *NeuroImage*, 99, 301–311. <https://doi.org/10.1016/j.neuroimage.2014.05.023>
- Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage*, 59(3), 3050–3059. <https://doi.org/10.1016/j.neuroimage.2011.10.005>
- Timmers, I., Park, A. L., Fischer, M. D., Kronman, C. A., Heathcote, L. C., Hernandez, J. M., & Simons, L. E. (2018). Is empathy for pain unique in its neural correlates? A meta-analysis of neuroimaging studies of empathy. *Frontiers in Behavioral Neuroscience*, 12, 289. <https://doi.org/10.3389/fnbeh.2018.00289>
- Valk, S. L., Bernhardt, B. C., Böckler, A., Trautwein, F.-M., Kanske, P., & Singer, T. (2016). Socio-cognitive phenotypes differentially modulate large-scale structural covariance networks. *Cerebral Cortex*, 27(2), 1358–1368. <https://doi.org/10.1093/cercor/bhv319>
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829–858. <https://doi.org/10.1002/hbm.20547>
- Weisz, E., & Zaki, J. (2018). Motivated empathy: A social neuroscience perspective. *Current Opinion in Psychology*, 24, 67–71.
- Wen, X., Liu, Y., Yao, L., & Ding, M. (2013). Top-down regulation of default mode activity in spatial visual attention. *Journal of Neuroscience*, 33(15), 6444–6453.
- Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula. *Neuron*, 40(3), 655–664. [https://doi.org/10.1016/S0896-6273\(03\)00679-2](https://doi.org/10.1016/S0896-6273(03)00679-2)
- Zaki, J. (2013). Cue integration: A common framework for social cognition and physical perception. *Perspectives on Psychological Science*, 8(3), 296–312. <https://doi.org/10.1177/1745691613475454>
- Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, 15(5), 675–680. <https://doi.org/10.1038/nn.3085>
- Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences*, 106(27), 11382–11387. <https://doi.org/10.1073/pnas.0902666106>

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Shany, O., Greental, A., Gilam, G., Perry, D., Bleich-Cohen, M., Ovadia, M., Cohen, A., & Raz, G. (2021). Somatic engagement alters subsequent neurobehavioral correlates of affective mentalizing. *Human Brain Mapping*, 42(18), 5846–5861. <https://doi.org/10.1002/hbm.25640>