Behavioral/Cognitive

Decomposing Gratitude: Representation and Integration of Cognitive Antecedents of Gratitude in the Brain

Hongbo Yu,[1,2,3*] Xiaoxue Gao,[1,2*] Yuanyuan Zhou,[1,2,4] and Xiaolin Zhou[1,2,5,6,7,8]

[1]Center for Brain and Cognitive Sciences, Peking University, Beijing 100871, China, [2]School of Psychological and Cognitive Sciences, Peking University, Beijing 100871, China, [3]Department of Experimental Psychology, University of Oxford, OX1 3UD, Oxford, United Kingdom, [4]School of Economics and Management, Tsinghua University, Beijing 100084, China, [5]Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China, [6]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China, [7]Institute of Psychological and Brain Sciences, Zhejiang Normal University, Zhejiang 321004, China, and [8]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China

Gratitude is a typical social-moral emotion that plays a crucial role in maintaining human cooperative interpersonal relationship. Although neural correlates of gratitude have been investigated, the neurocognitive processes that lead to gratitude, namely, the representation and integration of its cognitive antecedents, remain largely unknown. Here, we combined fMRI and a human social interactive task to investigate how benefactor's cost and beneficiary's benefit, two critical antecedents of gratitude, are encoded and integrated in beneficiary's brain, and how the neural processing of gratitude is converted to reciprocity. A coplayer decided whether to help a human participant (either male or female) avoid pain at his/her own monetary cost; the participants could transfer monetary points to the benefactor with the knowledge that the benefactor was unaware of this transfer. By independently manipulating monetary cost and the degree of pain reduction, we could identify the neural signatures of benefactor's cost and recipient's benefit and examine how they were integrated. Recipient's self-benefit was encoded in reward-sensitive regions (e.g., ventral striatum), whereas benefactor-cost was encoded in regions associated with mentalizing (e.g., temporoparietal junction). Gratitude was represented in perigenual anterior cingulate cortex (pgACC), the strength of which correlated with trait gratitude. Dynamic causal modeling showed that the neural signals representing benefactor-cost and self-benefit passed to pgACC via effective connectivities, suggesting an integrative role of pgACC in generating gratitude. Moreover, gyral ACC plays an intermediary role in converting gratitude representation into reciprocal behaviors. Our findings provide a neural mechanistic account of gratitude and its role in social-moral life.

Key words: cognitive antecedents; dynamic causal modeling; fMRI; gratitude; integration; reciprocity

Significance Statement

Gratitude plays an integral role in subjective well-being and harmonious interpersonal relationships. However, the neurocognitive processes through which various components and antecedents of gratitude are integrated remain largely unknown. We developed a new interpersonal paradigm to independently and parametrically manipulate two antecedents of gratitude in a helping context, namely, the benefit to beneficiary and the cost to benefactor, to examine their representation and integration in the beneficiary's brain using fMRI. We found the neural encoding of self-benefit and benefactor-cost in reward- and mentalizing-related brain areas, respectively. More importantly, by examining effective connectivity, we showed that these componential signals are passed to perigenual anterior cingulate cortex, which tracks trial-by-trial gratitude levels. Our study thus provides a neural mechanistic account of gratitude.

Introduction

Gratitude plays a crucial role in social and moral life (McCullough et al. 2001; Harpham, 2004; McCullough and Tsang, 2004; Leithart, 2014; Kristjánsson, 2015; Manela, 2015) and is regarded as a typical social-moral emotion (Haidt, 2003; Algoe and Haidt, 2009). There are various ways of analyzing gratitude.

Received Oct. 26, 2017; revised April 7, 2018; accepted April 17, 2018.

Author contributions: H.Y. and X.G. edited the paper; Y.Z. wrote the first draft of the paper. H.Y., X.G., Y.Z., and X.Z. designed research; H.Y. and X.G. performed research; H.Y., X.G., and Y.Z. analyzed data; H.Y., X.G., Y.Z., and X.Z. wrote the paper.

This work was supported by the Natural Science Foundation of China Grants 31630034 and J1103602 and the National Basic Research Program of China 973 Program: 2015CB856400. H.Y. was supported by British Academy Newton International Fellowship NF160700. An earlier version of the manuscript and data analysis have been presented to the research groups led by Dr. Paul Glimcher, Dr. Matthew Rushworth, Dr. Molly Crockett, and Dr. Bolton

For instance, Tesser et al. (1968) proposed an influential model of the determinants (or cognitive antecedents) of gratitude: benefit to the beneficiary, cost to the benefactor, and intention of the benefactor. Gratitude also has desirable consequences for a "good life" (Watkins, 2014), such as increasing subjective well-being (Emmons and McCullough, 2003; Morgan et al., 2017), cultivating social relationship (Algoe et al., 2008; Algoe and Haidt, 2009; Algoe, 2012), and promoting reciprocal and cooperative behaviors (Bartlett and DeSteno, 2006; Tsang, 2006; Tangney et al., 2007; DeSteno et al., 2010; Yu et al., 2017; Tsang and Martin, 2018; Yost-Dubrow and Dunham, 2018).

The neural correlates of gratitude have been investigated in recent neuroimaging studies (e.g., Zahn et al., 2009; Fox et al., 2015; Kini et al., 2016; Karns et al., 2017; Yu et al., 2017). Fox et al. (2015), following the theoretical model of Tesser et al. (1968), elicited gratitude by manipulating the benefactor's cost and the beneficiary's benefit through gratitude-inducing scenarios. They found a correlation between self-reported gratitude and the neural activation in medial prefrontal cortex (MPFC) and perigenual anterior cingulate cortex (pgACC), regions associated with value representation and integration (Bartra et al., 2013; Sescousse et al., 2013; Kolling et al., 2016). However, this study has not identified neural representation of the cognitive antecedents of gratitude. Moreover, the scenario-based paradigm made it difficult to investigate the neural basis underlying the social consequences of gratitude, as the emotion was imagined and hypothetical, and no social interaction was presented. Others adopted social interactive tasks to elicit gratitude and subsequent prosocial behaviors (Kini et al., 2016; Karns et al., 2017; Yu et al., 2017). Using a pain alleviation task, Yu et al. (2017) manipulated the benefactor's intention (voluntary vs forced) in the costly reduction of pain stimulation delivered to the participants. They found that, compared with forced help, voluntary help had stronger effects on participants' subsequent prosocial behaviors (e.g., reciprocity). Consistent with Fox et al. (2015), situations associated with higher gratitude elicited activations in ventral MPFC/pgACC regions (see also Kini et al., 2016; Karns et al., 2017; Yu et al., 2017). These findings echo the theoretical construal of gratitude and other positive emotions as a form of subjective value (Fredrickson, 2004; Todd, 2014).

While the previous neuroscience studies of gratitude have identified the neural correlates of gratitude, none has adequately addressed the questions of how the brain encodes the cognitive antecedents of gratitude (e.g., self-benefit, benefactor-cost) and integrates them to give rise to gratitude and subsequent reciprocal/prosocial behaviors. According to the appraisal theory of emotion, these appraisal processes characterize a specific emotion and differentiate it from other emotions (Lazarus and Smith, 1988; Frijda, 1993; Ellsworth and Scherer, 2003). To address these questions, we developed a novel social interactive task based on the pain alleviation task (Hu et al., 2017; Yu et al., 2017).

We independently manipulated the benefit to the participants (i.e., pain reduction) and the cost to the benefactor (i.e., monetary cost for pain reduction) in a helping situation. This allowed us to identify neural representations of benefit and cost and to examine the brain processes through which they are integrated and give rise to gratitude. We predicted that self-benefit and benefactor-cost are encoded in regions associated with valuation (e.g., ventral striatum [VS]) (Bartra et al., 2013) and mentalizing (e.g., temporoparietal junction) (Van Overwalle and Baetens, 2009; David et al., 2017) networks, respectively, whereas the gratitude is represented in pgACC (Yu et al., 2017).

Dynamic causal modeling (DCM) was used to directly test this hypothesized integration process, namely, signals related to benefactor's cost and beneficiary's benefit pass from their dedicated regions to pgACC for integration. Finally, we examined the neural process by which gratitude is converted into reciprocal behaviors.

Materials and Methods

Participants

Thirty-six healthy right-handed undergraduate students took part in the experiment. Five participants were excluded from data analysis due to misunderstanding of instructions (1 participant) or excessive head movements (>3 mm of locomotion or 3 degrees of rotation; 4 participants), leaving 31 participants (15 males; mean age 23.0 ± 1.9 years) for data analysis. None of the participants reported any history of psychiatric, neurological, or cognitive disorders. Informed written consent was obtained from each participant before experiment. The study was performed in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the School of Psychological and Cognitive Sciences, Peking University. The data for this study and that for Yu et al. (2017) were collected in two different cities and was separated by ~2 years. Participants who had taken part in any studies involving social interaction were excluded from participating in this study.

Experimental design and statistical analyses

Overview. The experiment consisted of three phases. In the first, pain titration phase, we measured each participant's pain threshold and determined the physical intensity (in mA) of the shocks that corresponded to four levels of subjective pain experience. In the second phase, the participants performed a decision-making task where they could earn a monetary bonus by taking pain stimulations. The purpose of this task was twofold: (1) it familiarized the participants with the pain stimulation; and (2) through this task, the participants (and their coplayers, ostensibly) earned endowments for the main task (see below). In the third phase, the participants performed the social interactive task in the scanner, which was our main task.

Randomization and pain titration. Each participant came to the scanning room individually. Upon arrival, he/she met three confederates and was told that they would later perform an interactive task together through internet. At least one confederate was of the same sex as the participant, and at least one was of the opposite sex. The player undergoing MRI scanning (i.e., the participant) had a different role in the game than the others (i.e., the confederates). Together, they were told that their roles in the game were randomly determined upon their signup.

Pain titration. After the participants met the three confederates, they were led to a separate testing room to do pain titration. Intraepidermal electrical stimulation (Inui et al., 2002) was delivered during the pain titration phase. Three stainless-steel concentric bipolar needle electrodes were attached to the left-hand dorsum of the participant. These three electrodes are separated by an equal distance of 6 mm. Each electrode has a needle cathode (length: 0.1 mm, Ø: 0.2 mm) surrounded by an acylindrical anode (Ø: 1.4 mm) (Inui et al., 2002). The reliability of the pain delivery system has been demonstrated in a number of fMRI studies on pain perception (Hu et al., 2015) and social-affective processing (Yu et al., 2014, 2015). Pain calibration begun with 12 repeated pulses, each of which was 0.2 mA and lasted for 0.5 ms. A 10 ms interval was inserted between pulses. Then we gradually increased the intensity of each single pulse until the participant reported 6 on an 8-point pain scale (1 = not painful, 8 = intolerable). The participants reported that they could only experience the whole train of pulse as a single stimulation, rather than separate shocks. Then we delivered stimulations that contained 8, 4, and

Chau. The authors have received invaluable comments and suggestions concerning behavioral modeling and imaging analysis from discussions with these groups. We thank Joshua Brown for sharing the activation maps from Kini et al. (2016) for comparison; and Dr. Philip Blue for commenting on an earlier version of the manuscript.

The authors declare no competing financial interests.

*H.Y. and X.G. contributed equally to this work.

Correspondence should be addressed to Dr. Xiaolin Zhou, School of Psychological and Cognitive Sciences Peking University, Beijing, 100871 China. E-mail: xz104@pku.edu.cn.

DOI:10.1523/JNEUROSCI.2944-17.2018

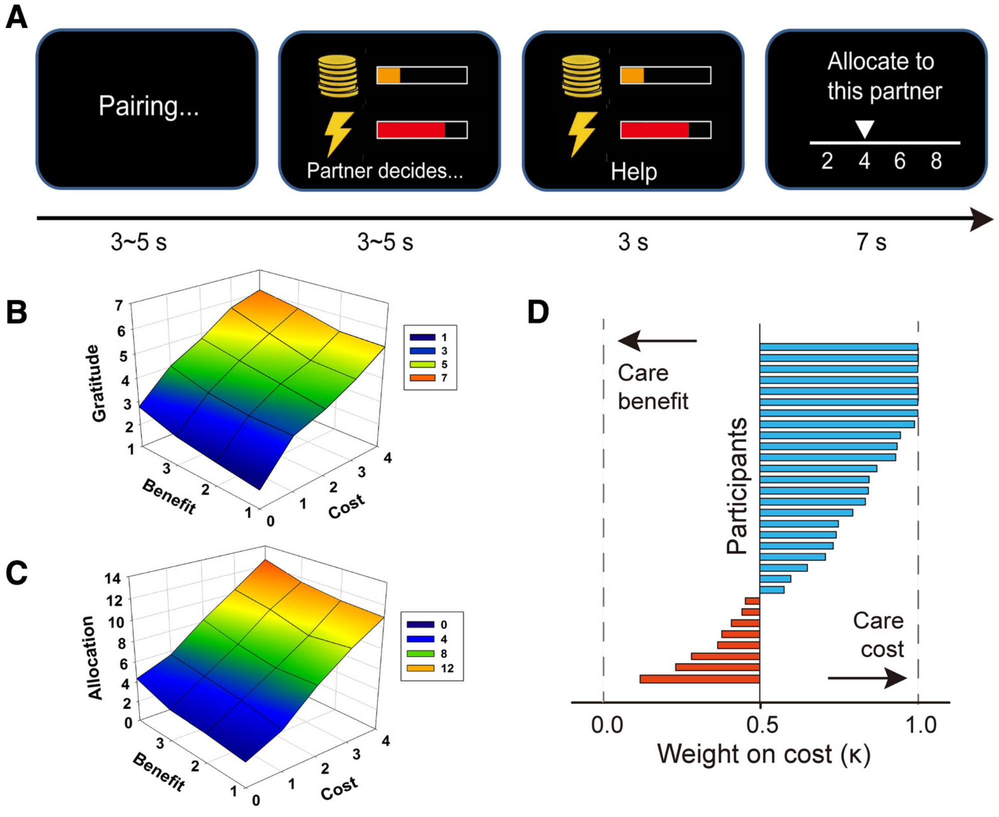

Figure 1. Experiment procedure and behavioral results. ***A***, At the beginning of each trial, the participants were (anonymously and ostensibly) paired with 1 of 3 coplayers. Then the participants saw a pain-money pair and waited for the coplayer's decision. If the coplayer chose Help, then the coplayer lost the corresponding amount of bonus while the participants would be exempted from the pain stimulation on this trial. If the coplayer chose NoHelp, then the coplayer could keep the bonus while the participants had to receive the corresponding pain stimulation. The presentation of the coplayer's decision was defined as the critical events in the fMRI data analysis. At the end of the trial, the participant could allocate 20 Yuan (~$3 U.S.) between himself/herself and the coplayer, with the knowledge that the coplayer was not aware of this procedure. ***B***, ***C***, Postscan gratitude rating and allocation during scanning (i.e., reciprocity) as a function of self-benefit and benefactor-cost. ***D***, Relative weight of benefactor-cost over self-benefit in gratitude rating.

2 repeated pulses. All the participants reported that the four levels of pain stimulation (2, 4, 8, 12 repeated pulses) were clearly distinguishable, and were instructed that these four levels of pain stimulation would be used in the later tasks.

Pain-money exchange task (behavioral). After the pain calibration phase, the participants performed a pain-money exchange task, where they made a series of decisions as to whether to accept electric shocks in exchange for a certain amount of money. Each of the four pain levels, as determined in the calibration, were paired with different amount of monetary gains. If the participants chose "accept," they would receive the indicated money in that trial but would also receive the corresponding pain stimulation immediately. There were 15 blocks; each contained one trial for each pain level. The order of pain levels in each block was randomized. In the first block, the monetary gains paired with the 4 levels of pain was 0.5, 1.0, 1.5, and 2.0 Yuan (1 Yuan ~ $0.16 U.S.), respectively. These were the baseline monetary bonuses. From the second block onward, the monetary gain paired with a certain pain level would increase/decrease if the participants had rejected/accepted the offer on the same pain level in the preceding block. The length of the incremental step for each pain level was the baseline monetary bonus multiplying a converging factor (round to the nearest tenths), starting from 1.5. For example, the incremental steps after the first block were 0.8, 1.5, 2.3, and 3.0 Yuan for the four levels respectively. Once the participants' preference reversed (i.e., "reject" to "accept" or "accept" to "reject"), the converging factor reduced by 0.5, until it reached 0.5. After that, the incremental step decreased by 0.1 when preference reverse occurred (compare Shen et al., 2016). The participant was told that all the other coplayers completed the same task and earned their own payoffs during this phase. Unbeknownst to the participant, the payoff for him/her in this task was predetermined to be 18 Yuan (~$2.8 U.S.). Because the behaviors in this task was not relevant to the aim of this study, we skipped this for brevity.

Help-receiving task (fMRI). In the scanning phase, the participants performed the help-receiving task while their BOLD responses were measured with MRI (Fig. 1*A*). In each trial, the participants were paired with a partner and then a pain-money pair was presented, indicating the level of pain the participant potentially had to receive and the monetary cost the partner needed to spend to eliminate the pain for the participants if he/she chose so. The pain stimulation had four levels (1–4, corresponding to four pain levels in the titration and pain-money exchange task) and the monetary cost had 5 levels (0–4, corresponding to 0%, 25%, 50%, 75%, and 100% of the partner's monetary bonus in the pain-money exchange task). The participants first saw the pain-money pair and then saw the partner's decision (Help or NoHelp). If the partner helped, then the partner would lose the corresponding amount of bonus while the participants would be exempted from the pain stimulation for this trial. We told the participants that, at the end of the task, 20 trials would be randomly selected and actualized. For each selected trial, if the partner chose to help, the participants would not receive the pain stimulation for that trial; otherwise, the participants had to receive the indicated number of shocks. The monetary bonus and shocks were implemented before the participants left the MRI testing room. At the end of each trial, the participants were asked to divide 20 Yuan (~$3 U.S.) between themselves and the partner paired in this trial. They could adjust the amount of allocation in increments of 2 Yuan. The participants' final payoff was the payoff gained in the Pain-money exchange task and the average amount of endowment the participants left for themselves across the randomly selected trials in the fMRI task. The participants were told that the partner was unaware of the allocation procedure. This was to prevent the participant from suspecting the partner's intention to help (e.g., strategically helping the participant for the sake of getting more allocation from the participant). To avoid any influence of the previous encounter and reputation concerns, we explicitly told the participants

Table 1. Distribution of NoHelp trials in different cost-benefit conditions

	Cost				
No. of NoHelp trials	0	1	2	3	4
Benefit					
1	0	3	3	5	5
2	0	3	3	4	5
3	0	2	3	3	3
4	0	1	2	3	3

that the partner in each trial may or may not be the same partner as in the last trial. The partner's help decision was binary so that the partner either accepted the cost as indicated in that trial and thus reduced the participant's pain to 0, or rejected the cost and left the pain stimulation to the participant as indicated.

The two independent variables we manipulated in the main task were the intensity of the pain stimulation that the partner took on (i.e., self-benefit) and the cost of the partner for doing so (i.e., benefactor-cost). The partner's decision to help or not was predetermined. There were 20 possible combinations of self-benefit and benefactor-cost for the Help trials, thus forming a 4 (Benefit: 1, 2, 3, 4) × 5 (Cost: 0, 1, 2, 3, 4) within-subject design. The NoHelp trials were included as fillers. The experiment thus consisted of 111 trials (3 trials for each of the above 20 Help conditions and 51 filler trials for NoHelp condition, the distribution of which can be found in Table 1). When determining the order of conditions, we first created a randomized sequence for all the 111 trials (both Help and NoHelp trials included). We then divided the sequence into 3 parts with equal number of trials. These 3 parts were assigned to 3 runs of MRI scanning in a Latin-square manner across participants. Each run consisted of 37 trials and lasted for ~15 min.

After the experiment, the participants recalled and rated their gratitude feeling for all the Help conditions on a scale of 1 (not at all) to 7 (very strong), one score for each cost-benefit combination. We therefore had 20 gratitude scores from each participant. The participants then completed the Gratitude Trait questionnaire (McCullough et al., 2002). A postscan interview was conducted to examine whether the participants had any suspicion about our experimental manipulation. No participant reported any suspicion of experiment manipulation or the existence of partners.

Analysis of the behavioral data. We analyzed the behavioral data using R (www.r-project.org). To obtain the standard coefficients and to enable comparison of parameters between participants, all the data were normalized within participant before analysis. First, to test whether and how self-benefit and benefactor-cost contributed to gratitude and reciprocity, we fit four general linear mixed models for the monetary allocation in the main task and the postscan gratitude rating (Tables 2, 3) separately with participant as a random effect. By-subject random slopes for each fixed effect were also included in the models (Barr et al., 2013). Model 1 included cost as single predictor. Model 2 included benefit as single predictor. Model 3 included both cost and benefit as predictors. Model 4 included cost, benefit, and the interaction between these two predictors as predictors. Model goodness of fit was assessed using the Bayesian information criterion (BIC; Lewandowsky and Farrell, 2010), which takes into account both model fitness and complexity. Parameters were estimated based on the best model (lowest BIC).

Second, to test the relationship between gratitude and allocation, we fit a general linear mixed model with gratitude rating as the predictor, online allocation as the dependent variable, and participant as random effect. For each participant, we conducted linear regression with gratitude rating as predictor and online allocation as dependent variable to characterize the influence of gratitude on reciprocity individually. The β value of this model indicates the extent to which gratitude feeling is translated into reciprocal behaviors. In other words, the β value reflects the exchange rate between gratitude for each participant.

Third, to characterize the weights of cost and benefit in the generation of gratitude feelings, we performed a linear regression with cost and benefit as predictors and postscan gratitude rating as the dependent variable. Because both cost and benefit contributed to gratitude and because no interaction was observed, we assumed that the weights of benefit and cost could be summed up to one in each model. Based on the weights of cost and benefit (Fig. 1D), we calculated a trial-by-trial "constructed gratitude" for each participant individually. The conducted gratitude on trial *i* was defined as follows:

$$\text{Constructed gratitude}_i = k * \text{cost}_i + (1 - k) * \text{benefit}_i \qquad (1)$$

where cost_i and benefit_i are cost and benefit on trial *i* and *k* is the individual weight for cost. To check the extent to which the constructed gratitude index captures gratitude ratings, we regressed constructed gratitude in each cost-benefit combination against participants' gratitude rating in the same condition using mixed linear effect model with by-subject random slopes for the fixed effect. We found that the constructed gratitude reliably tracked gratitude rating ($\beta = 0.88 \pm 0.09$, marginal $R^2 = 0.44$).

MRI data acquisition and preprocessing. Images were acquired using a 3.0 T MR scanner (GE MR750) with a standard head coil at the Peking University. T2*-weighted functional images were acquired in 35 axial slices parallel to the anterior commissural–posterior commissural line with no interslice gap, affording full-brain coverage. Images were acquired using an EPI pulse sequence (TR = 2000 ms; TE = 30 ms; flip angle = 90°; FOV = 192 mm × 192 mm; slice thickness = 4 mm). An ascending, interleaved slice acquisition order was used.

Image preprocessing and analysis were conducted with the Statistical Parametric Mapping software SPM8 (Wellcome Trust Department of Cognitive Neurology, London). Images were slice-time corrected (with the middle slice as the reference), motion corrected, normalized to MNI space, spatially smoothed using an 8 mm FWHM Gaussian filter, and temporally filtered using a high-pass filter with 1/128 Hz cutoff frequency.

fMRI data analysis. We performed event-related fMRI analyses of participants' neural responses at the time at which they viewed the partner's decisions. For the first-level (within-participant) statistical analysis, 2 separate GLMs were created to address different questions and avoid issues of colinearity between regressors.

To identify brain areas that encode benefactor-cost and self-benefit processing, in GLM1, we modeled Help_condition as five separate regressors, starting from the time the benefactor's choice was revealed and spanning the duration of this event, which was 3 s: LowCost_LowBenefit (including conditions where Cost = 1 or 2 and Benefit = 1 or 2), HighCost_LowBenefit (including conditions where Cost = 3 or 4 and Benefit = 1 or 2), LowCost_HighBenefit (including conditions where Cost = 1 or 2 and Benefit = 3 or 4), HighCost_HighBenefit (including conditions where Cost = 3 or 4 and Benefit = 3 or 4), and NoCost (including conditions where Cost = 0 and Benefit = 1, 2, 3, or 4). Regressors of no interest included NoHelp_condition (onset of the NoHelp decision), Pair_presentation (onset of the money-shock pair), Allocation (the decision period for allocation), Pain_cue (the cue for pain stimulation threat), and Miss_allocation (the missing decision period for allocation). Six movement parameters were included as regressors of no interest. We defined two contrasts for the main effect of cost and main effect of benefit as follows:

Main effect of Cost (Contrast 1):

$$(\text{HighCost\_HighBenefit} + \text{HighCost\_LowBenefit}) > (\text{LowCost\_HighBenefit} + \text{LowCost\_LowBenefit})$$

Main effect of Benefit (Contrast 2):

$$(\text{HighCost\_HighBenefit} + \text{LowCost\_HighBenefit}) > (\text{HighCost\_LowBenefit} + \text{LowCost\_LowBenefit}).$$

To assess the neural correlates of gratitude and subsequent reciprocity, we created GLM2 in which all Help trials were grouped into a single regressor with two first-level parametric modulators: the constructed gratitude (see Eq. 1) and monetary allocation. Contrasts 3 and 4 were

Table 2. Models of postscan gratitude rating[a]

Model	Predictors	df	$-2 \ln L$	BIC	ΔBIC	w	Term	Beta	SE	t
1	Cost	6	2086.0	2124.6	273.6	0.0	Cost	0.83	0.06	13.06
2	Benefit	6	2426.0	2464.5	613.5	0.0	Benefit	0.33	0.11	2.96
3	Cost + benefit	10	1856.4	1920.7	69.7	0.0	Cost	0.83	0.06	13.06
							Benefit	0.33	0.11	2.96
4[b]	Cost + benefit	15	1754.5	1851.0	0.0	1.0	Cost	0.80	0.16	5.02
	+ cost × benefit						Benefit	0.30	0.07	4.58
							Interaction	0.01	0.05	0.26

[a] L, Likelihood; ΔBIC = BIC − min(BIC); w (exceedance probability) = exp(−0.5 × ΔBIC)/sum(exp(−0.5 × ΔBIC)) (Lewandowsky and Farrell, 2010).

[b] Best model.

Table 3. Models of monetary allocation[a]

Model	Predictors	df	$-2 \ln L$	BIC	ΔBIC	w	Term	Beta	SE	t
1	Cost	6	17291.6	17340.0	43.0	0.0	Cost	0.77	0.10	7.96
2	Benefit	6	17390.8	17439.0	142.0	0.0	Benefit	0.62	0.09	7.10
3[b]	Cost + benefit	10	17217.4	17297.0	0.0	1.0	Cost	0.78	0.11	7.43
							Benefit	0.64	0.10	6.11
4	Cost + benefit	15	17213.0	17333.0	36.0	0.0	Cost	0.62	0.17	3.63
	+ cost × benefit						Benefit	0.47	0.16	2.98
							Interaction	0.07	0.06	1.17

[a] L, Likelihood; ΔBIC = BIC − min(BIC); w (exceedance probability) = exp(−0.5 × ΔBIC)/sum(exp(−0.5 × ΔBIC)) (Lewandowsky and Farrell, 2010).

[b] Best model.

defined as the positive effect of the parametric modulators, respectively. Regressors of no interest were the same as GLM1.

To investigate how the neural processing of receiving help could predict subsequent reciprocal behavior, we defined a first-level contrast Help > NoHelp in GLM2 to capture the main effect of receiving help. At the second (group) level, we defined a one-sample t test based on the first level contrasts maps and included the individual's "exchange rate" (i.e., regression coefficient between gratitude and reciprocity for each participant) as a covariate (or parametric modulator). Positive effect of this second-level parametric modulator could identify brain areas whose main effect of receiving help (Help > NoHelp) positively correlated with individual exchange rate between gratitude and reciprocity. In other words, participants who show a larger contrast effect in these areas more readily translate their grateful feelings into reciprocal behaviors.

For all GLMs, the events in each regressor were convolved with the canonical hemodynamic response function. Second-level models were constructed as one-sample t tests using contrast images from the first-level models. All results were corrected for multiple comparisons using the threshold of peak-level $p < 0.001$ (uncorrected) combined with cluster-level $p < 0.05$ (FWE-corrected). This cluster-level threshold was determined using a Monte Carlo simulation as implemented in the AFNI AlphaSim package (http://afni.nimh.nih.gov/pub/dist/doc/manual/AlphaSim.pdf). Statistic parametric maps are presented at this threshold unless otherwise noted.

Effective connectivity analysis. To address the question of how the dynamic neural network generates the signal of gratitude based on the processing of benefactor-cost and self-benefit, we used dynamic causal modeling (DCM; Friston et al., 2003) to examine the effective connectivity between brain areas that encode gratitude, benefactor-cost, and self-benefit. Our hypothesis was that the brain areas encoding benefactor-cost (e.g., right temporoparietal junction [rTPJ]) and self-benefit (e.g., VS) feed information to pgACC, where the processing of the two antecedents is integrated and a gratitude signal is generated.

Thus, we selected the volumes of interest (VOIs) based on the Contrasts 1, 2, and 3. Specifically, we chose right VS (rVS; peak MNI: 12, 20, −15), rTPJ (peak MNI: 48, −52, 31), and pgACC (peak MNI: 9, 50, 1) as our VOIs for self-benefit, benefactor-cost, and gratitude, respectively. The VOI data were the first eigenvalues of the signals within a 3-mm-radius sphere centered at these peak voxels. We set the slice timings of the VOIs as the reference slice in the slice time correction during the preprocessing (i.e., applying previous slice time correction) (cf. Kiebel et al., 2007).

We built and compared 7 families of models (33 single models) differing in the direction of intrinsic connectivity (bilateral or unilateral), the presence or absence of the intrinsic connectivity between rVS and rTPJ (see Fig. 5). Four Help conditions (i.e., HighCost_HighBenefit, HighCost_LowBenefit, LowCost_HighBenefit, LowCost_LowBenefit) were combined into one single regressor and used as model input. For each single model, we assumed that the input was entered through both rVS and rTPJ. Within each model family, modulatory effects (i.e., the four discrete conditions) were placed on different intrinsic connectivities in different individual models. The same set of models were built and compared where we used left VS instead of rVS as VOI. Model comparison showed a similar pattern as that using rVS. Therefore, in the interest of space, we only reported the results based on rVS VOI. These models were compared using Bayesian model selection, which uses a Bayesian framework to calculate the "model evidence" of each model. The model evidence represents the trade-off between model simplicity and accuracy (Penny et al., 2004). Here, Bayesian model selection was implemented using a random-effects analysis (i.e., assuming that the model structure might vary across participants) that is robust to the presence of outliers (Stephan et al., 2009). When comparing model families, all models within a family were averaged using Bayesian model averaging, and the exceedance probabilities were calculated for each model family (Penny et al., 2010). Model parameters (i.e., connectivity strength) were estimated based on the winning model through Bayesian model averaging.

Functional connectivity. To complement the analysis of effective connectivity, a correlation-based connectivity measure, psychophysiological interaction), was performed (Friston et al., 1997). The rVS VOI, which was extracted for the DCM analysis, was used as seed region. Four separate psychophysiological interaction models were estimated, each having one of the four critical conditions as the psychological factor. We then extracted the connectivity strength (or correlation) from an rTPJ ROI, which consisted of 27 voxels around the peak rTPJ cluster defined by Contrast 1.

Results

Behavioral results

As a manipulation check, we first tested whether gratitude and reciprocity increased as a function of both self-benefit and benefactor-cost. For the gratitude rating (Table 2), the model with two main effects and the interaction was the best model. However, as can be seen from the table, only the main effects of benefit and cost were significant (benefit: $\beta = 0.30 \pm 0.07$, $t = 4.58$; cost: $\beta = 0.80 \pm 0.16$, $t = 5.02$); the interaction term did not reach significance ($\beta = 0.01 \pm 0.05$, $t = 0.26$). For allocation (Table 3), the model with the two main effects was the best model. Parameters estimated based on this model showed that both benefit and cost were predictive of allocation (benefit: $\beta = 0.64 \pm 0.10$, $t = 6.11$; cost: $\beta = 0.78 \pm 0.11$, $t = 7.43$). Both gratitude rating and allocation increased monotonically with benefit and cost (Fig. 1*B*,*C*).

To examine whether the participants' allocation was influenced by trial history, namely, trial features (cost, benefit) and the benefactor's decision from the previous trial, we performed a separate regression model for allocation (Help trials alone) with this information included the following:

$$\text{Allocation}_n = \beta_0 + \beta_1\text{Cost}_n + \beta_2\text{Benefit}_n + \beta_3\text{Cost}_{n-1} + \beta_4\text{Benefit}_{n-1} + \beta_5\text{Decision}_{n-1} + \beta_6\text{Decision}_{n-1} \times \text{Cost}_{n-1} + \beta_7\text{Decision}_{n-1} \times \text{Benefit}_{n-1} \quad (2)$$

We found that the contribution of cost and benefit on the trial$_n$ remained significant ($\beta_1 = 2.07 \pm 0.18$, $t = 11.18$; $\beta_2 = 0.55 \pm 0.14$, $t = 3.96$). Interestingly, the contribution of benefit on the last trial was also significant ($\beta_4 = 0.18 \pm 0.08$, $t = 2.26$), and it was qualified by a significant interaction with benefactor's decision ($\beta_7 = 0.27 \pm 0.10$, $t = 2.65$). These results indicate that the participants allocated more on the current trial if the benefit on the last trial was high and the benefactor chose "Help." Benefactor's sacrifice on trial$_{n-1}$ did not influence participants' allocation on trial$_n$, nor did its interaction with benefactor's decision. These findings shed light on how the impacts of different cognitive antecedents on gratitude and reciprocity persist and decay over time. Decisive conclusion in this regard is beyond the scope of the current study because this study was not designed to address the question; thus, it did not balance the distribution of cost, benefit, and benefactor's decision over time.

fMRI results

Neural representations of cost and benefit

Our first aim was to examine how the brain encodes benefit and cost when receiving help. Contrasts corresponding to the main effect of benefactor-cost and self-benefit in the Help conditions were defined based on the regressors in GLM 1 (see Materials and Methods). As we predicted, the main effect of benefactor-cost (Contrast 1) revealed activations in dorsomedial PFC, precuneus, and bilateral TPJs, the regions implicated in empathy and mentalizing (Table 4; Fig. 2*A*). The main effect of self-benefit (Contrast 2) revealed activations in a network related to value representation, including the ventromedial PFC, bilateral VS, and dorsal striatum (Table 4; Fig. 3*A*). Regional activation patterns were extracted from our hypothesized ROIs for illustrative purposes (Figs. 2*B*, 3*B*). As a comparison, the same set of contrasts defined for the NoHelp trials revealed no suprathreshold activation at the brain areas revealed by the corresponding contrasts in the Help trials (Figs. 2*A*, 3*A*). It is worth noting, however, that the null effect of the NoHelp contrasts is not sufficient to demonstrate that the neural processes observed here are specific to receiving help. To demonstrate specificity, one needs to show "separate modifiability" (e.g., Woo et al., 2014) of two constructs (e.g., Help vs NoHelp), which is beyond the scope of the current study.

Because our primary interest here is the neurocognitive processes underlying receiving help and feeling grateful, we included the NoHelp conditions only as fillers. As a result, the distribution of NoHelp trials was not balanced across different levels of cost and benefit, neither was it matched with the Help trials (Table 1). Future studies are needed to reveal the cognitive and affective response to other's withdrawal of help.

Table 4. Results of whole-brain analysis of fMRI data[a]

Regions	Hemisphere	t	Cluster size (voxels)	MNI coordinates		
				x	y	z
Constructed gratitude						
pgACC	L	3.96	60	−9	44	4
Calcarine	R	6.10	275	12	−82	13
Cuneus	L	5.48	137	−9	−73	28
Main effect benefit						
VS	R	4.40	30	12	20	−15
	L	5.87	117	−18	20	−17
Putamen	R	4.19	45	33	11	13
MCC	L	4.13	30	−12	−19	43
	R	4.01	31	12	−19	43
Temporal pole	L	4.09	27	−51	8	−14
Main effect of cost						
MPFC	R	6.25	723	3	47	7
TPJ	R	4.70	144	48	−52	31
	L	4.84	152	−48	−70	25
PCC/PreCu	R	5.00	266	12	−46	34
MTG	L	4.39	33	−57	−7	−26
Insula	R	5.81	122	33	26	−17
Calcarine	R	4.38	47	9	−79	4
Superior frontal gyrus	R	4.28	30	21	35	52

[a] MCC, Middle cingulate cortex; PCC, Posterior cingulate cortex; PreCu, Precuneus; MTG, middle temporal gyrus. Clusters survive $p < 0.001$ at voxel level and $p_{\text{FWE}} < 0.05$ at cluster level.

Neural representation of gratitude

Directly examining the representation of gratitude required us to have for each participant a trial-by-trial measure of gratitude and perform a parametric regression against brain activity elicited by observing benefactor's Help decision. Although the participants did not provide gratitude ratings during scanning, we assumed that this could be reconstructed from features of the trials (cost, benefit) and participants' gratitude ratings in the postscan gratitude recall. Thus, we derived for each participant a parameter, k, which reflects the relative contribution or weight of benefactor's cost over beneficiary's benefit in gratitude ratings (see Eq. 1). We found that k (0.72 ± 0.27; Fig. 1*D*) was significantly >0.5 ($t(30) = 4.67$, $p < 0.001$). It seems therefore that benefactor's cost may play a more prominent role in the generation of grateful feeling than the benefit one receives. We then applied these weights to each trial in the scanning task (compare Crockett et al., 2017), which served as a link between features of the trials, which were predetermined in our experimental design, and participants' trial-by-trial gratitude, which was otherwise latent.

Parametric contrast of the "constructed gratitude" revealed an activation cluster in pgACC (Fig. 4*A*), indicating that the activation in pgACC tracked gratitude throughout the task (GLM 2, see Materials and Methods). This pattern replicated previous findings that pgACC/MPFC activation is associated with gratitude (Fox et al., 2015; Kini et al., 2016; Yu et al., 2017). Moreover, ROI analysis showed that the neural responses to the constructed gratitude in pgACC were not only significantly above 0 (Fig. 4*B*, blue bar) but also positively correlated with participants' trait gratitude (Fig. 4*C*, blue dots and line). The latter finding indicates that participants with higher trait gratitude (i.e., more frequently ex-

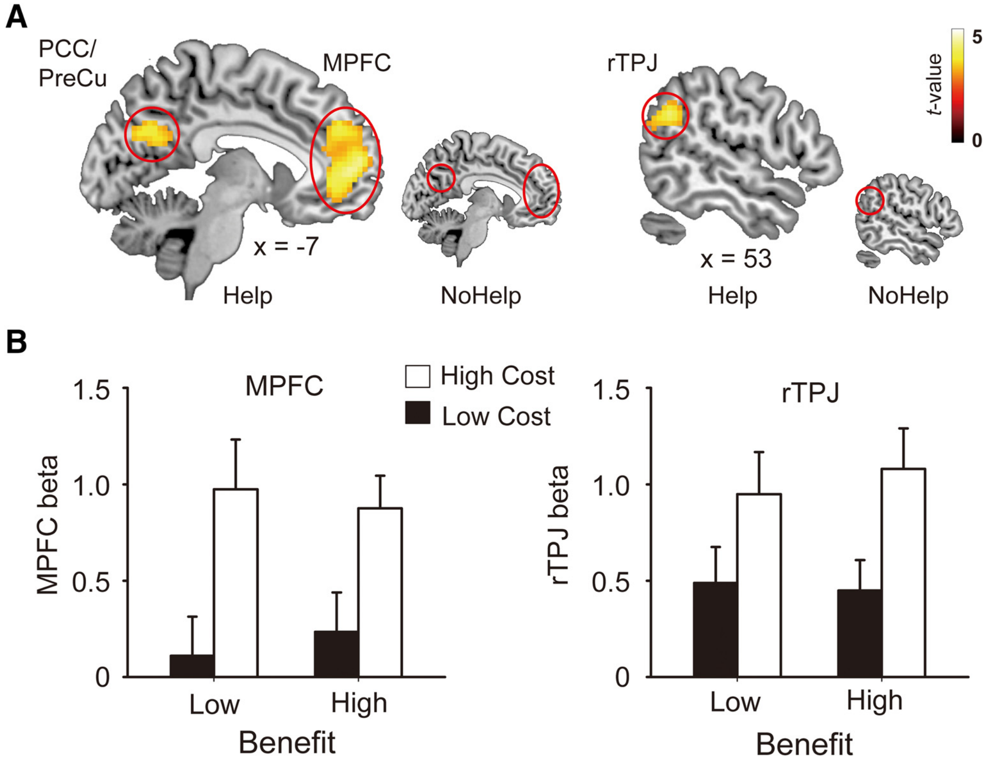

Figure 2. Encoding of benefactor-cost. ***A***, Whole-brain contrast of high versus low cost in Help conditions (larger figure). The same contrast in the NoHelp conditions was inserted for comparison (smaller figure). ***B***, Parameter estimates (β values) corresponding to the four Help conditions were extracted from MPFC and rTPJ for illustrative purposes. Error bars indicate standard error of means.

perience and express gratitude in their life) had stronger pgACC responses to constructed gratitude.

It is an interesting question concerning whether pgACC found to represent constructed gratitude is also responsible for the allocation decisions. We extracted the β estimate of the parametric contrast with trial-by-trial allocation (Contrast 4) from pgACC. Interestingly, this area was not sensitive to the amount of allocation (Fig. 4*B*, green bar). Moreover, the trait gratitude score did not correlate with pgACC's responses to allocation (Fig. 4*C*, green dots and line), indicating that the neural computation in pgACC (at least, in the current task) is specific to gratitude rather than allocation decisions.

Neural integration of cost and benefit

Once we identified the brain structures that represent benefit and cost (e.g., VS for benefit and rTPJ for cost), we could then examine the information flow between the brain areas encoding gratitude and its cognitive antecedents. We predicted that the benefit and cost information, which are represented by the neural activity in VS and rTPJ, respectively, should pass to pgACC to be integrated into an overall gratitude signal. We built and compared 33 models varying in their intrinsic connectivity, modulatory effect, and input. They were further grouped into 7 model families. Models within the same family shared the same intrinsic connectivity patterns (Fig. 5*A*). Bayesian model comparison on the family level showed that Model Family 1 had the highest exceedance probability (0.33; Fig. 5*B*). In this model family, rVS and rTPJ had unidirectional intrinsic connectivity to pgACC. This connectivity pattern is in line with our hypothesis that the brain representations of cost and benefit are fed to and integrated in the brain structure that closely track gratitude (i.e., pgACC). The connectivity strength estimated based on the Bayesian average of Model Family 1 indicated that the intrinsic connectivities from both rVS and rTPJ to the pgACC were significant (Table 5). Moreover, one of the high cost conditions, the HighCost_LowBenefit, significantly enhanced the connectivity from rTPJ to pgACC (Fig. 5*D*). In contrast, the high benefit conditions did not significantly increase the connectivity from rVS to pgACC. Overall, the DCM results partially supported our hypothesis about the neural integration of the cognitive antecedents of gratitude. The fact that the modulation of high benefit conditions was relatively weak is consistent with the behavioral finding that the benefactor's cost was weighted more by the participants, at least in the context of our task (Fig. 1*D*). A logical empirical question is what contextual factors may modulate the relative weights of benefactor's cost and one's own benefit in driving the neurocognitive processes underlying gratitude, and whether the connectivity between the benefit-related area and the gratitude-related area plays a more important role in the context that individuals weigh self-benefit more.

As can be seen from Figure 5*B*, Model Family 2 also has relatively high exceedance probability (0.28). Given that both Family 1 and Family 2 contain unidirectional connectivity from rTPJ and rVS to pgACC, the fact that their exceedance probabilities are close to each other does not seem to threaten our argument about the neural integration of cognitive antecedences in generating gratitude. The only difference between the two model families is that Family 2 contains connectivity from rVS to rTPJ. To further examine whether the connectivity between these areas plays a critical role in encoding cost and/or benefit, we performed an ROI-based psychophysiological interaction analysis focusing on rVS and rTPJ. Specifically, we defined rVS as the seed region and examined its functional connectivity with rTPJ (see Materials and Methods). The functional connectivity between rVS and rTPJ does not vary with cost ($F_{(1,30)} = 0.18, p = 0.68$) or benefit ($F_{(1,30)} =$

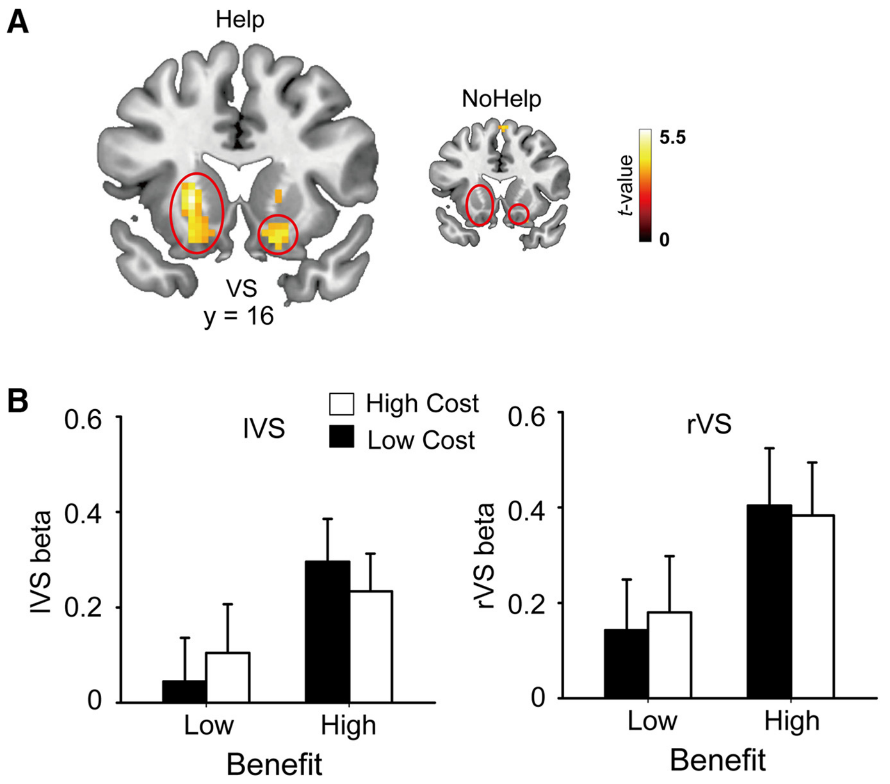

Figure 3. Encoding of self-benefit. ***A***, Whole-brain contrast of high versus low benefit in Help conditions (larger figure). The same contrast in the NoHelp conditions was inserted for comparison (smaller figure). ***B***, Parameter estimates (β values) corresponding to the four Help conditions were extracted from left VS and rVS for illustrative purposes. Error bars indicate standard error of means.

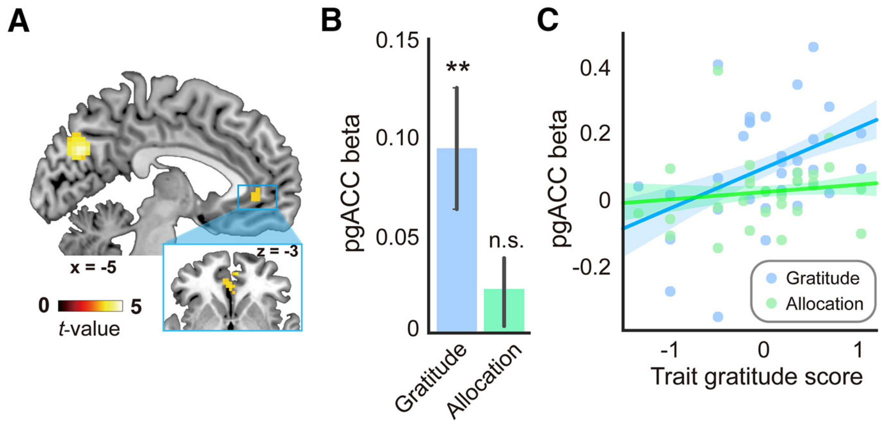

Figure 4. Representation of gratitude. ***A***, Whole-brain parametric contrast of constructed gratitude. ***B***, pgACC responses to constructed gratitude (blue) and allocation (green). ***C***, Relation between trait gratitude score and pgACC responses to constructed gratitude (blue) and allocation (green). $^{**}p < .005$. Error bars indicate standard error of means.

0.00, $p = 0.97$) (Table 6), indicating that the connectivity between these two areas does not play a critical role in generating gratitude. This finding is in line with a recent study about social-affective default network, which does not observe a connectivity between VS and TPJ in resting state BOLD signals (Amft et al., 2015).

From gratitude to reciprocity

Not surprisingly and consistent with previous findings (Yu et al., 2017), gratitude ratings correlated with allocation, both at dispositional and at situational levels. Specifically, at the dispositional or individual difference level, participants who generally gave higher ratings in the postscan gratitude recall also allocated more to the benefactor ($r = 0.47$, $p = 0.029$; Fig. 6*A*). At the situational level, the more grateful a participant felt in a given condition, as indicated by the postscan gratitude rating, the more money he/she would allocate to the benefactor in that situation ($t = 11.68$; Fig. 6*B*). However, as can be seen from Figure 6*C*, the exchange rate (regression weight) between gratitude and allocation/reciprocity varied across participants, reflecting individual differences in the prosocial behavioral motivation of gratitude. To pinpoint the neural basis of this prosocial behavioral motivation, we correlated the individual regression weights with the whole-brain contrast of Help > NoHelp based on GLM 2 (Fig. 6*D*). Two theoretical hypotheses could be proposed concerning the motivation underlying the reciprocal behavior after receiving help: it could be motivated by a self-focused concern, such as guilt-aversion and reputation; or by another regarding concern, such as goodwill for the benefactor's welfare (Batson, 1987; Fehr and Schmidt, 2006). On the one hand, receiving costly help and not giving back may generate feelings of guilt in the beneficiary, and those who are more susceptible to guilt-aversion motivation may convert gratitude to reciprocity to a larger extent than those who are less susceptible to such a motivation. Previous neural research on guilt-aversion motivation has identified anterior cingulate cortex as a critical structure for representing guilt-aversion (Chang et al., 2011). We thus performed a small-volume correction with the above contrast around the ACC coordinates reported by Chang et al. (2011). We found a significant cluster within this area ([−5, 23, 28], $t = 3.03$, $p_{\mathrm{FWE}} = 0.041$, voxel-level corrected), indicating that guilt-aversion could be a motivation of the subsequent reciprocal behavior in the current study. On the other hand, the beneficiary's reciprocity could also be motivated by a positive consideration, namely, the active concern for the benefactor's welfare. To test this possibility, we performed another small-volume correction with the above contrast based on the coordinates reported by Apps and Ramnani (2014). This study has demonstrated the role of gyral ACC in encoding value of others' rewards. We found a significant cluster within this area as well ([−3, 20, 22], $t = 3.19$, $p_{\mathrm{FWE}} = 0.030$, voxel-level corrected), indicating that other-regarding concern could be another motivation for the subsequent reciprocal behavior in the current study. Thus, based on the current data alone, we cannot choose between the two proposals. It is also possible that both motivations exist simultaneously, and their relative weights vary across individuals.

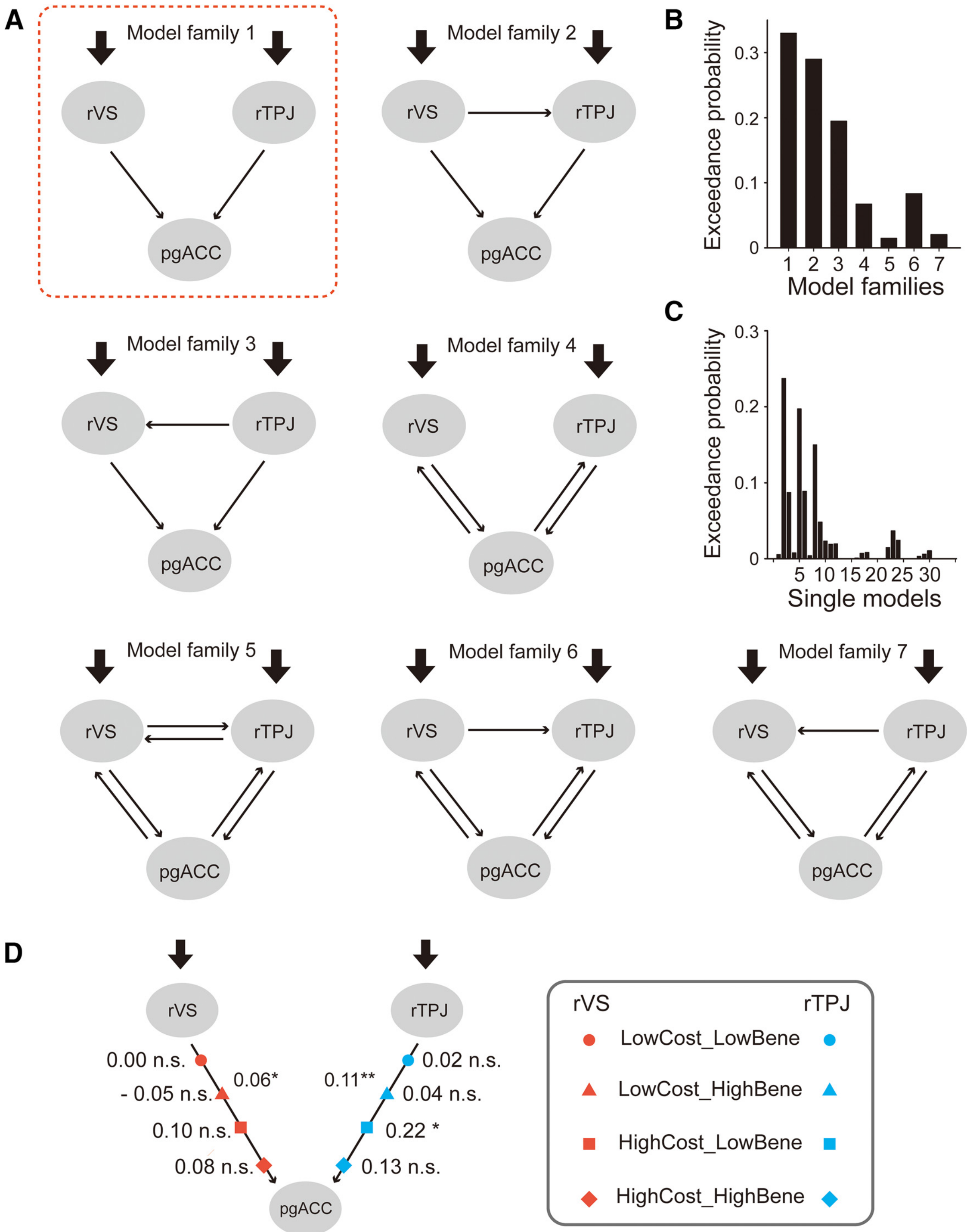

Figure 5. Results of effective connectivity (DCM) analysis. ***A***, Thirty-three individual models, grouped into 7 model families, were constructed and compared using Bayesian Model Comparison. The exceedance probability of each family (***B***) and each individual model (***C***) are shown. Model Family 1, enclosed in the red square, has the highest exceedance probability. ***D***, Strength of intrinsic and modulatory connectivities estimated based on the winning family. $^*p < 0.05$, $^{**}p < .005$.

Future research with more advanced neuroimaging methods, such as multivariate pattern classification, could be applied to distinguish these motivations.

Discussion

"Gratitude is a feeling that depends on thinking" (Visser, 2012, p. 271). Research in social psychology have provided many postulates regarding what type of "thinking," or cognitive antecedents, contribute to the feeling of gratitude (e.g., Tesser et al., 1968). However, the neural mechanism through which encoding of the cognitive antecedents gives rise to gratitude and reciprocity is far from clear. Here, by combining a social interactive paradigm with fMRI, we provide an account of how cognitive antecedents of gratitude, namely, benefactor's cost and beneficiary's benefit, are represented and integrated in the brain to give rise to gratitude and reciprocity. By analyzing effective connectivity, we showed that the representations of the cognitive antecedents, especially other's cost, were integrated in pgACC, a structure that has been

Table 5. Model parameters estimated based on Model Family 1

Parameter	Mean ± SD
Intrinsic connectivity	
VS → pgACC	0.06 ± 0.14*
rTPJ → pgACC	0.11 ± 0.18**
Modulation on VS → pgACC	
LowCost_LowBene	0.00 ± 0.06
LowCost_HighBene	−0.05 ± 0.42
HighCost_LowBene	0.10 ± 0.30
HighCost_HighBene	0.08 ± 0.67
Modulation on rTPJ → pgACC	
LowCost_LowBene	0.02 ± 0.06
LowCost_HighBene	0.04 ± 0.39
HighCost_LowBene	0.22 ± 0.59*
HighCost_HighBene	0.13 ± 0.60
Driving input to VS	
Help decision	0.04 ± 0.08*
Driving input to TPJ	
Help decision	0.06 ± 0.15*

VS, ventral striatum; TPJ, temporoparietal junction; pgACC, perigenual anterior cingulate cortex. $^*p < 0.05$; $^{**}p < 0.005$.

Table 6. Functional connectivity (PPI) between rVS and rTPJ

Condition	Connectivity (mean ± SD)
LowCost_LowBene	1.51 ± 4.10
LowCost_HighBene	1.17 ± 2.69
HighCost_LowBene	1.44 ± 3.25
HighCost_HighBene	1.83 ± 4.95

consistently implicated in representing gratitude, both by the current data (Fig. 4*A*) and a few previous neuroimaging studies on gratitude (Fox et al., 2015; Kini et al., 2016; Yu et al., 2017). These findings provide a neurocognitive account of gratitude that is in line with the appraisal approach to gratitude (Tesser et al., 1968; Weiner et al., 1979; Naito et al., 2005).

Compared with previous neuroscience research on gratitude, this study has a few novel contributions to our understanding of the neural mechanisms that give rise to gratitude and reciprocity. First, this study has adapted a theoretical model of gratitude (Tesser et al., 1968) into a computational model and, based on this model, derived a trial-by-trial index of gratitude. This allows us to pinpoint neural encoding of gratitude, as a first step to delineate its neural representation. Second, we are among the first to investigate how the neural representations of antecedents of gratitude are integrated neurally to give rise to gratitude. Finally, this study more precisely characterized the processes, at both behavioral and neural levels, through which gratitude motivates reciprocal behavioral toward the benefactor.

Appraisal theory has provided a framework to formalize our understanding of how gratitude arises from cognitive processing of relevant social information, such as benefactor's cost and beneficiary's benefit (Tesser et al., 1968). These processes may not be gratitude-specific but are rather likely to be domain-general building blocks upon which more specific and complex functions/representations could be constructed (compare Ferguson and Bargh, 2003; Lindquist and Barrett, 2012). Neural research along this line could contribute to the understanding of gratitude by first mapping out how the "building blocks" are represented neurally and then explicating how they are integrated according to certain algorithmic account (e.g., gratitude ≈ benefactor-cost + self-benefit).

This approach has been adopted in a previous neuroimaging study on gratitude. Fox et al. (2015) aimed to identify the neural correlates of benefactor's cost and beneficiary's benefit in a scenario-based gratitude imagination task. Specifically, the participants read stories depicting a helping situation, where cost (or effort) of the benefactor and benefit to the beneficiary varied. However, the authors found that neither the self-reported effort nor benefit significantly correlated with brain activity in any region.

This null effect may have resulted from some characteristics of the paradigm and data analysis. First, in a scenario-based paradigm, it is difficult to determine the onsets of the various cognitive processes leading to gratitude while the narrative is unfolding. Moreover, the self-reported effort and benefit were obtained after scanning when the participants read the scenarios again with the explicit task of evaluating effort and benefit. It is hard to know whether and to what extent such reflection captures the cognitive processes going on while the participants first read and imagined those scenarios during MRI scanning. In contrast, in the help-receiving task adopted here, cost and benefit were explicitly given to the participants at the time when the helping was happening and were independently manipulated. This allowed us to dissociate the contributions of these cognitive antecedents.

We found that the representation of self-benefit was associated with increased activations in a pain-relief and reward related network, including bilateral VS and ventromedial PFC (Bartra et al., 2013; Ruff and Fehr, 2014). Similarly, benefactor-cost was tracked by the neural activation in a mentalizing network, including bilateral TPJ and dorsomedial PFC (Van Overwalle and Baetens, 2009). It should be noted that many of these areas have not been implicated in previous studies of gratitude (Zahn et al., 2009; Fox et al., 2015; Kini et al., 2016; Yu et al., 2017), indicating that the neural substrates of gratitude and its antecedents are not identical processes. That neither the neural substrates of pain relief or mentalizing by itself gives rise to gratitude, nor the neural processing of gratitude necessarily involves brain areas associated with pain relief and mentalizing, lends support to a core characteristic of the appraisal account, namely, that the cognitive antecedents of gratitude are domain-general building blocks not inherent in gratitude but that contribute to gratitude when they are integrated and interpreted in a specific manner (Frijda, 1993; McConnell, 1993; Ellsworth and Scherer, 2003).

Consistent with previous neuroimaging studies on gratitude (Fox et al., 2015; Kini et al., 2016; Karns et al., 2017; Yu et al., 2017), we found that the pgACC was sensitive to gratitude. More specifically, our results showed that the pgACC tracked gratitude parametrically on a trial-by-trial manner (Fig. 2). An important question remains as to how the brain constructs gratitude from the component processes dedicated to its antecedents (e.g., cost, benefit). This question is analogous to a broader question in neuroeconomics, namely, how the brain constructs subjective value from various attributes associate with an item/product. (Receiving help could be seen as a socially valuable item/product.) For example, using fMRI and connectivity analysis Lim et al. (2013) found that the attributes (e.g., visual appearance, meaning) of an item are computed in specialized brain regions corresponding to the attributes, and then these specialized signals are projected to medial prefrontal/cingulate region for integration and generation of an overall value (see also Domenech et al., 2018). Here we drew on a similar analysis strategy to delineate the functional network involved in integration of antecedents. DCM analysis partially confirmed our hypothesis: although the intrinsic connectivities from the area encoding benefactor-cost and the area encoding recipient-benefit to pgACC were significant, only the benefactor-cost pathway exhibited a significant modulation by the presence of a high benefactor-cost condition (Fig. 5*D*). This finding is in

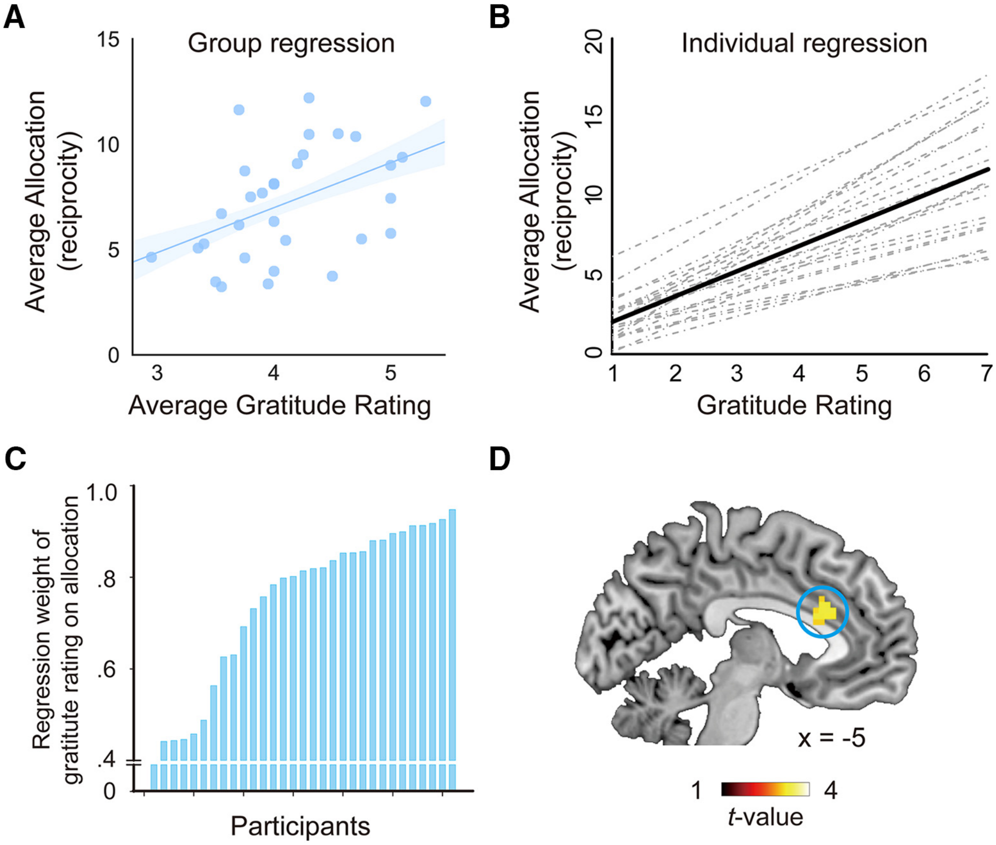

Figure 6. From gratitude to reciprocity. ***A***, Average gratitude rating in the postscan gratitude recall of an individual participant predicts average monetary allocation (i.e., reciprocity) of that participant. ***B***, Within each individual, variation in gratitude ratings predicts variation in allocation. The correlation reported here is the correlation between the postscan gratitude ratings in each of the 20 Help conditions and the average amount of allocation in the 20 Help conditions. Each dotted line indicates the regression line of a single participant, Solid line indicates the group effect. ***C***, Individual differences in the exchange rate between gratitude and reciprocity (i.e., the slopes of the dotted lines in ***B***). ***D***, Neural correlates of individual differences in the exchange rate. This map is thresholded with $p < 0.005$ for illustrative purposes.

line with the constructive nature of social emotions (Ferguson and Bargh, 2003) and sheds light on where the antecedent signals of gratitude come from and how they are integrated to give rise to the overall value of gratitude. It thus bridges the gap between the theoretical hypothesis concerning how gratitude is constructed (Tesser et al., 1968), on the one hand, and the neural evidence of how the brain represents gratitude (Fox et al., 2015; Kini et al., 2016; Karns et al., 2017; Yu et al., 2017), on the other hand.

The reciprocal motivation in gratitude has been emphasized as a core feature of this emotion, both by ancient authors and modern philosophers (e.g., Seneca, 1935; Berger, 1975; Card, 1988; McConnell, 1993; Herman, 2012; Gulliford et al., 2013). However, to our knowledge, the pathway through which such motivation emerges from the processing of gratitude has not been investigated in previous neuroscience research on gratitude. Our findings provide a preliminary attempt to answer this question. We found that those participants who were most willing to translate their grateful feelings into actual reciprocation or recompense showed higher gyral ACC response to the benefactor's help (Fig. 6D). This area has been shown to play a critical role in encoding the value of other's reward (Apps and Ramnani, 2014). In this context, the activation may reflect recipients' genuine goodwill for the benefactor's welfare and the motivation of actively seeking reward for the benefactor. Another possibility is that this activation encodes a self-focused motivation underlying reciprocity, such as avoiding guilt and indebtedness (Fisher et al., 1982; Nadler and Fisher, 1986; Watkins et al., 2006; Manela, 2016). If this is the case, then the recipients do not reciprocate because they desire to reward the benefactor, but because they are averse to the anticipated guilt of not doing so (Fehr and Schmidt, 2006). The finding that this part of ACC is sensitive to guilt-aversion (Chang et al., 2011) lends support to this interpretation. Future studies are needed to distinguish these feelings and motivations more rigorously and examine their shared and specific neural and personality profiles.

In conclusion, gratitude and other social emotions are ubiquitous and play a critical role in our social-moral life (Elfers and Hlava, 2016), the deficit of which incurs tremendous psychological, economic, and societal costs to the individuals involved, their close social relationships, and society at large (Viding et al., 2009; Blair, 2013). It is thus crucial to understand the neurocognitive basis of the function and dysfunction of social emotions. By combining an interpersonal paradigm with fMRI, we demonstrate how the brain encodes cognitive antecedents of gratitude and integrates them to give rise to gratitude. We show that the antecedent-specific signals are generated in dedicated brain structures and pass to pgACC for integration, the activity of which tracks the variation of gratitude both within (i.e., constructed gratitude) and across individuals (i.e., trait gratitude). Our approach focuses on delineating the neurocognitive pathway through which emotion is constructed and converted into behaviors, rather than charting the brain areas elicited by various categories of emotional states. We believe that, by combining emotion theories, computational modeling, interpersonal tasks, and appropriate neurobiological measures, this approach not only helps us achieve a mechanistic account of gratitude, but also serves as a role model for investigation of the neurobiological basis of other complex emotions and their significance in social-moral life.

References

Algoe SB (2012) Find, remind, and bind: the functions of gratitude in everyday relationships. Soc Personal Psychol Compass 6:455–469. CrossRef

Algoe SB, Haidt J (2009) Witnessing excellence in action: the 'other-praising' emotions of elevation, gratitude, and admiration. J Posit Psychol 4:105–127. CrossRef Medline

Algoe SB, Haidt J, Gable SL (2008) Beyond reciprocity: gratitude and relationships in everyday life. Emotion 8:425–429. CrossRef Medline

Amft M, Bzdok D, Laird AR, Fox PT, Schilbach L, Eickhoff SB (2015) Definition and characterization of an extended social-affective default network. Brain Struct Funct 220:1031–1049. CrossRef Medline

Apps MA, Ramnani N (2014) The anterior cingulate gyrus signals the net value of others' rewards. J Neurosci 34:6190–6200. CrossRef Medline

Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 68:255–278. CrossRef Medline

Bartlett MY, DeSteno D (2006) Gratitude and prosocial behavior: helping when it costs you. Psychol Sci 17:319–325. CrossRef Medline

Bartra O, McGuire JT, Kable JW (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76:412–427. CrossRef Medline

Batson CD (1987) Prosocial motivation: is it ever truly altruistic? In: Advances in experimental social psychology, Vol 20, pp 65–122. San Diego, CA: Academic.

Berger FR (1975) Gratitude. Ethics 85: 298–309. CrossRef

Blair RJ (2013) The neurobiology of psychopathic traits in youths. Nat Rev Neurosci 14:786–799. CrossRef Medline

Card C (1988) Gratitude and obligation. Am Philos Q 25:115–127.

Chang LJ, Smith A, Dufwenberg M, Sanfey AG (2011) Triangulating the neural, psychological, and economic bases of guilt aversion. Neuron 70: 560–572. CrossRef Medline

Crockett MJ, Siegel JZ, Kurth-Nelson Z, Dayan P, Dolan RJ (2017) Moral transgressions corrupt neural representations of value. Nat Neurosci 20: 879–885. CrossRef Medline

David B, Hu Y, Krüger F, Weber B (2017) Other-regarding attention focus modulates third-party altruistic choice: an fMRI study. Sci Rep 7:43024. CrossRef Medline

DeSteno D, Bartlett MY, Baumann J, Williams LA, Dickens L (2010) Gratitude as moral sentiment: emotion-guided cooperation in economic exchange. Emotion 10:289–293. CrossRef Medline

Domenech P, Redouté J, Koechlin E, Dreher JC (2018) The neuro-computational architecture of value-based selection in the human brain. Cereb Cortex 28:585–601. CrossRef Medline

Elfers J, Hlava P (2016) The spectrum of gratitude experience. New York, NY: Springer.

Ellsworth PC, Scherer KR (2003) Appraisal processes in emotion. In: Handbook of affective sciences (Davidson RJ, Scherer KR, Goldsmith HH, eds), pp 572–595. Oxford, UK: Oxford UP.

Emmons RA, McCullough ME (2003) Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 84:377–389. CrossRef Medline

Fehr E, Schmidt KM (2006) The economics of fairness, reciprocity and altruism: experimental evidence and new theories. In: Handbook of the economics of giving, altruism and reciprocity (Kolm S, Ythier JM, eds), pp 615–691. Amsterdam, the Netherlands: Elsevier.

Ferguson MJ, Bargh JA (2003) The constructive nature of evaluation. In: The psychology of evaluation: affective processes in cognition and emotion (Musch J, Klauer K, eds), pp 169–188. Mahwah, NJ: Erlbaum.

Fisher JD, Nadler A, Whitcher-Alagna S (1982) Recipient reactions to aid. Psychol Bull 91:27–54. CrossRef

Fox GR, Kaplan J, Damasio H, Damasio A (2015) Neural correlates of gratitude. Front Psychol 6:1491. CrossRef Medline

Fredrickson BL (2004) Gratitude, like other positive emotions, broadens and builds. In: The psychology of gratitude (Emmons RA, McCullough ME, eds), pp 145–166. New York, NY: Oxford UP.

Frijda NH (1993) The place of appraisal in emotion. Cogn Emot 7:357–387. CrossRef

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229. CrossRef Medline

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273–1302. CrossRef Medline

Gulliford L, Morgan B, Kristjánsson K (2013) Recent work on the concept of gratitude in philosophy and psychology. J Value Inquiry 47:285–317. CrossRef

Haidt J (2003) The moral emotions. In: Handbook of affective sciences (Davidson RJ, Scherer KR, Goldsmith HH, eds), pp 852–870. Oxford, UK: Oxford UP.

Harpham EJ (2004) Gratitude in the history of ideas. In: The psychology of gratitude (Emmons RA, McCullough ME, eds), pp 19–36. New York, NY: Oxford UP.

Herman B (2012) Being helped and being grateful: imperfect duties, the ethics of possession, and the unity of morality. J Philos 109:391–411. CrossRef

Hu J, Li Y, Yin Y, Blue PR, Yu H, Zhou X (2017) How do self-interest and other-need interact in the brain to determine altruistic behavior? Neuroimage 157:598–611. CrossRef Medline

Hu L, Zhang L, Chen R, Yu H, Li H, Mouraux A (2015) The primary somatosensory cortex and the insula contribute differently to the processing of transient and sustained nociceptive and non-nociceptive somatosensory inputs. Hum Brain Mapp 36:4346–4360. CrossRef Medline

Inui K, Tran TD, Hoshiyama M, Kakigi R (2002) Preferential stimulation of Aδ fibers by intra-epidermal needle electrode in humans. Pain 96:247–252. CrossRef Medline

Karns CM, Moore WE 3rd, Mayr U (2017) The cultivation of pure altruism via gratitude: a functional MRI study of change with gratitude practice. Front Hum Neurosci 11:599. CrossRef Medline

Kiebel SJ, Klöppel S, Weiskopf N, Friston KJ (2007) Dynamic causal modeling: a generative model of slice timing in fMRI. Neuroimage 34:1487–1496. CrossRef Medline

Kini P, Wong J, McInnis S, Gabana N, Brown JW (2016) The effects of gratitude expression on neural activity. Neuroimage 128:1–10. CrossRef Medline

Kolling N, Wittmann MK, Behrens TE, Boorman ED, Mars RB, Rushworth MF (2016) Value, search, persistence and model updating in anterior cingulate cortex. Nat Neurosci 19:1280–1285. CrossRef Medline

Kristjánsson K (2015) An Aristotelian virtue of gratitude. Topoi 34:499–511. CrossRef

Lazarus RS, Smith CA (1988) Knowledge and appraisal in the cognition–emotion relationship. Cogn Emot 2:281–300. CrossRef

Leithart PJ (2014) Gratitude: an intellectual history. Waco, TX: Baylor UP.

Lewandowsky S, Farrell S (2010) Computational modeling in cognition: principles and practice. Thousand Oaks, CA: Sage.

Lim SL, O'Doherty JP, Rangel A (2013) Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. J Neurosci 33: 8729–8741. CrossRef Medline

Lindquist KA, Barrett LF (2012) A functional architecture of the human brain: emerging insights from the science of emotion. Trends Cogn Sci 16:533–540. CrossRef Medline

Manela T (2015) Gratitude. In The Stanford Encyclopedia of Philosophy (Spring 2015 Edition), Edward N. Zalta (ed.). Available at https://plato.stanford.edu/archives/spr2015/entries/gratitude.

Manela T (2016) Negative feelings of gratitude. J Value Inquiry 50:129. CrossRef

McConnell T (1993) Gratitude. Philadelphia, PA: Temple UP.

McCullough ME, Tsang J (2004) Parent of the virtues? The prosocial contours of gratitude. In: The psychology of gratitude (Emmons RA, McCullough ME, eds), pp 123–141. New York, NY: Oxford UP.

McCullough ME, Kilpatrick SD, Emmons RA, Larson DB (2001) Is gratitude a moral affect? Psychol Bull 127:249–266. CrossRef Medline

McCullough ME, Emmons RA, Tsang JA (2002) The grateful disposition: a conceptual and empirical topography. J Pers Soc Psychol 82:112–127. CrossRef Medline

Morgan B, Gulliford L, Kristjánsson K (2017) A new approach to measuring moral virtues: the multi-component gratitude measure. Pers Individ Dif 107:179–189. CrossRef

Nadler A, Fisher JD (1986) The role of threat to self-esteem and perceived control in recipient reaction to help: theory development and empirical validation. In: Advances in experimental social psychology, Vol 19 (Berkowitz L, ed), pp 81–122. San Diego, CA: Academic.

Naito T, Wangwan J, Tani M (2005) Gratitude in university students in Japan and Thailand. J Cross Cult Psychol 36:247–263. CrossRef

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. Neuroimage 22:1157–1172. CrossRef Medline

Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP (2010) Comparing families of dynamic causal models. PLoS Comput Biol 6:e1000709. CrossRef Medline

Ruff CC, Fehr E (2014) The neurobiology of rewards and values in social decision making. Nat Rev Neurosci 15:549–562. CrossRef Medline

Seneca (1935) De Beneficiis. In: Seneca moral essays (Basore JW, translator), Vol III. Cambridge, MA: Loeb Classical Library, Harvard.

Sescousse G, Caldú X, Segura B, Dreher JC (2013) Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. Neurosci Biobehav Rev 37:681–696. CrossRef Medline

Shen B, Yin Y, Wang J, Zhou X, McClure SM, Li J (2016) High-definition tDCS alters impulsivity in a baseline-dependent manner. Neuroimage 143:343–352. CrossRef Medline

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017. CrossRef Medline

Tangney JP, Stuewig J, Mashek DJ (2007) Moral emotions and moral behavior. Annu Rev Psychol 58:345–372. CrossRef Medline

Tesser A, Gatewood R, Driver M (1968) Some determinants of gratitude. J Pers Soc Psychol 9:233–236. CrossRef Medline

Todd C (2014) Emotion and value. Philos Compass 9:702–712. CrossRef

Tsang JA (2006) Gratitude and prosocial behaviour: an experimental test of gratitude. Cogn Emot 20:138–148. CrossRef

Tsang JA, Martin SR (2018) Four experiments on the relational dynamics and prosocial consequences of gratitude. J Positive Psychol. CrossRef

Van Overwalle F, Baetens K (2009) Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage 48:564–584. CrossRef Medline

Viding E, Simmonds E, Petrides KV, Frederickson N (2009) The contribution of callous-unemotional traits and conduct problems to bullying in early adolescence. J Child Psychol Psychiatry 50:471–481. CrossRef Medline

Visser M (2012) The gift of thanks: the roots, persistence, and paradoxical meanings of a social ritual. Boston, MA: Harper Collins.

Watkins PC (2014) Gratitude and the good life. New York, NY: Springer.

Watkins PC, Scheer J, Ovnicek M, Kolts R (2006) The debt of gratitude: dissociating gratitude and indebtedness. Cogn Emot 20:217–241. CrossRef

Weiner B, Russell D, Lerman D (1979) The cognition-emotion process in achievement-related contexts. J Pers Soc Psychol 37:1211–1220. CrossRef

Woo CW, Koban L, Kross E, Lindquist MA, Banich MT, Ruzic L, Andrews-Hanna JR, Wager TD (2014) Separate neural representations for physical pain and social rejection. Nat Commun 5:5380. CrossRef Medline

Yost-Dubrow R, Dunham Y (2018) Evidence for a relationship between trait gratitude and prosocial behaviour. Cogn Emot 32:397–403. CrossRef Medline

Yu H, Hu J, Hu L, Zhou X (2014) The voice of conscience: neural bases of interpersonal guilt and compensation. Soc Cogn Affect Neurosci 9:1150–1158. CrossRef Medline

Yu H, Li J, Zhou X (2015) Neural substrates of intention-consequence integration and its impact on reactive punishment in interpersonal transgression. J Neurosci 35:4917–4925. CrossRef Medline

Yu H, Cai Q, Shen B, Gao X, Zhou X (2017) Neural substrates and social consequences of interpersonal gratitude: intention matters. Emotion 17: 589–601. CrossRef Medline

Zahn R, Moll J, Paiva M, Garrido G, Krueger F, Huey ED, Grafman J (2009) The neural basis of human social values: evidence from functional MRI. Cereb Cortex 19:276–283. CrossRef Medline