

# Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving

 Anita Tusche,<sup>1,2</sup> Anne Böckler,<sup>1</sup>  Philipp Kanske,<sup>1</sup>  Fynn-Mathis Trautwein,<sup>1</sup> and Tania Singer<sup>1</sup>

<sup>1</sup>Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany, and <sup>2</sup>Emotion and Social Cognition Lab, Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125

Altruistic behavior varies considerably across people and decision contexts. The relevant computational and motivational mechanisms that underlie its heterogeneity, however, are poorly understood. Using a charitable giving task together with multivariate decoding techniques, we identified three distinct psychological mechanisms underlying altruistic decision-making (empathy, perspective taking, and attentional reorienting) and linked them to dissociable neural computations. Neural responses in the anterior insula (AI) (but not temporoparietal junction [TPJ]) encoded trial-wise empathy for beneficiaries, whereas the TPJ (but not AI) predicted the degree of perspective taking. Importantly, the relative influence of both socio-cognitive processes differed across individuals: participants whose donation behavior was heavily influenced by affective empathy exhibited higher predictive accuracies for generosity in AI, whereas those who strongly relied on cognitive perspective taking showed improved predictions of generous donations in TPJ. Furthermore, subject-specific contributions of both processes for donations were reflected in participants' empathy and perspective taking responses in a separate fMRI task (EmpaToM), suggesting that process-specific inputs into altruistic choices may reflect participants' general propensity to either empathize or mentalize. Finally, using independent attention task data, we identified shared neural codes for attentional reorienting and generous donations in the posterior superior temporal sulcus, suggesting that domain-general attention shifts also contribute to generous behavior (but not in TPJ or AI). Overall, our findings demonstrate highly specific roles of AI for affective empathy and TPJ for cognitive perspective taking as precursors of prosocial behavior and suggest that these discrete routes of social cognition differentially drive intraindividual and interindividual differences in altruistic behavior.

**Key words:** fMRI; mentalizing; multivariate pattern analysis (MVPA); prosocial decision-making; social cognition; theory of mind (ToM)

## Significance Statement

Human societies depend on the altruistic behavior of their members, but teasing apart its underlying motivations and neural mechanisms poses a serious challenge. Using multivariate decoding techniques, we delineated three distinct processes for altruistic decision-making (affective empathy, cognitive perspective taking, and domain-general attention shifts), linked them to dissociable neural computations, and identified their relative influence across individuals. Distinguishing process-specific computations both behaviorally and neurally is crucial for developing complete theoretical and neuroscientific accounts of altruistic behavior and more effective means of increasing it. Moreover, information on the relative influence of subprocesses across individuals and its link to people's more general propensity to engage empathy or perspective taking can inform training programs to increase prosociality, considering their "fit" with different individuals.

## Introduction

Altruism, which involves costly other-regarding behavior, is a complex phenomenon resulting from numerous mental pro-

cesses (Fehr and Fischbacher, 2003; Penner et al., 2005; Tankersley et al., 2007; Zaki and Mitchell, 2011; Hutcherson et al., 2015), such as fairness preferences (Fehr and Schmidt, 2006) or received

Received Sept. 9, 2015; revised Jan. 27, 2016; accepted March 7, 2016.

Author contributions: A.T. and T.S. designed research; A.T., A.B., P.K., and F.-M.T. performed research; A.T. analyzed data; A.T. and T.S. wrote the paper.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Anita Tusche, California Institute of Technology, 1200 E. California Blvd, MC 228-77, Pasadena, CA 91125. E-mail: anita.tusche@gmail.com.

DOI:10.1523/JNEUROSCI.3392-15.2016

Copyright © 2016 the authors 0270-6474/16/364719-14\$15.00/0

satisfaction when helping the needy (Harbaugh et al., 2007). With the emergence of social neuroscience (Adolphs, 2010; Singer, 2012), the role of social processes such as empathy and perspective taking for driving differences in prosociality, have moved into the focus of attention. Empathy refers to an affective state that is elicited by and is isomorphic to another person's affective state (e.g., suffering) (Decety and Jackson, 2004; de Vignemont and Singer, 2006). Thus, empathizing refers to affect sharing ("feeling with") of others' states. In contrast, perspective taking (mentalizing or theory of mind) refers to a cognitive process of inferring and reasoning about others' beliefs, thoughts, or intentions (Frith and Frith, 2006) and does not require affective involvement. Empathy and perspective taking have been repeatedly linked to altruistic behavior (Hare et al., 2010; Hein et al., 2010; Mathur et al., 2010; Masten et al., 2011; Telzer et al., 2011; Morishima et al., 2012; Rameson et al., 2012; Waytz et al., 2012; Morelli et al., 2014). However, although both social capacities likely act in concert (Lamm et al., 2011; Kanske et al., 2015b), they are almost always studied in isolation, making it impossible to disentangle their behavioral and neural contributions to prosociality. This poses a challenge to our understanding of altruistic behavior, its determinants, and more effective means of increasing it.

Interestingly, neuroimaging evidence suggests that both processes draw on distinct brain networks: Empathy with others' suffering has been suggested to involve the anterior insula (AI) and mid cingulate cortex (mCC) (Singer and Lamm, 2009; Kurth et al., 2010; Fan et al., 2011), whereas perspective taking recruits the temporoparietal junction (TPJ), posterior superior temporal sulcus (pSTS), temporal poles, and medial prefrontal cortex (mPFC) (Van Overwalle, 2009; Mar, 2011; Bzdok et al., 2012; Schurz et al., 2014). Consistent with the notion of functional segregation, patients with autistic spectrum disorders often exhibit deficits in perspective taking (Frith, 2001; Hoffmann et al., 2015) but not necessarily empathy (Bird et al., 2010), whereas psychopaths demonstrate a lack of empathy but are not impaired in perspective taking (Blair, 2008; Meffert et al., 2013). Likewise, in a paradigm (EmpaToM) that simultaneously assesses and independently manipulates both processes (Kanske et al., 2015b), AI reflected self-reported empathy, whereas ventral TPJ was selectively linked to perspective taking performance. Interestingly, TPJ (but also AI) has also been linked to low-level, stimulus-driven reorienting of attention (Decety and Lamm, 2007; Corbetta et al., 2008; Mitchell, 2008; Krall et al., 2015). Thus, the role of TPJ during altruistic choices might reflect domain-general attentional shifts rather than social computations related to mentalizing (for debate, see Mitchell, 2008; Scholz et al., 2009; Young et al., 2010).

We used an fMRI donation task together with multivariate decoding techniques to disentangle conceptually related, but functionally distinct, mechanisms underlying altruism. We hypothesized that both empathy and perspective taking contribute to trial-by-trial differences in generosity but are encoded in dissociated response patterns in AI (or mCC; mediating empathy) and TPJ (or pSTS, mPFC; mediating perspective taking). Moreover, given robust evidence of individual differences in both processes (Lamm et al., 2011; Kanske et al., 2015b), we examined whether their relative influence on donations reflects differences in participants' general propensity to either empathize or mentalize in complex social settings. This was tested using cross-task decoding together with the novel EmpaToM task that independently measures empathy and perspective taking. Finally, additional behavioral and neural measures were used to distinguish alternative computational accounts of the function of TPJ and AI

during altruistic giving, such as low-level, domain-general attentional processes, or the mediating role of social processes, such as judgments of similarity, closeness, or deservingness, (Hare et al., 2010; Batson, 2011; Telzer et al., 2011).

## Materials and Methods

### Participants

Thirty-three volunteers (15 female, mean age = 26 years, range: 20–34 years) participated in the fMRI donation task, fMRI EmpaToM task (Kanske et al., 2015b), and a behavioral post-test (counterbalanced order of donation task and EmpaToM task across participants). A subset of 23 participants took part in an additional fMRI attention task, which was performed after the main donation task. All participants had normal or corrected-to-normal vision, were free of psychiatric or neurological history, and were German native speakers. Subjects were paid €8 per hour for their participation, plus 20% of the money that was not donated in a randomly selected donation trial. The donation task data for one participant was excluded from further analyses due to insufficient generosity (<1%). The attention data of one participant was excluded from the analysis because of exceedingly high error rates (90%). Moreover, due to technical difficulties with the scanner, EmpaToM data of 2 participants were not available. We obtained written informed consent from all participants, and the local ethics committee approved the study.

### Task design

**Donation task.** Altruistic behavior was measured using a charitable donation task (Böckler et al., 2016) (see Fig. 1A). Before the fMRI session, participants were given an endowment of €50. In each trial, they were first presented with a short description of a charitable organization, including its name, main goal, beneficiaries, and examples of the charity's activities (reading phase; up to 30 s). Participants indicated via a button press when they had finished reading. A continuous rating scale was then presented at the bottom of the screen (range of €0 to €50), and participants had to decide how much to donate to that charity (decision phase; up to 10 s). Participants responded by moving a slider (randomized initial position) using the index and middle finger of the right hand and confirmed their donation choice with a button press of the right ring finger. Intertrial intervals varied from 4 to 8 s, during which a black fixation cross was presented against a white background. Before scanning, participants were informed that one donation trial would be randomly selected at the end of the experiment and that their donation would be given to the charity (full amount donated in that trial). Notably, donations were costly to the subjects because they could keep 20% of the €50 that was not given to the randomly chosen charity. These rules ensured that participants treated each trial as if it was the only one (instead of dividing their endowment among different charities) and made donations consistent with their actual preferences. Partial payout was implemented to provide a moderate incentive for donations (Tusche et al., 2013). The 60 charities used in the task were selected based on pretests using independent samples. The charities were presented in 5 blocks, with the presentation order of charities randomized across participants, within and across donation blocks. Task blocks were divided by breaks of ~30 s in which participants were allowed to rest.

**Behavioral post-test.** Subsequent to scanning, participants completed a behavioral task designed to measure the extent to which participants deployed several social processes that have previously been associated with altruistic behavior (Eisenberg and Miller, 1987; Eckel and Grossman, 1996; Moll et al., 2006; Hare et al., 2010; Hein et al., 2010; Mathur et al., 2010; Batson, 2011; Ma et al., 2011; Waytz et al., 2012). During this task, participants were again presented with the mission statements of each charitable organization and were asked to provide the following ratings for each charity: (1) empathic responses ("how much empathy did you experience for the beneficiaries of the charity?"); (2) perspective taking ("how much did you try to take the perspective of the beneficiaries of the charity?"); (3) similarity ("how similar to you are the beneficiaries?"); (4) closeness ("how likely is it that you or someone close to you would benefit from the organization?"); (5) deservingness ("how much do the beneficiaries deserve to be supported?"); and (6) opposition ("how

much are you opposed to the general cause of the organization”). Importantly, participants were instructed to reconstruct and report their engagement in these processes during the previous donation decisions in the scanner. For each charity, the ratings were collected in randomized order using a continuous rating scale that ranged from 0 (not at all) to 5 (very much). Self-report ratings aimed to investigate the role of empathy and perspective taking in altruistic giving (on the within- and between-subject level). Additional ratings were included to examine the mediating role of social processes, such as judgments of similarity, closeness, or deservingness, on prosocial behavior and allowed investigating the specificity of neural responses in, for example, AI and TPJ.

**EmpaToM task.** An additional fMRI task measured participants' empathy- and perspective taking-related responses independent of the donation task (Kanske et al., 2015b). These data were used to identify subject-specific neural markers of empathy and perspective taking, respectively. We also compared its behavioral data with self-reported empathy and perspective taking in the donation task as measured subsequent to scanning in the behavioral post-test. The EmpaToM task refers to cognitive perspective taking as “Theory of Mind” (ToM).

In each of the 48 trials, participants saw a short video (~15 s) of an actor recounting an autobiographical episode that was either highly emotionally negative (high emotionality) or neutral (low emotionality). The assignment of videos to both conditions was defined a priori based on extensive validations of the EmpaToM paradigm (Kanske et al., 2015b). After each video, participants rated their own affect to measure spontaneous empathic responding (“How do you feel?”; “very negative” to “very positive,” scale from -3 to 3) and their compassion for the person in the video (“How much compassion do you feel?”; “none” to “very much,” scale from 0 to 6) (4 s per rating, fixed order). Participants responded by moving a slider using the index and middle finger of the right hand. After a variable delay of 1–3 s, participants were asked to answer a multiple-choice question concerning the content of the previous video clip (up to 15 s). Questions required either perspective taking (ToM) (e.g., “Anna thinks that . . .”) or factual reasoning (noToM) (e.g., “It is correct that . . .”). Participants responded by pressing one of three buttons, assigned to the three choice options. Trials were separated by variable delays of 1–5 s (plus a 2 s screen that displayed the name of the person in the following video). For a detailed description of task validation and example stories and questions for each experimental condition, see Kanske et al. (2015b).

**Attention task.** Previous meta-analyses point to a substantial overlap of the neural networks involved in domain-general processes, such as reorienting of attention, and those involved in social processes, such as perspective taking and empathy (Decety and Lamm, 2007; Mitchell, 2008; Krall et al., 2015; but see Silani et al., 2013). Based on this evidence, it has been hypothesized that shifts of attention might be responsible for some of the neural activity in areas such as the TPJ during altruistic decision-making (Hare et al., 2010). If this is true and attentional shifts indeed contribute to altruistic giving (and hence account for, at least part of, the predictive information in the TPJ or other regions), neural responses that encode attentional reorienting should be reactivated during generous donations. To investigate this possibility, we asked a subset of participants to also perform a version of the Posner location-cueing task (Posner and Cohen, 1984; Corbetta and Shulman, 2002). The task was structured as follows: In each of the 240 trials, a centrally presented visual cue (white arrow pointing to the left or the right, 200 ms) indicated the spatial location of a following target stimulus that would appear at the left or the right side of the screen (ISI of 200–800 ms between cue and target). Targets were displayed for 200 ms and required participants to respond with a corresponding button press of their right index finger (target arrow upward) or middle finger (target arrow downward). Trials were separated by an average of 2925 ms. Compared with validly cued targets (80%, 192 trials), invalidly cued targets (20%, 48 trials) require reorienting of attention as reflected in increased reaction times (Muller and Rabbitt, 1989).

### Functional image acquisition

Functional imaging was performed on a 3T Verio scanner (Siemens Medical Systems) equipped with a 12-channel head coil. T2\*-weighted functional images were obtained using an EPI sequence (TR = 2 s, TE =

27 ms, flip angle = 90°, 3 × 3 × 3 mm, 1 mm interslice gap, matrix size 70 × 70, 37 slices tilted at ~30° from axial orientation, ipat = 2). For the donation task, a maximum of 932 volumes were acquired. For the attention task, we acquired 470 volumes. For the EmpaToM task, a maximum of 1028 volumes were acquired.

### fMRI data analysis

Functional images were analyzed using the statistical parametric mapping software SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) implemented in MATLAB. Preprocessing consisted of slice-time correction, spatial realignment, and normalization to the MNI brain template. Head movement of all participants was found to be below the criterion of 3 mm/3° throughout the scanning session. For each subject, we estimated several GLMs of the BOLD responses, using a canonical hemodynamic response function, and a 128 s high-pass cutoff filter to eliminate low-frequency drifts in the data.

**GLM 1 (high and low donations).** This GLM was used to obtain a measure of BOLD responses for generous and selfish donation decisions at the task-block level. To this end, for each participant, we defined high and low generosity trials based on participants' average donations (mean split). For each of the five donation blocks, a regressor for high (R1-R5) and low donation decision (R6-R10) was estimated. Each of these regressors lasted from the onset of the donation scale to the button press confirming the donation for the trial (decision phase). Similar block-wise estimates were created for the corresponding reading phases of high (R11-R15) and low (R16-R20) donations (with the modeled duration equal to the time between the presentation of the mission statement and the button press indicating that participants had completed reading). Movement parameters (R21-R26) were included as regressors of no interest. The estimated responses for the regressors of interest (R1-R10) were then used in the multivariate pattern analyses (MVPA) described below. GLM 1 estimated regressors for each of the five donation blocks separately, ensuring that the number of training and test data in the MVPA were perfectly balanced for participants' generous and selfish decisions.

**GLM 2 (trial-wise donations).** This GLM was used to obtain a measure of BOLD responses during the donation task at the trial level. The model included a regressor for each decision phase (R1-R60), a separate regressor for each reading phase (R61-R120), and movement parameters (R121-R126). The duration of the regressors matched the durations modeled in GLM 1. The estimated responses for the donation decisions (R1-R60) were used as inputs for the multivariate support vector regressions (SVRs) described below.

**GLM 3 (EmpaToM task).** This GLM was used to investigate whether the subject-specific influence of empathy and perspective taking in the donation task is related to participants' empathy and perspective taking responses in an independent task. In particular, we tested whether empathy- and perspective taking-related neural signatures obtained in the EmpaToM task can predict the degree to which these processes contribute to donation behavior. To this end, GLM 3 estimated four regressors of interest for (R1) high emotional video phases, (R2) low emotional video phases, (R3) question phases with perspective taking demands (ToM), and (R4) question phases without perspective taking demands (noToM). The two rating phases (R5, R6) and six movement parameters (R7-R12) were included as regressors of no interest. As in previous studies (Kanske et al., 2015b), two contrast images were estimated for each participant: To identify empathy-related brain responses, we contrasted “high emotional” minus “low emotional” video phases (R1-R2). To obtain perspective taking-related brain responses, we contrasted “ToM” – “noToM” question phases (R3-R4). Individual contrast images were entered into two separate random-effects group analyses (simple *t* test in SPM8). Only voxels that were significant at a statistical threshold of  $p < 0.05$  (FWE-corrected, small volume corrected for the respective main effect in a sample of  $n = 178$ ) (Kanske et al., 2015b) were considered relevant for the neural encoding of empathy and perspective taking and were used as input (i.e., features) for the MVPA. All voxels identified in the respective contrast were used together as input for the MVPA described below. The selection of voxels (i.e., features) used for the subsequent MVPA based on statistical criteria, such as *t* tests, is a standard

approach in MVPA (Mitchell et al., 2004; Polyn et al., 2005) to identify “relevant” voxels (i.e., potentially high level of signal and low level of noise) for the neural encoding of a condition/variable of interest and to reduce the dimensionality of the MVPA (for detailed introductions to this topic see, e.g., Haynes and Rees, 2006; Misaki et al., 2010). The selection of voxels based on comparisons of task-related experimental conditions of interest is conceptually well grounded, data-driven, and completely independent from the donation task.

**GLM 4 (attention task).** This GLM was used to investigate whether activity related to the generosity of donations is related to stimulus-driven shifts of attention. GLM 4 estimated the average responses in each target period of the Posner cueing task (R1–R240). Cueing periods and movement parameters were modeled as regressors of no interest. Estimated responses for the 48 invalidly cued target trials and for 48 validly cued target trials (randomly drawn from the total of 192 valid trials) were used as input in the MVPA described below. Only correct trials were considered as input for the MVPA and the number of validly and invalidly cued trials was matched.

### MVPA

**Neural decoding of generous donation decisions.** This decoding analysis was designed to identify brain regions that decode generous versus selfish giving decisions as measured by a mean split of donations at the subject level. To this end, we used GLM 1’s parameter estimates for brain responses during high (R1–R5) and low (R6–R10) donations for each participant. We applied a whole-brain “searchlight” decoding approach that does not depend on a priori assumptions about informative brain regions and ensures unbiased information mapping throughout the whole brain (Kriegeskorte et al., 2006; Haynes et al., 2007). The searchlight decoding was done as follows: For each participant, we defined a sphere (radius = 4 voxels) around a given voxel  $v_i$  of the measured brain volume (Bode and Haynes, 2009; Kahnt et al., 2011a; Tusche et al., 2014; Wisniewski et al., 2015). For each of the  $N$  voxels within this sphere, we extracted parameter estimates and transformed them into an  $N$ -dimensional pattern vector. Pattern vectors were created separately for high and low donations and for each donation block. Pattern vectors of all blocks but one (“training data”) were used to train a linear kernel support vector machine (SVM) classifier (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>) using a fixed cost parameter  $c = 1$ . This provided the basis of the following classification of the pattern vectors of the remaining donation block (“test data”) as representing either high or low donations. The procedure was repeated five times, always using a different donation block as test dataset (leave-one-pair-out cross-validation). The amount of predictive information on generosity was represented by the average percentage of correct classifications across the fivefold cross-validation and was assigned to the central voxel  $v_i$  of the searchlight sphere. This support vector classification (SVC) was successively performed for all spherical clusters created around every voxel of the measured brain volume, resulting in a 3D map of average classification accuracies for each participant. These accuracy maps were then smoothed (8 mm FWHM) and entered into a random-effects group analysis (simple  $t$  test in SPM8) to identify brain regions that encoded generosity across participants. Chance level of the binary classification (high vs low donations) was 50%. In this and later searchlight decoding analyses, brain regions are reported as significant only if they yielded decoding accuracies significantly above chance at a stringent statistical threshold of  $p < 0.05$  (FWE-corrected for multiple comparisons at the voxel level).

**Post hoc permutation tests** were performed for each identified cluster (using all voxels within a cluster) to illustrate how likely cluster-wise decoding accuracies were achieved by chance, compared with data-driven permutation-based null distributions. For each participant, permutation distributions were created by breaking up the mapping of labels (high or low donation) and response patterns (1000-fold). To allow for permutation tests at the group level, we then compared average “real” decoding accuracies (cluster-wise mean across participants) to sampled permutation null distribution (averaged across participants for each of the 1000-fold). Cluster-wise permutation-based null distributions were also created for all other whole-brain searchlight decoding analyses, verifying theoretical chance levels (e.g., of 50% for binary clas-

sifications) without exception. Complementary permutation tests, which further illustrated that cluster-wise decoding accuracies are unlikely achieved under the null hypothesis, are only displayed for the whole-brain decoding of high and low donations.

To compare these decoding results with classic univariate effects, we also estimated a GLM 5 that was identical to GLM 1 (including parameter estimates for “high” and “low” donations) with the exception that spatially smoothed functional data were used (Gaussian kernel of 8 mm FWHM). Individual contrast images were then used to compute random-effects group analyses using paired  $t$  tests as implemented in SPM8. Similar to the multivariate decoding approach, we applied a stringent statistical threshold of  $p < 0.05$  (FWE-corrected).

In addition to binary classifications of high and low donations, we also examined whether trial-by-trial variations in donation size were encoded in multivoxel response patterns, using a support vector regression (SVR) analysis in combination with a searchlight approach (as outlined above). This analysis used trial-wise donations as labels and trial-wise response patterns (GLM 2) as input features of the prediction (for a detailed description of SVR, see below).

**Neural decoding of social processes at the time of donation decisions.** Next, we examined whether predictive neural information on generous donations can be linked to the recruitment of social processes (measured after scanning) that, as reported below, were found to impact donations at the behavioral level. In particular, we tested whether the amount of reported empathy and perspective taking for the beneficiaries of each charity were encoded in brain regions that encoded the generosity of donation decisions. This was done as follows: For each rating obtained in the behavioral post-test, we applied a separate SVR analysis in combination with a searchlight approach as described above. To take full advantage of stimulus-specific ratings (e.g., the degree of perspective taking), we used trial-wise parameter estimates (R1–R60, GLM 2) as input for all linear kernel SVRs ( $\nu$ -SVR; LIBSVM as used for binary classifications; <http://www.csie.ntu.edu.tw/~cjlin/libsvm>) together with participants’ ratings as labels. A slightly reduced cost parameter of  $c = 0.01$  was used (preselected based on previous studies using a similar whole-brain searchlight SVR approach) (Kahnt et al., 2011a, 2014) to account for substantially increased computational costs compared with block-wise, binary SVC. Similar to the decoding of high and low donations, pattern vectors of four of the five donation blocks were used to train the regression model. This provided the basis for the prediction of the ratings of the 12 trials of the remaining donation block (independent test data) based on their neural response patterns. This procedure was repeated several times, always using trials of a different task block as test data (fivefold cross-validation). Information about the mental process of interest (e.g., perspective taking) was defined as the average Fisher’s  $z$ -transformed correlation coefficient between the rating predicted by the SVR model and participants’ actual ratings in these trials (Kahnt et al., 2011a, b, 2014; Gross et al., 2014). Similar to the SVC, this accuracy value was assigned to the central voxel of the searchlight cluster, and the procedure was repeated for every voxel of the measured brain volume. For each rating, resulting decoding maps of all participants were smoothed (8 mm FWHM) and submitted to random-effects group analyses. Statistical tests were restricted to brain regions predictive of the generosity of donations (see Table 2) using the inclusive masking option as implemented in SPM8. Only regions that passed the statistical threshold of  $p < 0.05$  (FWE-corrected at voxel level) are reported.

As before, we compared these results with those of a univariate analysis by estimating a GLM 6 (based on smoothed data, 8 mm FWHM) with the following regressors: reading phases (R1), decision phases (R2), and the decision phase interacted with empathic responses (R3), perspective taking (R4), deservingness (R5), similarity (R6), closeness (R7), and opposition (R8). Movement parameters were included as regressors of no interest (R9–R14). Contrast images were calculated for each of the parametric regressors (R3–R8) and used for separate random-effects group analyses using single  $t$  tests as implemented in SPM8 ( $p < 0.05$ , FWE-corrected). We also repeated this GLM without the serial orthogonalization of R3–R8 as implemented in SPM.

**Neural decoding of the subject-specific influence of empathy and perspective taking in donation decisions.** Behavioral regressions in the donation

task (described below) revealed that the relative influence of empathy and perspective taking in donation decisions varied across individuals. Following up on this evidence, we examined whether these subject-specific contributions of empathy and perspective taking for donation decisions are related to participants' more general empathy and perspective taking responses, independent of altruistic giving. In particular, we tested whether task-elicited neural signatures of empathy and perspective taking can predict the degree to which participants rely on these processes in a separate donation task. Contrary to the previously described decoding analyses that tested for predictive neural information within each participant (across trials or conditions), this analysis focused on interindividual differences and hence used a cross-subject decoding approach. Moreover, instead of focusing on information contained in local spatially distributed response patterns (as identified in searchlight decoding approaches), this analysis used multivoxel representations across brain regions to characterize affective (empathy) and cognitive (perspective taking) mental states in individual subjects.

For each participant, based on functional data in the EmpaToM task, we created two independent contrast images that represented neural signatures for empathy and perspective taking (GLM 3). Two separate group-level analyses were implemented to identify voxels relevant for the encoding of both processes (i.e., feature selection independent of the donation task). For the 119 identified empathy-relevant voxels (see Table 5), we extracted parameter estimates from the subject-specific contrast image for empathy (yielding a 119-dimensional pattern vector for each participant). Subject-specific empathy-patterns were then used to train a linear support vector regression ( $\nu$ -SVR,  $c = 1$ ) together with participants' measures of the relative influence of empathy in the donation task as labels ( $\beta_{\text{empathy}}$  obtained in the behavioral regression of donations). Cross-validation was achieved by a split-half of the sample (i.e.,  $n_{\text{train}} = 15$ ,  $n_{\text{test}} = 15$ ), and repeated 1000 times. Predictive accuracies were defined as the average Fisher's  $z$ -transformed correlation coefficient between the regression weights predicted by the SVR model and actual regression weights of participants in the test sample. To assess the statistical significance of the predictions, accuracies were contrasted against permutation-based null distributions (1000-fold) (paired-sample Wilcoxon signed rank test as implemented in MATLAB, The MathWorks). The decoding procedure was then repeated for perspective taking and the other post-test measures (separately), based on both the empathy-related and perspective taking-related responses in the EmpaToM task. Bonferroni correction was used to adjust  $p$  values for multiple comparisons across the total of 12 independent tests.

**Neural decoding of stimulus-driven reorientation of attention.** To investigate the role of domain-general attentional shifts in altruistic decisions, we performed two different types of decoding analyses for attention data: within-attention task and cross-task decoding.

The within-attention task MVPA identified brain regions that encode stimulus-driven reorienting of attention using a whole-brain searchlight approach similar to the binary classification of generous and selfish donations described above. Parameter estimates of validly and invalidly cued target trials from GLM 4 were used to create trial-specific pattern vectors. Pattern vectors for 48 valid and 48 invalid trials were then assigned to four sets according to their presentation order. Matching the decoding of high and low donations, a linear SVC ( $c = 1$ ) was trained on data of three sets and tested on independent data of the remaining fold. Average correct classification across the fourfold cross-validation was assigned to the central voxel of the searchlight. Participants' decoding maps were smoothed (8 mm FWHM) and tested at the group level against the chance level of 50% ( $p < 0.05$ , FWE-corrected).

The cross-task MVPA tested whether overlaps in areas predicting attention shifts and generous donations are indicative of shared neural code across both tasks. Thus, assuming that attentional shifts indeed contribute to altruistic giving, we hypothesized that neural responses that encode attentional reorienting might be reactivated during generous donations. Comparisons of neural activation patterns across tasks (Tusche et al., 2014; Corradi-Dell'Acqua et al., 2016) allowed us to explicitly test for common neural codes across both tasks. To this end, we performed ROI-based cross-task decoding using clusters in right TPJ and right pSTS as described in Table 2. Decoding parameters and neural response pat-

terns were identical to the within-task decoding of donations and attention shifts, with the exception that we now trained a classifier on all response patterns of the attention task (validly vs invalidly cued targets) and tested it on response patterns of one donation block (low vs high donations). Decoding accuracies obtained in this fivefold out-of-sample prediction (i.e., of independent data of one particular donation block) were then averaged and reflected the amount of shared predictive information in the ROI.

For each ROI, we then tested whether participants' decoding accuracies were statistically significant above the chance level of 50% using permutation tests (1000-fold). Significant results (corrected for the number of ROIs) would point to shared information across both tasks and support the notion that predictive information on generosity in this brain region is, at least partly, due to reinstatement of neural codes recruited for domain-general processes, such as reorienting of attention.

## Results

### Donation behavior

Average donations varied substantially across participants (minimum = €6.6; maximum = €37.68; mean  $\pm$  SD: €22.30  $\pm$  9.09) (Fig. 1B). There was also considerable variation in donations to different charities within participants (range of individuals' SDs: €4.47 to €15.71). This latter variation was exploited in defining "high" and "low" donations at the individual level for the neural decoding of generous and selfish donations. High (€31.38  $\pm$  2.95) and low (€13.95  $\pm$  2.36) donation trials were identified based on participants' average donations (mean split) (Fig. 1C). Decision times were comparable for high (4.67  $\pm$  1.78 s) and low (4.63  $\pm$  1.88 s) donation conditions (paired  $t$  test,  $p = 0.63$ ), ensuring that the neural decoding of generous donations is not confounded by differential durations of decision periods. Likewise, the number of trials for high and low donations was matched (paired  $t$  test,  $p = 0.20$ ). Moreover, there was substantial variation in donations to specific charities (Fig. 1D). This finding makes it unlikely that the neural decoding of generous donations is driven merely by properties of particular charity stimuli.

### Behavioral post-test measures

To identify potential sources of the variation in observed donations, participants rated the extent to which each charity activated six different social mental processes that have previously been implicated in altruistic behavior (Eisenberg and Miller, 1987; Eckel and Grossman, 1996; Moll et al., 2006; Hare et al., 2010; Hein et al., 2010; Mathur et al., 2010; Batson, 2011; Ma et al., 2011; Waytz et al., 2012). Self-report measures were obtained subsequent to scanning (i.e., after the donation task) using continuous rating scales ranging from 0 (not at all) to 5 (very much). On average, participants reported a considerable degree of perceived empathy (mean  $\pm$  SD: 3.22  $\pm$  0.64) and perspective taking (2.84  $\pm$  0.66) for beneficiaries of a charity. Thus, self-report data indicate that participants engaged in empathic responses and perspective taking, to various degrees, when making donation decisions. Self-reports also indicated that beneficiaries were perceived as highly deserving (mean  $\pm$  SD: 3.71  $\pm$  0.53), moderately similar (1.88  $\pm$  0.60), and close (2.11  $\pm$  0.61) to themselves. Closeness referred to the extent to which participants thought that they, or someone close to them, might benefit from the charity (Hare et al., 2010). Participants also reported relatively little opposition to the general goals of the charities (1.56  $\pm$  0.59) (for the correlation matrix of average post-test measures, see Table 1).

### Behavioral prediction of donations

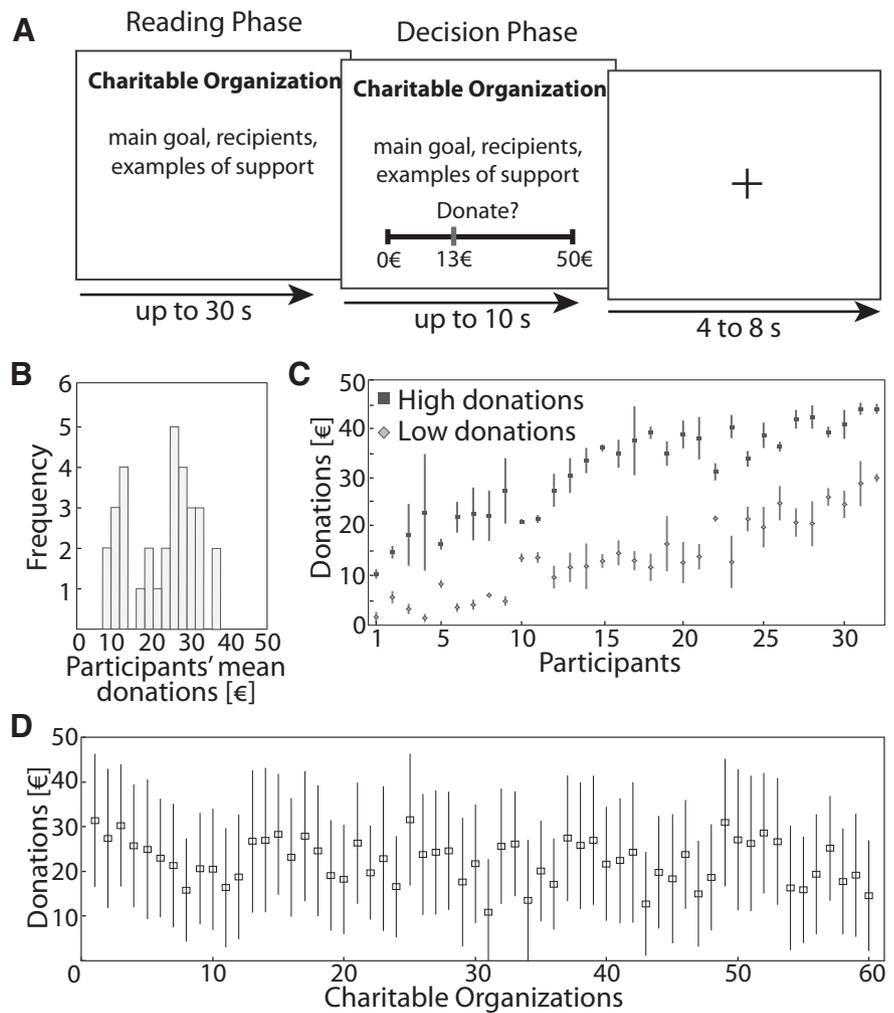
Next, we estimated a linear mixed regression model to investigate the relationship between the six socio-affective/socio-cognitive

ratings and donations. The dependent variable was the donation made in a given trial. The independent variables were the six ratings provided by the same participant for that charity, which were entered as fixed effects, whereas subjects were modeled as random effects. Self-reported empathy ( $\beta = 1.97, p < 0.001$ ), perspective taking ( $\beta = 0.78, p < 0.003$ ), deservingness ( $\beta = 4.52, p < 0.001$ ), and opposition to the general cause of the charity ( $\beta = 1.97$ , rescaled so positive regression parameter represent less opposition,  $p < 0.001$ ) were found to be significant predictors of donations. In contrast, similarity ( $p = 0.14$ ) and closeness ( $p = 0.885$ ) were not significant predictors. Including interaction terms in the linear mixed regression model did not yield any significant interaction effects ( $p$  values  $> 0.05$ ). Overall, these results demonstrate that empathy and perspective taking for beneficiaries of the charities (as well as a number of control variables) are associated with increased levels of generous behavior.

For use in later analyses, we also estimated an individual version of these models. For each participant, we estimated a separate linear regression of donations on the six predictor variables (mean  $\pm$  SD:  $\beta_{\text{perspective taking}} = 0.47 \pm 1.82$ ;  $\beta_{\text{empathy}} = 1.86 \pm 2.47$ ;  $\beta_{\text{opposition}} = 2.59 \pm 2.98$ ;  $\beta_{\text{closeness}} = -0.06 \pm 1.44$ ;  $\beta_{\text{similarity}} = -0.07 \pm 3.76$ ;  $\beta_{\text{deservingness}} = 4.46 \pm 3.01$ ). The estimated coefficients from these regressions were used as individual measures of the extent to which these different processes are associated with donations. Behavioral regression weights were also used to test for neural correlates that reflect the relative level of influence of (particularly) empathy and perspective taking in donation decisions across individuals. Subject-specific contributions of affective empathy were significantly stronger than those for cognitive perspective taking (paired  $t$  test,  $p = 0.03$ ).

**Behavior in a separate empathy and perspective taking task (EmpaToM)**

Participants reported significantly more negative affect after emotionally negative (mean  $\pm$  SD:  $-1.16 \pm 0.88$ , scale from  $-3$  to  $3$ ) than after neutral ( $0.36 \pm 0.76$ ) videos (paired  $t$  test,  $p < 0.001$ ). This finding suggests that participants shared the emotional state of the persons depicted in the videos (i.e., responded empathically). Importantly, participants' average empathic responses in the EmpaToM task (rescaled so that higher values correspond to increased empathic responses) correlated with average empathy ratings for the beneficiaries of the charities in the separate donation task ( $r = 0.34, p = 0.045$ , two-tailed;  $p = 0.55$  for average perspective taking in the donation task). Correlation coefficients were comparable ( $r = 0.35, p = 0.064$ , two-tailed)



**Figure 1.** fMRI donation task and behavior. **A**, In each trial, participants were presented with a mission statement of a real-life charity (reading phase, up to 30 s). Subsequently, a continuous rating scale (€0 to €50) was presented, and participants were asked to decide how much of their endowment of €50 they were willing to donate (decision phase, up to 10 s). Intertrial intervals varied from 4 to 8 s. **B**, Histogram of participants' average donations (range of €6.69 to €37.68; mean  $\pm$  SD: €22.30  $\pm$  9.09), confirming considerable variations in participants' average generosity. **C**, Taking advantage of the variability of generosity across trials, we defined high and low donations for each participant based on their average donation. Graph represents mean  $\pm$  SD for individuals' high and low donations used to decode participants' generous and selfish giving decisions from neural response patterns. **D**, Stimuli were selected based on independent behavioral pretests. Graph represents mean  $\pm$  SD of donations to each charity for the fMRI sample. Substantial variations in donations to each charity ensured that neural decoding of generosity did not merely reflect properties of individual stimuli.

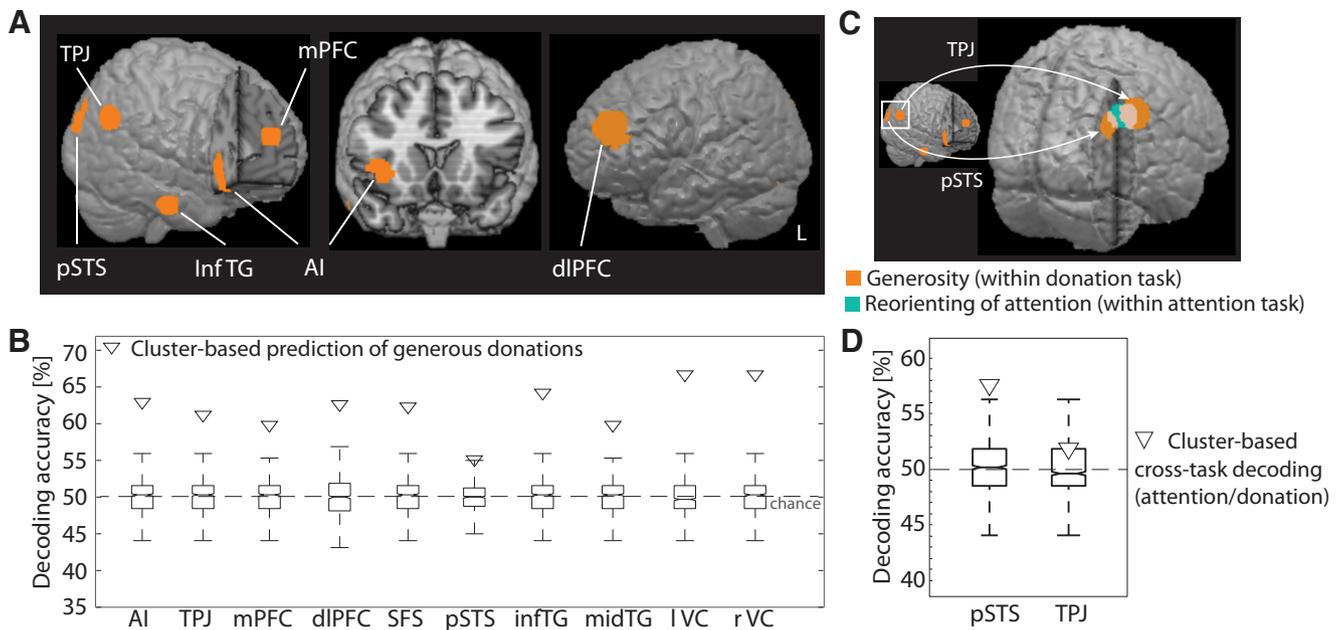
**Table 1. Correlations of average self-reported empathy, deservingness, perspective taking, opposition, similarity, and closeness**

Empathy	Deservingness	Perspective taking	Opposition	Similarity	Closeness
E	0.82*	0.74*	0.73*	0.50 <sup>†</sup>	0.46 <sup>†</sup>
D		0.61*	0.92*	0.35	0.33
PT			0.57*	0.26	0.25
O				0.39 <sup>†</sup>	0.29
S					0.72*
C					

\* $p < 0.05$ , Bonferroni corrected; <sup>†</sup> $p < 0.05$  uncorrected.

when we controlled for social desirability as measured in a questionnaire score (Stöber, 2001), indicating that this effect was not due to self-report biases or demand characteristics.

Participants also reported significantly more compassion after emotionally negative ( $4.00 \pm 1.05$ , scale from 0 to 6) than after neutral videos ( $1.80 \pm 0.93$ ) (paired  $t$  test;  $p < 0.001$ ). Partici-



AI: anterior insula; TPJ: temporoparietal junction; mPFC: medial prefrontal cortex; dlPFC: dorsolateral PFC; SFS: superior frontal sulcus; pSTS: posterior superior temporal sulcus; inf/midTG: inferior/mid temporal gyrus, l/r VC: left/right visual cortex/cerebellum

**Figure 2.** Neural decoding of generous donations and of stimulus-driven reorienting of attention. **A**, A whole-brain searchlight decoding approach was used to identify neural signatures that reliably encoded participants' generous and selfish giving decisions across participants ( $p < 0.05$ , FWE-corrected at voxel level,  $k > 5$  voxels). For details, see Table 2. **B**, *Post hoc* permutation tests further illustrated that decoding accuracies in predictive brain regions (displayed as triangles), based on response patterns of all voxels in a cluster, are unlikely achieved by chance (all  $p$  values  $< 0.007$ ). Boxplots represent null distributions of classification accuracies in the cluster across 1000 permutations. Central marks of boxplots indicate medians, which were found to be 50% (chance level) for all clusters. Edges of boxes indicate the 25th–75th percentiles. Whiskers extend to extreme data points. **C**, Cluster in the right TPJ/pSTS that encoded (nonsocial) stimulus-driven reorienting of attention (displayed in turquoise) in a whole-brain searchlight analysis (within attention task decoding,  $p < 0.05$ , FWE-corrected at voxel level,  $k > 5$  voxels). The predictive cluster overlapped with two clusters in the right TPJ and pSTS that predicted generous and selfish donations in a separate donation task (displayed in orange, within donation task decoding, see **A**). **D**, ROI-based cross-task decoding examined whether neural responses that encode attention shifts are reactivated during generous donation decisions. Activation patterns in the right pSTS coding for stimulus-driven attention shifts contained significant information on generous giving decisions (illustrated by triangle,  $p < 0.05$ , corrected, permutation test). This finding suggests that the domain-general process of reorienting of attention toward (socially relevant) stimuli may account for part of the predictive information in the pSTS in the donation task. No evidence for shared neural code was found for the TPJ cluster ( $p > 0.05$ ). Graph represents cross-task decoding accuracies in both clusters (triangles) and for cluster-based null-distributions (1000 permutations). Boxplots represent medians and 25th–75th percentiles. Whiskers indicate ranges of decoding accuracies.

pants' average self-reported compassion in the EmpaToM task was positively correlated with their average affective empathic responses in both the EmpaToM ( $r = 0.62$ ,  $p < 0.001$ ) and the donation task ( $r = 0.49$ ,  $p = 0.005$ ).

As in previous implementations of this task (Kanske et al., 2015b), EmpaToM questions with perspective taking demands had lower error rates ( $28 \pm 13\%$ ) and slightly faster reaction times ( $8073 \pm 1213$  ms) compared with those without perspective taking requirements (errors:  $42 \pm 13\%$ ; reaction times:  $8286 \pm 1113$  ms; paired  $t$  test for errors:  $p < 0.001$ , for reaction times:  $p = 0.196$ , two-tailed). Difference scores in both measures were  $z$ -scored and averaged to obtain a composite measure of participants' perspective taking performance. Composite scores of participants' perspective taking performance in the EmpaToM were marginally positively correlated with participants' average perspective taking ratings in the donation task ( $r = 0.32$ ,  $p = 0.08$ , two-tailed;  $p = 0.96$  for average empathy for beneficiaries in donation task).

Notably, behavioral measures of empathy and perspective taking are comparable with previous implementations of the EmpaToM task (Kanske et al., 2015b), suggesting that the implementation of the donation task did not affect these independent process measures. The positive correlation of empathy and perspective taking responses in the donation task with independent process-measures suggests that self-reports in the donation task indeed reflect these specific social processes.

### Neural decoding of high and low donations within participants

In the first step, we looked for brain regions that reflect the degree of generosity in the donation task (within participants across donation choices). To this end, for each participant, we used a whole-brain searchlight decoding approach to identify local activation patterns that reliably decode participants' high versus low donations. High and low donation trials might differ in numerous ways, including, for example, in terms of their overall value or saliency. Thus, this initial decoding analysis explicitly allowed for the possibility that predictive neural information is unrelated to socio-cognitive and socio-affective processes. Neural signatures in the right AI and right TPJ reliably encoded participants' generous (vs selfish) donation choices ( $p < 0.05$ , FWE-corrected for multiple comparisons at the voxel level, cluster threshold of 5 voxels). Activation patterns in the bilateral mPFC, dlPFC, right pSTS, right superior frontal sulcus (SFS), and left mid and inferior temporal gyrus also predicted high and low donations (Fig. 2A; Table 2). This analysis was performed at the level of block-wise estimates of conditions of interest (for details, see Materials and Methods), comparable with previous implementations of this MVPA approach in various task domains (Haynes et al., 2007; Tusche et al., 2010, 2013; Bogler et al., 2011; Bode et al., 2013; Wisniewski et al., 2015). The decoding accuracies generated by response patterns in these different clus-

**Table 2. Whole-brain searchlight decoding of generous versus selfish donation decisions<sup>a</sup>**

Brain region	Side	BA	T	k	MNI		
					x	y	z
mPFC	L/R	10/32	5.92	12	0	50	16
dIPFC	L	46	7.20	529	-26	46	22
SFS	R	46/9	5.81	6	26	52	32
AI	R		6.23	108	42	24	0
Mid temporal gyrus	R	21	6.38	13	46	-2	-20
Inferior temporal gyrus	R	20	6.07	14	56	-10	-28
TPJ	R	39/40/22	6.04	20	56	-54	34
pSTS	R	39	6.28	17	38	-66	26
Visual cortex/cerebellum	L	18/19	8.96	2351	-14	-62	0
	R	18/19	7.29	322	14	-72	2

<sup>a</sup>Results are reported at a statistical threshold of  $p < 0.05$ , FWE-corrected for multiple comparisons at voxel level (cluster threshold of 5 voxels); only peak activations of clusters are reported. L, Left hemisphere, R, right hemisphere; BA, Brodmann area; k, cluster size in voxels.

ters are depicted in Figure 2B (triangles). These accuracies are comparable with those found in previous studies using a similar decoding procedure (Hampton and O'Doherty, 2007; Soon et al., 2008; Tusche et al., 2010, 2014; Bode et al., 2013; Wisniewski et al., 2015). *Post hoc* permutation tests illustrate that decoding accuracies in all clusters are unlikely achieved by chance alone (all  $p$  values  $< 0.007$ ; Fig. 2B) (Tusche et al., 2010, 2014; Wisniewski et al., 2015). Moreover, null distributions of classification accuracies based on permutation analyses in all clusters (1000 repetitions) varied  $\sim 50\%$  (Fig. 2B, boxplots), confirming the chance level expected for a binary classification.

Several control analyses confirmed that predictive accuracies for this decoding analysis are unrelated to differential variance in donations for high and low donation trials. For instance, difference scores of variance ( $\delta_{\text{high}}^2 - \delta_{\text{low}}^2$ ) used as a covariate in the random-effects group analysis (simple  $t$  test in SPM) found that decoding accuracies in all clusters were unrelated to differential variance, even at a very liberal statistical threshold of  $p < 0.001$ , uncorrected. Similar results were obtained for using unsigned difference scores ( $|\delta_{\text{high}}^2 - \delta_{\text{low}}^2|$ ) as a covariate ( $p < 0.001$ , uncorrected).

For comparison with previous work, we also estimated a univariate GLM of the BOLD responses to look for areas that respond differently in high versus low donation trials. We did not find any significant regions exhibiting larger responses during high donations at a statistical threshold of  $p < 0.05$ , FWE-corrected. For the reverse contrast (low  $>$  high donations), only one cluster in the early visual cortex was significantly more activated ([MNI -16, -62, -2],  $t = 6.13$ ). These findings suggest that information on the generosity of donations seems to be encoded in spatially distributed response patterns.

Together, these findings provide converging evidence that response patterns in several brain regions, including the right TPJ and right AI, contained reliable information about the level of generosity in donation choices. Further support for this notion was provided by the results of a supplemental support vector regression analysis that tested for response patterns that encode trial-by-trial variations in donation size (for a full list of significant clusters, see Table 3) ( $p < 0.05$ , FWE-corrected at the whole-brain voxel level,  $k = 5$  voxels).

### Neural decoding of high and low donations varies systematically across participants

Next, having established that several brain regions contain information about the generosity of donations, we investigated

whether their neural codes might be related to distinct mental processes of interest, such as empathy or perspective taking. Behavioral regression estimates of all six post-test measures (see Behavioral prediction of donations) were used as covariates in a random-effects group analysis on accuracy maps (simple  $t$  test in SPM, small volume [SV] correction for predictive clusters as listed in Table 2). We found that increased contributions of empathic responses as input for donation decisions were reflected in higher predictive accuracies for generosity obtained in the AI (peak at [MNI 46, 22, -6],  $t = 3.29$ ) (height threshold of  $p < 0.001$ , FWE SV-corrected). Increased contributions of perspective taking, on the other hand, covaried with higher decoding accuracies in the right TPJ (peak at [MNI 64, -50, 32],  $t = 4.09$ ) (height threshold of  $p < 0.001$ , FWE SV-corrected). Decoding accuracies in all other cluster showed no significant interaction with interindividual differences in empathy or perspective taking, even at this more liberal statistical threshold. Notably, these results were found when controlling for the impact of all other post-test measures (using serial orthogonalization as implemented in SPM in different GLMs that only differed in the order of covariates). Individual differences in other post-test measures were not found to systematically covary with decoding accuracies at this statistical threshold. Overall, the results suggest that empathy and perspective taking might be the two psychological processes that most closely resemble the computations performed in the AI and TPJ during generous donations.

### Neural decoding of trial-by-trial variations in empathy and perspective taking for beneficiaries during donations

In this set of analyses, we further investigated the specific link between empathy and neural activation in the AI, and perspective taking and neural responses in the TPJ. In particular, we tested whether brain responses obtained during donation decisions reflected participants' self-reported levels of empathy and perspective taking (as measured after scanning) on a trial-by-trial basis.

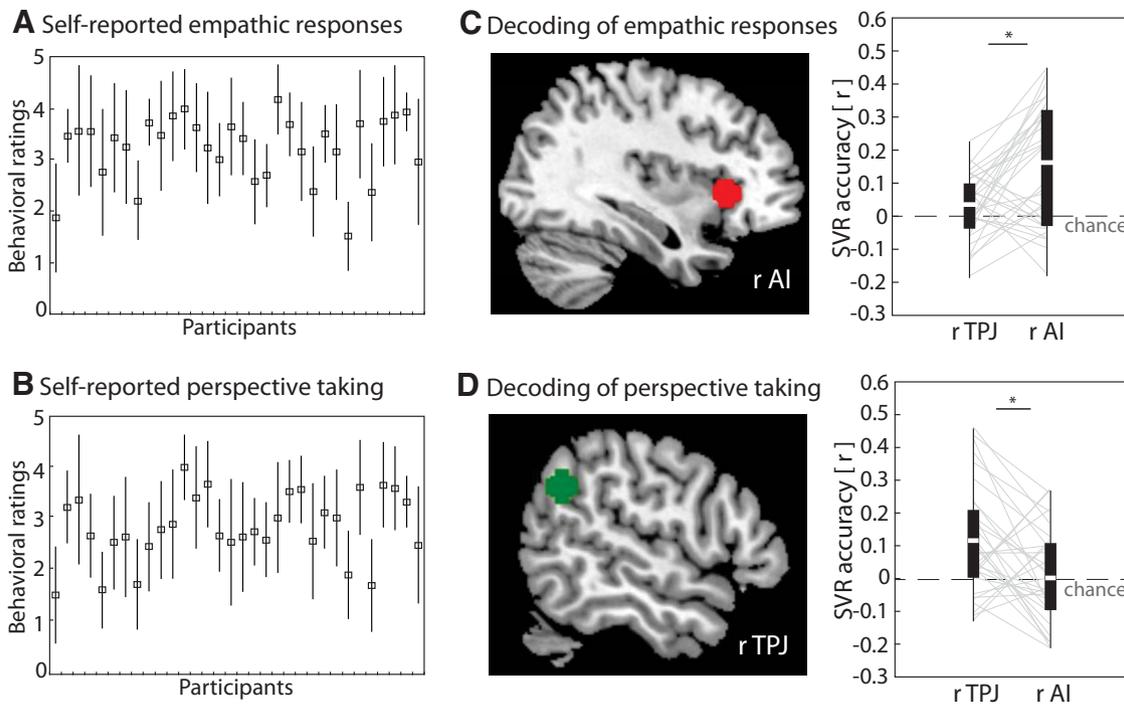
First, we used multivariate regressions (SVRs) together with a whole-brain searchlight approach to identify brain areas that encode the levels of participants' empathy across donation trials (Fig. 3A). Consistent with the results above, we found that trial-by-trial variations in empathic responses for the beneficiaries of different charities were encoded in the right AI (height threshold of  $p < 0.05$ , FWE-corrected at voxel level) (average  $r = 0.14$ ; for illustration of cluster identified in this whole-brain analyses, see Fig. 3C). Response patterns in the mPFC, dIPFC, SFS, and bilateral visual cortex also predicted trial-wise variations in empathy (for details, see Table 4). Notably, while the right AI was found to contain information on empathic responses during donation decisions, predictions based on response patterns in the right TPJ varied around chance (Fig. 3C). A *post hoc* paired  $t$  test of predictive accuracies in both clusters (average across voxels) confirmed that predictive information on participants' empathic responses in the right AI was significantly higher than those found in the right TPJ ( $p = 0.002$ ). *Post hoc* cluster-based permutation tests (using all voxels within a cluster to define response-patterns, 1000-fold) further illustrated that predictive accuracies in all clusters were unlikely to arise by chance ( $p$  values  $\leq 0.001$ ).

Second, we performed a similar whole-brain SVR analysis to look for clusters that encode participants' levels of perspective taking across trials (Fig. 3B). We found that spatially distributed response patterns in the right TPJ encoded the degree of perspective taking during donation decisions ([MNI 58, -54, 32],  $t = 4.23$ , average  $r = 0.11$ ,  $p < 0.05$ ; Fig. 3D). No other brain region (including the AI) was predictive at a stringent statistical thresh-

**Table 3. Whole-brain searchlight regression (SVR) of donations<sup>a</sup>**

Brain region	Side	BA	T	k	MNI		
					x	y	z
mPFC/dIPFC/SFS	R/L	9/10/32/46	9.63	8595	20	36	32
Ventromedial PFC	L	11	6.23	23	-6	36	-8
Inferior frontal gyrus/AI	R	47	6.16	33	38	40	-2
Inferior frontal gyrus/AI	R	44	5.90	23	50	10	12
Inferior/mid temporal gyrus	R	20/21	6.98	77	54	-38	-12
TPJ	R	39/40/22	5.78	28	60	-56	30
Precentral gyrus	L	6	5.88	14	-32	0	46
Superior frontal gyrus/SMA	R	6/8	6.19	104	38	6	66
Superior frontal gyrus/SMA	R	6	5.71	11	24	0	68
Visual cortex/cerebellum (VC)	R	19	6.13	338	22	-86	28
VC/precuneus/superior parietal cortex	R/L	7/18/19/39/40	9.41	7107	6	-70	52

<sup>a</sup>Results are reported at a statistical threshold of  $p < 0.05$ , FWE-corrected for multiple comparisons at voxel level (cluster threshold of 5 voxels); only peak coordinates of clusters are reported. L, Left hemisphere; R, right hemisphere; BA, Brodmann area; k, cluster size in voxels; SMA, supplemental motor area.



**Figure 3.** Neural decoding of empathy and perspective taking for beneficiaries of charities. Left, Data are mean  $\pm$  SD for participants' (A) empathy and (B) degree of perspective taking across donation trials as reported after scanning, indicating participants' engagement in both processes and comparable degrees of variance. C, During donation decisions, empathy for beneficiaries was encoded in response patterns in the right AI (red), but not in the right TPJ ( $p < 0.05$ , FWE-corrected). Pairwise comparisons confirmed that predictive information on empathic responses in AI was significantly higher than those in TPJ ( $p = 0.002$ ) that varied around chance. D, The degree to which participants represented the perspective of the charities' beneficiaries was encoded in the right TPJ (green), but not the right AI ( $p < 0.05$ , FWE-corrected). Information on perspective taking in the TPJ was significantly higher than those decoded from the AI that varied around chance ( $p = 0.008$ ). Graphs represent median predictive accuracies (white bar) achieved in the right AI and TPJ. Boxes represent the 25th–75th percentiles. Whiskers indicate range of multivariate regression accuracies. \* $p < 0.01$ .

**Table 4. Whole-brain searchlight regression (SVR) of empathic responses<sup>a</sup>**

Brain region	Side	BA	T	k	MNI		
					x	y	z
mPFC	L/R	10/32	4.37	6	2	50	18
dIPFC	L	46	4.52	85	-32	58	26
SFS	R	46/9	5.40	6	28	52	34
AI	R		4.82	14	44	22	2
Visual cortex/cerebellum	L	18/19	5.20	194	-14	-62	0
Visual cortex	R	19	7.05	52	28	-80	36
Visual cortex	L/R	18	3.68	15	4	-78	28

<sup>a</sup>Results are reported at a statistical threshold of  $p < 0.05$ , FWE-corrected for multiple comparisons at voxel level (cluster threshold of 5 voxels); only peak coordinates of clusters are reported. L, Left hemisphere; R, right hemisphere; BA, Brodmann area; k, cluster size in voxels.

old of  $p < 0.05$ , FWE-corrected. *Post hoc* permutation tests further illustrated that regression accuracies in the TPJ were unlikely the result of chance ( $p < 0.001$ ). A *post hoc* paired *t* test of decoding accuracies (average across voxels in cluster) confirmed that predictive information on the level of perspective taking encoded in the right TPJ was significantly higher than that found in the right AI ( $p = 0.008$ ) (Fig. 3D).

Importantly, no other brain region was found to encode trial-wise variations in empathy or perspective taking at the whole-brain level, even when using a more liberal statistical threshold (height threshold of  $p < 0.001$ , FWE-corrected at cluster level at  $p < 0.05$ ). This finding, together with the close match of predictive clusters found in multivariate regressions and binary classification of high versus low donations, indicates that the decoding

of high and low donations provides a reasonable data-driven approach to identify regions involved in socio-cognitive process during donation decisions.

Third, for the sake of completeness, we ran separate SVRs for each of the other four ratings. Response patterns in the mPFC ([MNI 0, 48, 16],  $t = 4.91$ ,  $r = 0.10$ ), SFS ([MNI 28, 54, 32],  $t = 4.12$ ,  $r = 0.08$ ), dlPFC ([MNI -30, 46, 22],  $t = 4.08$ ,  $r = 0.08$ ), and visual cortex ([MNI 2, -64, 8],  $t = 5.70$ ,  $r = 0.13$ ; [MNI 16, -64, 12],  $t = 5.62$ ,  $r = 0.14$ ) were found to encode participants' judgments of deservingness ( $p < 0.05$ , FWE-corrected). Perceived similarity, closeness, and opposition to the charities general causes could not be decoded at this stringent statistical threshold, suggesting that these variables are not reliably encoded in localized response patterns during donation decisions. This matches behavioral results, because similarity and closeness were not found to be significant predictors of donations. Notably, response patterns in the TPJ did not contain information about any other rating. Likewise, searchlight SVRs for all other self-report measures revealed that brain responses in the AI were only predictive of affective empathic responses. These findings suggest a highly specific role of the right TPJ for the neural encoding of perspective taking and of the right AI in encoding empathic responses during charitable giving.

For completeness, we also estimated related univariate analysis looking for brain areas in which the BOLD responses at the time of the donation decisions were parametrically modulated by empathy and perspective taking. No such areas were found at a stringent statistical threshold of  $p < 0.05$  (FWE-corrected).

Together, these findings suggest a double dissociation of information encoded in neural activity, with AI reflecting the degree of empathy elicited by the different charities, and the TPJ reflecting the degree of perspective taking associated with each decision. The findings also suggest that these computations might be represented in spatially distributed response patterns that encode trial-by-trial variations in the variable of interest in a continuous fashion. In contrast to conventional mass-univariate analysis of BOLD-response in single voxels, MVPA takes advantage of information contained in differential response biases across multiple voxels (which in turn are driven by population of neurons whose activity increases or decreases as a function of a variable/condition of interest) that are spatially distributed. Interestingly, distributed activation patterns of fMRI data have even been suggested to reflect the underlying neural population code (Kamitani and Tong, 2005).

### Neural decoding of the subject-specific influence of empathy and perspective taking in donation decisions

The relative influence of empathy and perspective taking in altruistic choices varied across individuals, as indicated by subject-specific parameter estimates in the behavioral regression of donations. Here, we examined whether multivariate neural representations of empathic responses and perspective taking in individual subjects, obtained in a separate task that did not require altruistic choices, can predict the degree to which participants rely on these two processes in the donation task. Neural signatures of empathy and perspective taking were created for each participant based on the respective contrast in another task (EmpaToM) that did not require altruistic choices (Table 5). This ensured that the feature selection (i.e., selection of process-relevant voxels) for these decoding analyses was completely independent of the donation task.

We found that participants' perspective taking-related neural signatures in the EmpaToM task (recruiting TPJ, mPFC, STS,

**Table 5. Independent empathy and perspective taking responses (EmpaToM task) used as input for cross-subject decoding analyses<sup>a</sup>**

Brain region	Side	BA	<i>T</i>	<i>k</i>	MNI		
					<i>x</i>	<i>y</i>	<i>z</i>
<b>Perspective taking (ToM &gt; noToM)</b>							
mPFC	L/R	10/9	6.99	236	-9	56	31
TPJ	L	22/39	5.73	177	-48	-52	28
STS	L	21/20	4.82	67	-51	-25	-5
STS	R	21/20	5.84	137	45	-28	-2
Temporal pole	R	20/21/38	7.19	75	45	14	-35
Temporal pole	L	20/21/38	4.86	19	-48	5	-29
<b>Empathy (high &gt; low emotional)</b>							
AI	L		4.62	18	-30	23	-17
Inferior frontal gyrus	R	47	5.00	56	48	47	-11
Supramarginal gyrus	L	40	5.27	45	-54	-55	40

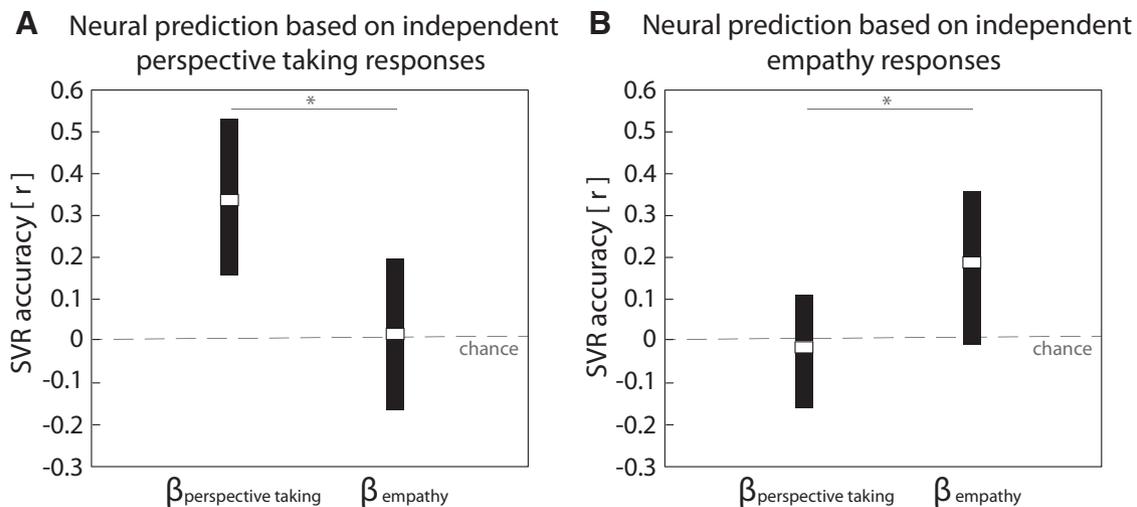
<sup>a</sup>Results are reported at a statistical threshold of  $p < 0.05$ , FWE-corrected; only peak activations of clusters are reported. L, Left hemisphere; R, right hemisphere; BA, Brodmann area; *k*, cluster size in voxels.

and temporal poles; Table 5) predicted the relative weight of perspective taking as input into independent donation decisions ( $\beta_{\text{perspective taking}}$ ) (average Fisher *z*-transformed  $r = 0.34$ ,  $p < 0.001$ , Bonferroni corrected). Consistent with our predictions, neural decoding of the relative influence of empathy in the donation task varied around chance ( $\beta_{\text{empathy}}$ ) ( $r = -0.001$ ,  $p > 0.05$ , corrected) (Fig. 4A). An additional paired-sample Wilcoxon signed rank test verified that predictions of subject-specific contributions of perspective taking were also significantly higher than those for empathy ( $p < 0.001$ ). In contrast, participants' empathy-related activation patterns in the EmpaToM task (engaging AI, inferior frontal gyrus, and supramarginal gyrus; Table 5) reliably predicted the degree to which individuals relied on empathy in the separate donation task ( $\beta_{\text{empathy}}$ ) ( $r = 0.16$ ,  $p < 0.001$ , corrected) (Fig. 4B), but not of perspective taking ( $\beta_{\text{perspective taking}}$ ) ( $r = -0.04$ ,  $p > 0.05$ , corrected). A complementary paired-sample Wilcoxon signed rank test also confirmed that predictive accuracies for empathy were significantly higher than those for perspective taking ( $p < 0.001$ ). For the sake of completeness, we repeated these decoding analyses for all control ratings (deservingness, closeness, opposition, similarity), but comparisons with permutation-based null-distributions did not yield any other significant predictions above those achieved by chance only ( $p$  values  $> 0.05$ , corrected).

Overall, the results suggest that the relative influence of empathy and perspective taking for altruistic choices might reflect participants' more general propensity to recruit these processes when engaging in complex social cognition. Moreover, the high specificity of these predictions provides further validation and evidence that the respective behavioral regression weights for empathic responses and perspective taking in the donation task indeed reflect information on these specific socio-cognitive capacities.

### Neural decoding of stimulus-driven reorienting of attention

To explicitly test whether predictive information in TPJ (or AI) for generous donations is related to domain-general processes, such as reorienting of attention (Corbetta et al., 2008), we asked a subset of participants to perform an additional fMRI Posner location-cueing task (Posner and Cohen, 1984). Participants had to respond to validly and invalidly cued target stimuli (arrows), with the latter requiring reorienting of attention. As expected, participants' response times in invalid (mean  $\pm$  SD: 754  $\pm$  101 ms) trials were found to be significantly longer (paired *t* test,  $p < 0.001$ ) than in valid trials (644  $\pm$  84 ms), reflecting attention



**Figure 4.** Neural decoding of the relative influence of empathy and perspective taking in donation decisions. **A**, Neural signatures for perspective taking obtained in an independent fMRI task (EmpaToM) predicted the relative weight of perspective taking as input into donation decisions ( $\beta_{\text{perspective taking}}$ ) (mean Fisher z-transformed  $r = 0.34$ ,  $p < 0.001$ , Bonferroni corrected), but not of empathy ( $r = -0.001$ ,  $p > 0.05$ , corrected). Graphs represent median predictive accuracies (white bar) achieved in these cross-subject decoding analyses (SVR). Boxes represent the 25th–75th percentiles. **B**, Independent empathy-related neural signatures (EmpaToM) reliably predicted the degree to which participants relied on affective empathy during donation decisions ( $\beta_{\text{empathy}}$ ) ( $r = 0.16$ ,  $p < 0.001$ , corrected), but not contributions of cognitive perspective taking ( $\beta_{\text{perspective taking}}$ ) that varied around chance ( $r = -0.04$ ,  $p > 0.05$ , corrected). \* $p < 0.001$ .

**Table 6. Whole-brain searchlight decoding of attentional reorienting (Posner task)<sup>a</sup>**

Brain region	Side	BA	T	k	MNI		
					x	y	z
Mid frontal gyrus	L	44	7.20	20	-44	12	34
Mid frontal gyrus/precentral gyrus	R	44/6	6.97	8	40	4	32
Superior/mid temporal gyrus	R	22/21	7.31	100	62	-44	8
TPJ/pSTS	R	39	6.75	5	52	-62	32
Posterior cingulate cortex	L/R	23	6.99	19	2	-28	34
Posterior cingulate cortex/paracentral lobe	R	23/4	6.83	16	10	-32	50
Occipital cortex	R	18/19	8.50	202	20	-88	20
	L	19	7.87	87	-28	-62	28
	L	19	6.96	9	-30	-80	4
	L	18	6.68	5	-16	-76	28

<sup>a</sup>Results are reported at a statistical threshold of  $p < 0.05$ , FWE-corrected for multiple comparisons at voxel level (cluster threshold of 5 voxels); only peak coordinates of clusters are reported. L, Left hemisphere; R, right hemisphere; BA, Brodmann area; k, cluster size in voxels.

shifts. The fMRI data obtained during the attention task was analyzed in two steps.

First, we identified brain regions that decode stimulus-driven reorienting of attention within the attention task (i.e., establishing that there is local information on attention shifts). We found that the attention shifts could be decoded from neural responses in the bilateral mid frontal gyrus and posterior cingulate cortex (and adjacent mid cingulate cortex). Moreover, response patterns in the right TPJ/pSTS and right superior/mid temporal gyrus were found to decode attention shifts ( $p < 0.05$ , FWE-corrected at voxel level,  $k > 5$  voxels; Table 6). Significant results at this stringent statistical threshold demonstrate that the sample size is sufficient to detect neural responses related to stimulus-driven reorienting of attention. Moreover, the results are consistent with previous findings on the neural network recruited for the Posner cueing paradigm (Corbetta et al., 2008). Interestingly, as shown in Figure 2C, brain responses in the right TPJ and pSTS were found to decode nonsocial attention shifts as well as generous donations (as identified in an independent within-donation task decoding; Table 2). This spatial overlap of predictive clusters demonstrates that similar brain regions, namely, the right TPJ

and adjacent pSTS, are recruited both during generous giving decisions and stimulus-driven reorienting of attention.

Second, clusters in right TPJ and pSTS (for details, see Table 2) were defined as ROIs, and we implemented cross-task classifications designed to explicitly test for shared neural codes for reorienting of attention and generous donations (Tusche et al., 2014). In particular, we tested whether predictive information on donations is due to reinstated neural codes relevant for domain-general reorienting of attention. Multivariate response patterns in the cross-task classifications were based on all voxels in an ROI (as defined by a predictive cluster identified in the independent within-task searchlight decoding). Interestingly, we found evidence for shared neural codes in the right pSTS across both tasks ( $p < 0.05$ , corrected, permutation tests, 1000-fold; Fig. 2D). This finding suggests that predictive information on generous donations in this brain area might indeed be related to domain-general attention shifts toward (socially relevant) stimuli. However, no evidence for shared neural codes was found in the TPJ cluster ( $p > 0.44$ , uncorrected, permutation tests, 1000-fold; Fig. 2D), suggesting differential neural processes coding for reorienting of attention and generosity that might co-occur in this functionally heterogeneous brain region. Complimentary analyses for all other clusters confirmed that no other brain cluster (Table 2), including the AI, contained shared neural code in the cross-task decoding (all  $p$  values  $> 0.13$ , uncorrected).

## Discussion

The computations and motivations that underlie the heterogeneity in human altruism are poorly understood. Neuroscience has started to shed light on the neural underpinnings of altruistic behavior, such as charitable giving (Moll et al., 2006; Harbaugh et al., 2007), revealing a prominent role of areas such as AI, TPJ, pSTS, and mPFC for differences in prosociality across people and decision contexts (Tankersley et al., 2007; Hare et al., 2010; Hein et al., 2010; Mathur et al., 2010; Masten et al., 2011; Morishima et al., 2012; Rameson et al., 2012; Waytz et al., 2012; Genevsky et al., 2013; Morelli et al., 2014; Zanon et al., 2014; FeldmanHall et al., 2015). While recruitment of these areas during altruistic behavior has been suggested to reflect socio-affective and socio-cognitive

processes, such as empathy or perspective taking, or domain-general attentional shifts, previous research has not distinguished specific contributions of these interrelated processes. To identify process-specific inputs into altruistic giving both behaviorally and neurally, we implemented three fMRI tasks (charitable giving paradigm; an empathy and perspective taking paradigm (Empa-ToM) (Kanske et al., 2015b); nonsocial attentional reorienting paradigm (Posner and Cohen, 1984)) as well as a battery of post-test ratings that capture social processes previously linked to prosocial behavior (Eisenberg and Miller, 1987; Hare et al., 2010; Batson, 2011). Using multivariate decoding techniques, we delineated three distinct psychological processes for altruistic decision-making (affective empathy, cognitive perspective taking, and domain-general attention shifts), linked them to dissociable neural computations, and identified their relative influence across individuals.

Consistent with previous evidence (e.g., Harbaugh et al., 2007; Hare et al., 2010; Hein et al., 2010), we demonstrate that neural responses in AI and TPJ (but also pSTS, mPFC, and dlPFC) encoded participants' generous and selfish giving decisions, but represent different computations: neural responses in AI (but not TPJ) encoded trial-wise experienced affective empathy for beneficiaries during donations, whereas activation in TPJ (but not AI) predicted the degree of cognitive perspective taking in the decision process. This double dissociation suggests distinct functional roles of both social capacities for driving altruistic choices. Individual differences in the relative influence of both processes further supports this functional segregation: participants whose donations were heavily influenced by affective empathic responses exhibited higher predictive accuracies for generosity in the AI. In contrast, an increased influence of cognitive perspective taking on donations correlated with better predictions of generous giving in the TPJ. Interestingly, independently measured empathic brain responses (EmpaToM) (Kanske et al., 2015b) predicted subject-specific contributions of empathy in the donation task (but not of perspective taking). Likewise, independent neural signatures for perspective taking (EmpaToM) decoded subject-specific inputs of perspective taking for altruistic choices (but not of empathy). This finding strongly suggests that participants' utilization of empathy and perspective taking during altruistic choices might reflect participants' more general propensity to engage in affective resonance or to take a more cognitive stance for understanding others in complex social interactions, extending previous links of prosociality and reflexive engagement of perspective taking (Waytz et al., 2012). Although effect sizes were generally small, they were achieved using conservative cross-validation approaches and stringent statistical thresholding, and were consistent with previous applications of unbiased searchlight decoding approaches (Haynes et al., 2007; Kahnt et al., 2011a, b, 2014; Tusche et al., 2014; Wisniewski et al., 2015).

The identified double dissociation of both socio-cognitive processes encoded in AI and TPJ is consistent with previous theoretical accounts (Singer, 2006, 2012; Kanske et al., 2015a) and meta-analytical evidence that empathy for others' suffering relies on a core network that includes the AI and mid cingulate cortex (Singer and Lamm, 2009; Kurth et al., 2010; Fan et al., 2011; Lamm et al., 2011), whereas perspective taking reliably recruits the TPJ, pSTS, mPFC, and temporal poles (Saxe et al., 2004; Amodio and Frith, 2006; Van Overwalle, 2009; Mar, 2011; Bzdok et al., 2012; Schurz et al., 2014). Evidence from clinical populations with marked social deficits also suggests that both capacities might draw on different mechanisms. Patients with autistic spectrum disorders frequently display impaired perspective taking

(Baron-Cohen et al., 1985; Hoffmann et al., 2015) but not necessarily impaired empathy (Silani et al., 2008; Bird et al., 2010), whereas psychopaths often lack empathy but are not impaired in perspective taking (Blair, 2008; Meffert et al., 2013). Our findings demonstrate, for the first time, that this functional segregation of empathy and perspective taking extends to prosocial decision-making settings and that both socio-cognitive routes contribute to variance in altruistic choices, within and across individuals. Notably, we also show that the neural underpinnings of empathy and perspective taking during donations are independent from other relevant variables, such as closeness, similarity, or deservingness, which have previously been linked to increased generosity (Hare et al., 2010; Batson, 2011; Telzer et al., 2011) and might mediate increased engagement in perspective taking and empathy. These findings demonstrate a high degree of functional specificity for AI (mediating empathy) and TPJ (mediating perspective taking, but not deservingness) (Hare et al., 2010) during altruistic behavior. Furthermore, evidence that flexible recruitment of empathy and perspective taking accounts for trial-by-trial variations of altruistic behavior suggests that the economic notion of mostly stable and context-insensitive social preferences (Fehr and Schmidt, 2006; Fehr and Camerer, 2007) should be extended to account for characteristics of the choice and charity. Thus, in addition to more trait-like altruistic or egoistic tendencies, or being generally more emotionally or rationally motivated when making altruistic choices, our results suggest that recruited mechanisms and their input into choices will vary depending on the context. Flexible recruitment of empathy might also account for people's tendency to preferentially give to identified versus anonymous victims (Genevsky et al., 2013).

In addition to affective and cognitive social processes, we also examined the role of stimulus-driven attention shifts in altruistic giving. Previous evidence points to a substantial overlap of neural networks involved in domain-general reorienting of attention with those involved in perspective taking and empathy (Decety and Lamm, 2007; Corbetta et al., 2008; Mitchell, 2008; Krall et al., 2015). Based on this evidence, it has been suggested that bottom-up shifts of attention of a predominant first-person to a third-person perspective, or toward (socially) relevant stimuli might underlie the involvement of the TPJ during altruistic giving (Hare et al., 2010). In line with this notion, we found that response patterns in the right TPJ/pSTS encoded stimulus-driven attentional reorienting in a separate (nonsocial) attention task (Posner and Cohen, 1984). Overlap of this cluster with portions of pSTS and TPJ that predicted generous donations might traditionally be interpreted as evidence for similar mental processes across both tasks. Advanced MVPA techniques allowed us to go beyond previous reports of colocalization and to explicitly test for common neural codes across attentional reorienting and generous giving. Cross-task decoding (Tusche et al., 2014) revealed that response patterns in the pSTS coding for reorienting of attention were indeed reinstated during generous donation choices, but no reactivation was observed in the TPJ (or AI). Evidence of shared neural codes in pSTS points to potential contributions of low-level, domain-general mechanisms to variance in altruistic decision-making and highlights the complexity of mechanisms that should be considered in theoretical and neural models of human altruism. However, the absence of shared neural code in the TPJ supports a socio-cognitive function of inferring other mental states during altruistic choices (Scholz et al., 2009; Young et al., 2010), rather than nonsocial attentional reorienting. MVPA-based comparisons of participants' brain responses across tasks offer a promising methodological approach

to investigate the generalization, or differentiation, of neural codes across tasks, conditions, and mental processes (Poldrack et al., 2009; Corradi-Dell'Acqua et al., 2011; Tusche et al., 2014). This allowed us to distinguish alternative computational accounts of the function of TPJ during altruistic giving, such as domain-general processes of attentional reorienting.

Decomposing altruistic choices into distinct mental and neural computations has direct implications for developing interventions and studying the malleability of human prosociality. For instance, comparisons of process-specific regression weights in donations suggest that, at the group level, affective empathic responses may provide a stronger input into altruistic decisions than cognitive perspective taking. We therefore predict that interventions that strengthen engagement in empathic responses might be more effective to increase altruistic behavior compared with those that modify perspective taking. However, examinations of subject-specific regression weights also found that a subset of participants relied more heavily on perspective taking as input into generous decisions, suggesting that there might be both “empathizers” and “mentalizers” in altruistic giving. Information on the relative influence of relevant subprocesses across participants might inform target-specific training programs to increase altruism, considering how different strategies “fit” with different individuals. While our results stress the relevance of basic social capacities for human altruism, future research needs to examine the generalization of these findings across different altruistic behaviors and decision contexts. Comparisons of recruited mechanisms across altruistic tasks used in economics, psychology, and neuroscience, combined with sensitive analysis techniques such as MVPA, will facilitate the development of more powerful computational models of altruism.

## References

- Adolphs R (2010) Conceptual challenges and directions for social neuroscience. *Neuron* 65:752–767. [CrossRef Medline](#)
- Amodio DM, Frith CD (2006) Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* 7:268–277. [CrossRef Medline](#)
- Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a “theory of mind?” *Cognition* 21:37–46. [CrossRef](#)
- Batson CD (2011) *Altruism in humans*. Oxford: Oxford UP.
- 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* 133:1515–1525. [CrossRef Medline](#)
- Blair RJ (2008) Fine cuts of empathy and the amygdala: dissociable deficits in psychopathy and autism. *Q J Exp Psychol (Hove)* 61:157–170. [CrossRef Medline](#)
- Böckler A, Tusche A, Singer T (2016) The Structure of Human Prosociality: Differentiating Altruistically Motivated, Norm Motivated, Strategically Motivated, and Self-Reported Prosocial Behavior. *Soc Psychol Personal Sci*. Advance Online Publication. Retrieved April 7, 2016. doi: 10.1177/1948550616639650
- Bode S, Haynes JD (2009) Decoding sequential stages of task preparation in the human brain. *Neuroimage* 45:606–613. [CrossRef Medline](#)
- Bode S, Bogler C, Haynes JD (2013) Similar neural mechanisms for perceptual guesses and free decisions. *Neuroimage* 65:456–465. [CrossRef Medline](#)
- Bogler C, Bode S, Haynes JD (2011) Decoding successive computational stages of saliency processing. *Curr Biol* 21:1667–1671. [CrossRef Medline](#)
- Bzdok D, Schilbach L, Vogeley K, Schneider K, Laird AR, Langner R, Eickhoff SB (2012) Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct Funct* 217:783–796. [CrossRef Medline](#)
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215. [CrossRef Medline](#)
- Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58:306–324. [CrossRef Medline](#)
- Corradi-Dell'Acqua C, Hofstetter C, Vuilleumier P (2011) Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *J Neurosci* 31:17996–18006. [CrossRef Medline](#)
- Corradi-Dell'Acqua C, Tusche A, Vuilleumier P, Singer T (2016) Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nat Commun* 7:10904. [CrossRef Medline](#)
- Decety J, Jackson PL (2004) The functional architecture of human empathy. *Behav Cogn Neurosci Rev* 3:71–100. [CrossRef Medline](#)
- Decety J, Lamm C (2007) The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13:580–593. [CrossRef Medline](#)
- de Vignemont F, Singer T (2006) The empathic brain: how, when and why? *Trends Cogn Sci* 10:435–441. [CrossRef Medline](#)
- Eckel CC, Grossman PJ (1996) Altruism in anonymous dictator games. *Games Econ Behav* 16:181–191. [CrossRef](#)
- Eisenberg N, Miller PA (1987) The relation of empathy to pro-social and related behaviors. *Psychol Bull* 101:91–119. [CrossRef Medline](#)
- Fan Y, Duncan NW, de Greck M, Northoff G (2011) Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neurosci Biobehav Rev* 35:903–911. [CrossRef Medline](#)
- Fehr E, Camerer CF (2007) Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn Sci* 11:419–427. [CrossRef Medline](#)
- Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* 425:785–791. [CrossRef Medline](#)
- Fehr E, Schmidt KM (2006) The economics of fairness, reciprocity and altruism: experimental evidence and new theories. *Handb Econ Giving Altruism Reciprocity* 1:615–691. [CrossRef](#)
- FeldmanHall O, Dalgleish T, Evans D, Mobbs D (2015) Empathic concern drives costly altruism. *Neuroimage* 105:347–356. [CrossRef Medline](#)
- Frith CD, Frith U (2006) The neural basis of mentalizing. *Neuron* 50:531–534. [CrossRef Medline](#)
- Frith U (2001) Mind blindness and the brain in autism. *Neuron* 32:969–979. [CrossRef Medline](#)
- Genevsky A, Västfjäll D, Slovic P, Knutson B (2013) Neural underpinnings of the identifiable victim effect: affect shifts preferences for giving. *J Neurosci* 33:17188–17196. [CrossRef Medline](#)
- Gross J, Woelbert E, Zimmermann J, Okamoto-Barth S, Riedl A, Goebel R (2014) Value signals in the prefrontal cortex predict individual preferences across reward categories. *J Neurosci* 34:7580–7586. [CrossRef Medline](#)
- Hampton AN, O'Doherty JP (2007) Decoding the neural substrates of reward-related decision making with functional MRI. *Proc Natl Acad Sci U S A* 104:1377–1382. [CrossRef Medline](#)
- Harbaugh WT, Mayr U, Burghart DR (2007) Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316:1622–1625. [CrossRef Medline](#)
- Hare TA, Camerer CF, Knopfle DT, Rangel A (2010) Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J Neurosci* 30:583–590. [CrossRef Medline](#)
- Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7:523–534. [CrossRef Medline](#)
- Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. *Curr Biol* 17:323–328. [CrossRef Medline](#)
- Hein G, Silani G, Preuschoff K, Batson CD, Singer T (2010) Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68:149–160. [CrossRef Medline](#)
- Hoffmann F, Koehne S, Steinbeis N, Dziobek I, Singer T (2015) Preserved self-other distinction during empathy in autism is linked to network integrity of right supramarginal gyrus. *J Autism Dev Disord* 46:637–648. [CrossRef Medline](#)
- Hutcherson CA, Bushong B, Rangel A (2015) A neurocomputational model of altruistic choice and its implications. *Neuron* 87:451–462. [CrossRef Medline](#)
- Kahnt T, Grueschow M, Speck O, Haynes JD (2011a) Perceptual learning and decision-making in human medial frontal cortex. *Neuron* 70:549–559. [CrossRef Medline](#)
- Kahnt T, Heinzle J, Park SQ, Haynes JD (2011b) Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *Neuroimage* 56:709–715. [CrossRef Medline](#)

- Kahnt T, Park SQ, Haynes JD, Tobler PN (2014) Disentangling neural representations of value and salience in the human brain. *Proc Natl Acad Sci U S A* 111:5000–5005. [CrossRef Medline](#)
- Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685. [CrossRef Medline](#)
- Kanske P, Böckler A, Singer T (2015a) Models, mechanisms and moderators dissociating empathy and theory of mind in: Current topics in behavioral neurosciences—social behavior from rodents to humans: neural foundations and clinical implications (Wöhr M, Krach S, eds), pp 1–14. Luxembourg: Springer.
- Kanske P, Böckler A, Trautwein FM, Singer T (2015b) Are empathizers mentalizers? Linking the neural mechanisms underlying affective and cognitive routes to understanding others. *Neuroimage* 122:6–19. [CrossRef Medline](#)
- Krall SC, Rottschy C, Oberwaland E, Bzdok D, Fox PT, Eickhoff SB, Fink GR, Konrad K (2015) The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Struct Funct* 220:587–604. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Kurth F, Zilles K, Fox PT, Laird AR, Eickhoff SB (2010) A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Struct Funct* 214:519–534. [CrossRef Medline](#)
- 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* 54:2492–2502. [CrossRef Medline](#)
- Ma Y, Wang C, Han S (2011) Neural responses to perceived pain in others predict real-life monetary donations in different socioeconomic contexts. *Neuroimage* 57:1273–1280. [CrossRef Medline](#)
- Mar RA (2011) The neural bases of social cognition and story comprehension. *Annu Rev Psychol* 62:103–134. [CrossRef Medline](#)
- Masten CL, Morelli SA, Eisenberger NI (2011) An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. *Neuroimage* 55:381–388. [CrossRef Medline](#)
- Mathur VA, Harada T, Lipke T, Chiao JY (2010) Neural basis of extraordinary empathy and altruistic motivation. *Neuroimage* 51:1468–1475. [CrossRef Medline](#)
- Meffert H, Gazzola V, den Boer JA, Bartels AA, Keysers C (2013) Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. *Brain* 136:2550–2562. [CrossRef Medline](#)
- Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53:103–118. [CrossRef Medline](#)
- Mitchell JP (2008) Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb Cortex* 18:262–271. [CrossRef Medline](#)
- Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang XR, Just M, Newman S (2004) Learning to decode cognitive states from brain images. *Mach Learn* 57:145–175. [CrossRef](#)
- Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci U S A* 103:15623–15628. [CrossRef Medline](#)
- Morelli SA, Rameson LT, Lieberman MD (2014) The neural components of empathy: predicting daily prosocial behavior. *Soc Cogn Affect Neurosci* 9:39–47. [CrossRef Medline](#)
- Morishima Y, Schunk D, Bruhin A, Ruff CC, Fehr E (2012) Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron* 75:73–79. [CrossRef Medline](#)
- Muller HJ, Rabbitt PMA (1989) Reflexive and voluntary orienting of visual-attention: time course of activation and resistance to interruption. *J Exp Psychol Human* 15:315–330. [CrossRef Medline](#)
- Penner LA, Dovidio JF, Piliavin JA, Schroeder DA (2005) Prosocial behavior: multilevel perspectives. *Annu Rev Psychol* 56:365–392. [CrossRef Medline](#)
- Poldrack RA, Halchenko YO, Hanson SJ (2009) Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychol Sci* 20:1364–1372. [CrossRef Medline](#)
- Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. *Science* 310:1963–1966. [CrossRef Medline](#)
- Posner MI, Cohen Y (1984) Components of visual orienting. *Attention Percept Mot* 10:531–556.
- Rameson LT, Morelli SA, Lieberman MD (2012) The neural correlates of empathy: experience, automaticity, and prosocial behavior. *J Cogn Neurosci* 24:235–245. [CrossRef Medline](#)
- Saxe R, Carey S, Kanwisher N (2004) Understanding other minds: linking developmental psychology and functional neuroimaging. *Annu Rev Psychol* 55:87–124. [CrossRef Medline](#)
- Scholz J, Triantafyllou C, Whitfield-Gabrieli S, Brown EN, Saxe R (2009) Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS One* 4:e4869. [CrossRef Medline](#)
- Schurz M, Radua J, Aichhorn M, Richlan F, Perner J (2014) Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* 42:9–34. [CrossRef Medline](#)
- Silani G, Bird G, Brindley R, Singer T, Frith C, Frith U (2008) Levels of emotional awareness and autism: an fMRI study. *Soc Neurosci* 3:97–112. [CrossRef Medline](#)
- Silani G, Lamm C, Ruff CC, Singer T (2013) Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *J Neurosci* 33:15466–15476. [CrossRef Medline](#)
- Singer T (2006) The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci Biobehav Rev* 30:855–863. [CrossRef Medline](#)
- Singer T (2012) The past, present and future of social neuroscience: a European perspective. *Neuroimage* 61:437–449. [CrossRef Medline](#)
- Singer T, Lamm C (2009) The social neuroscience of empathy. *Ann N Y Acad Sci* 1156:81–96. [CrossRef Medline](#)
- Soon CS, Brass M, Heinze HJ, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11:543–545. [CrossRef Medline](#)
- Stöber J (2001) The Social Desirability Scale-17 (SDS-17): convergent validity, discriminant validity, and relationship with age. *Eur J Psychol Assess* 17:222–232. [CrossRef](#)
- Tankersley D, Stowe CJ, Huettel SA (2007) Altruism is associated with an increased neural response to agency. *Nat Neurosci* 10:150–151. [CrossRef Medline](#)
- Telzer EH, Masten CL, Berkman ET, Lieberman MD, Fuligni AJ (2011) Neural regions associated with self control and mentalizing are recruited during prosocial behaviors towards the family. *Neuroimage* 58:242–249. [CrossRef Medline](#)
- Tusche A, Bode S, Haynes JD (2010) Neural responses to unattended products predict later consumer choices. *J Neurosci* 30:8024–8031. [CrossRef Medline](#)
- Tusche A, Kahnt T, Wisniewski D, Haynes JD (2013) Automatic processing of political preferences in the human brain. *Neuroimage* 72:174–182. [CrossRef Medline](#)
- Tusche A, Smallwood J, Bernhardt BC, Singer T (2014) Classifying the wandering mind: revealing the affective content of thoughts during task-free rest periods. *Neuroimage* 97:107–116. [CrossRef Medline](#)
- Van Overwalle F (2009) Social cognition and the brain: a meta-analysis. *Hum Brain Mapp* 30:829–858. [CrossRef Medline](#)
- Waytz A, Zaki J, Mitchell JP (2012) Response of dorsomedial prefrontal cortex predicts altruistic behavior. *J Neurosci* 32:7646–7650. [CrossRef Medline](#)
- Wisniewski D, Reverberi C, Tusche A, Haynes JD (2015) The neural representation of voluntary task-set selection in dynamic environments. *Cereb Cortex* 25:4715–4726. [CrossRef Medline](#)
- Young L, Dodell-Feder D, Saxe R (2010) What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48:2658–2664. [CrossRef Medline](#)
- Zaki J, Mitchell JP (2011) Equitable decision making is associated with neural markers of intrinsic value. *Proc Natl Acad Sci U S A* 108:19761–19766. [CrossRef Medline](#)
- Zanon M, Novembre G, Zangrando N, Chittaro L, Silani G (2014) Brain activity and prosocial behavior in a simulated life-threatening situation. *Neuroimage* 98:134–146. [CrossRef Medline](#)