[1]Center for Brain and Cognitive Sciences, Peking University, Beijing 100871, China
[2]School of Psychological and Cognitive Sciences, Peking University,

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., 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 Shaw 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 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 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 particcost were significant (benefit: 0.30 0.07, t 4.58; cost: 0.80 0.16, t 5.02); the interaction term did not reach significance (0.01 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: 0.64 0.10, t 6.11; cost: 0.78 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:

$$\text{Allocation}_n \quad {}_0 \quad {}_1\text{Cost}_n$$

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 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

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. 6A). 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. 6B). However, as can be seen from Figure 6C, 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. 6D). 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_{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

consistently implicated in representing gratitude, both by the current data (Fig. 4A) 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 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 b.alodu.wl7(actual)-374.1(reciprou(J T* [(pense)-179.7(sh ET /GS1euet)-326.9(ale 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.

R

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

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 reu.4092 cs 1 scnu81ao16(B,e1 scnu)-25:5381–1220.