Figure 1.

when ¼ 1, the sound would be attenuated by 100 dB, which resulted in silence. The VAS ranged from 0 (not unpleasant at all) to 10 (extremely unpleasant; Fig. 1B). For each participant, we used a power function to fit the relationships between subjective unpleasant ratings and objective noise intensity, and defined 10 different levels of intensity from weakest to strongest with equal subjective unpleasantness intervals (Fig. 1C). Specifically, the noise stimuli ranged from the objective intensity which was subjectively rated as 5 (level 1) to 9.5 (level 10) with an incremental step of 0.5. Therefore, although participants' subjective unpleasantness feelings about the noise stimuli may increase nonlinearly (exponentially) with objective noise intensity, their subjective unpleasantness feelings about the 10 selected noise stimuli would increase linearly from level 1 to level 10. The 10 selected noise stimuli would be used in the following tasks. Notably, participants did not know any information about upcoming tasks before they finished the noise rating task. Thus, their subjective feelings about the noise stimuli would not be biased by other irrelevant information or motives. Noise stimuli were delivered by AKG K271 MKII headphones, and controlled by software Presentation (Neurobehavioral System Inc.).

### Noise and visual stimuli association task

Since participants would virtually not hear any noise stimuli during the interpersonal helping task, we asked them to perform a noise and visual stimuli association task to associate different levels of noise stimuli with different visual cues before the helping task. In this way, we could use the conditioned visual cues to denote different selected noise stimuli in the interpersonal helping task. As individuals' subjective perceptions of unpleasantness for the same noise stimuli could vary from person to person, we defined 10 levels of

based on posterior probability (i.e., expected frequency) of each model within the model space (Rigoux et al., 2014).

We also performed cross-validation prediction analyses and model parameter recovery to validate the winning model. To assess the predictive accuracy of the models, we divided all the trials into even-numbered and odd-numbered trials to implement cross-validation prediction analyses. Specifically, for each participant, we first used even-numbered trials to estimate model parameters, and simulated 100 sets of response data with the estimated model parameters for the odd-numbered trials. We measured the predictive accuracy by calculating the proportion of simulated decisions that correctly predicted the observed decision for each trial. Then, we repeated this process by estimating parameters with odd-numbered trials and calculating the predictive accuracy for even-numbered trials. We computed the overall predictive accuracy by averaging the two predictive accuracy values.

In the fMRI experiment, participants went through different individual-specific sets of monetary cost amount and noise unpleasantness level pairings. To confirm that the winning model could reliably estimate parameters given different sets of choice, we pe(et)16.3(sr4(,)-2-4(if)17p265.3814 0 .5(-)0(a)16.318.2(,)d)14.(mo)21.7(dees)18.3(s)-r.2(r)tel ysis(ch)23

UðM; N Þ¼ $k$ $N^{a}$ 1

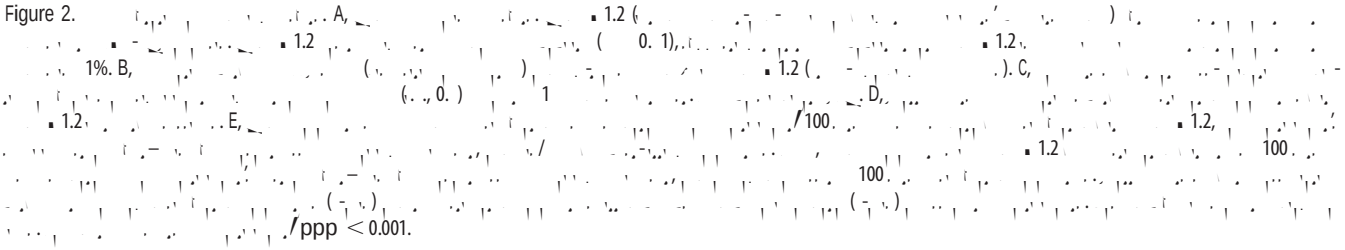

Figure 2.

$SV_{benefit}$. We observed that dACC [peak MNI: 9, 38, 19; T = 3.31, k = 2, p(SVC-FWE) = 0.013, SVC] and rIPL [peak MNI: 54, -58, 49; T = 4.15, k = 20, p(SVC-FWE) = 0.001, SVC] showed significant overlaps in neural valuation of $SV_{cost}$ and $SV_{benefit}$ (Fig. 3A, B, right panels). Contrasts of "$SV_{cost} > SV_{benefit}$" and "$SV_{benefit} > SV_{cost}$" did not reveal any region surviving a whole-brain corrected threshold or SVC.

In GLM 2, we identified regions encoding decision utility ($U_{chosen} - U_{unchosen}$), which underlies the helping decision. Whole-brain analysis revealed that activations in MPFC, left middle temporal gyrus (MTG), left angular gyrus and superior occipital gyrus (SOG) were positively associated with the decision utility, and that activations in middle cingulate cortex/supplementary motor area (MCC/SMA), left IFG, right DLPFC, and left angular gyrus were negatively associated with decision utility (Fig. 4, Table 3).

Consistent with prior studies, the finding that the activation in MPFC, especially VMPFC, was positively associated with decision utility confirmed the role of VMPFC in representing SV of decision (Levy and Glimcher, 2012; Bartra et al., 2013; Clithero and Rangel, 2014). Given that smaller decision utility increases decision difficulty, the findings that activations in cognitive control-related regions, including MCC, IFC, and DLPFC, were negatively associated with decision utility was also in line with previous studies suggesting that more extensive cognitive resources are recruited in more difficult decisions to resolve conflicts between choices with smaller utility differences (Zaki et al., 2010; Watanabe et al., 2014).

It is plausible that the regions identified in GLM 1 were also associated with decision utility or decision difficulty, but we did not observe any significant effect of decision utility or decision difficulty on dACC, rIPL as well daINS/IFG [all ps(FWE-SVC) > 0.05]. Therefore, we suggested that dACC, rIPL, and daINS/IFG identified in the previous GLM analyses were not involved in representing decision utility or difficulty.

INS, TPJ, and DLPFC) based on a 50-parcel whole-brain parcellation from Neurosynth database (http://neurovault.org/collections/2099/). We combined vaINS with daINS and mINS region in the parcellation template to form a mask covering the whole INS. Only neural responses to $SV_{benefit}$ in right vaINS/mINS [peak MNI: 45, 8, −5, $p_{(SVC\text{-}FWE)} < 0.05$] were correlated with participants' other-regarding preferences. Whole-brain analyses further confirmed stronger signal of $SV_{benefit}$ in right vaINS/mINS (peak MNI: 45, 8, −5, $T = 4.87$, cluster size = 108) in participants

Concerning the second analysis (i.e., correlation analysis), although the correlation comparison suggested that the difference between the correlation of other-regarding preferences (i.e., log-

preferences are more closely related to dispositional empathy for individuals whose positions are closer to the diagonal line in the general altruistic preference space than those who are farther away to the diagonal line. We implemented IS-RSA in the following procedure. First, we generated a parameter RDM by calculating the Euclidean distance in this general altruistic preference s(t)5.4(a(r)-2.(n).Tg6)-476.(c)-3.9(r ticipants. This parameter RDM could be taken as a measure of similarity in general altruistic preference between different individuals. Next, we creared, respectively, a ne respect to target regionsnby calculatingnthe correlation between multivoxelnpatternsof $SV_{cosnt}$ and $SV_{benefit}$ acrossnallpairsofnparticipants. Finally, we performed a Spearmarn rank correlation between these two RDMs (Fig. 7).

Results revealed significant intersubject

confirmed the functional dissociation between different subregions in INS with multivariate analyses, and indicated that neural activity patterns of $SV_{cost}$ in right DLPFC and activity patterns of $SV_{benefit}$ in bilateral vaINS/mINS were more similar between individuals who exhibited similar general altruistic preference than those who differed in general altruistic preference.

## Discussion

In this study, we provide a neurocomputational account of how benefactors weigh different attributes (i.e., one's own costs and others' benefits) to make altruistic decisions. Combining a novel task with model-based fMRI analyses, we clarify the algorithms of cost-benefit calculation underlying altruistic behaviors, the neural implementations of such calculation, and the neural basis of individual variations in altruistic preferences. Our findings implicate critical roles of a wide range of brain regions in altruistic decision-making and address how personality traits (i.e., dispositional empathy) and cognitive processes (i.e., cost-benefit calculation) interact to contributs,

welfare. The power exponent (i.e., $\alpha$) further differentiates individuals based on the magnitude of marginal utility of altruistic behaviors. Such a differentiation provides us with a new way to examine individual's altruistic preferences. One might argue that biased perceptions of noise stimuli (Shepard, 1978) and monetary magnitude (Namboodiri et al., 2014; Pardo-Vazquez et al., 2019) will render the observed integration of nonlinearly transformed attributes unreliable, and these confounding effects may not be easily addressed by our current design. Nevertheless, our findings highlight the importance of employing a nonlinear algorithm to examine cost-benefit integration of different dimensions of information underlying social decision-making.

Our model-based neuroimaging analyses further contribute to our understandings of neurocomputational basis underlying altruistic behaviors. First, our results suggest critical roles of dACC and rIPL in representing self-interest motives and other-regarding motives across different modalities. Second, univariate mediation analyses and multivariate IS-RSA provide convergent evidence for differentiating the roles of close but different subregions in INS underlying the helping behavior. Third, the IS-RSA further extend univariate analyses by revealing the role of DLPFC in altruistic preference and reconciling conflicts regarding the role of DLPFC in empathy-driven altruistic behaviors. We discuss these aspects in more detail in following sections.

First, the engagement of dACC and rIPL in evaluating benefactors' costs and recipients' benefits largely replicate our previous findings that dorsal part of MPFC encodes self-interest motives and rIPL encodes other-regarding motives when efs2.9(si)18(c)i22.6(o)]TJ 0 -1.1577 TD [(rpe)13.2(t)12.5(r)-248.4(I)13.4(oa

Batson CD, Shaw LL (1991) Evidence for altruism: toward a pluralism of prosocial motives. Psychol Inq 2:107–122.
Batson CD, Eklund JH, Chermok VL, Hoyt JL, Ortiz BG (2007) An additional antecedent of empathic concern: valuing the welfare of the person in need. J Pers Soc Psychol 93:65–74.
Bode NWF, Miller J, O'Gorman R, Codling EA (2015) Increased costs reduce reciprocal helping behaviour of humans in a virtual evacuation experiment. Sci Rep 5:15896.
Buckholtz JW, Marois R (2012) The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. Nat Neurosci 15:655–661.
Chang LJ, Yarkoni T, Khaw MW, Sanfey AG (2013) Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. Cereb Cortex 23:739–749.
Charpentier CJ, De Neve JE, Li X, Roiser JP, Sharot T (2016) Models of affective decision making: how do feelings predict choice? Psychol Sci 27:763–775.
Chong TTJ, Apps M, Giehl K, Sillence A, Grima LL, Husain M (2017) Neurocomputational mechanisms underlying subjective valuation of effort costs. PLoS Biol 15:e1002598.
Clithero JA, Rangel A (2014) Informatic parcellation of the network involved in the computation of subjective value. Soc Cogn Affect Neurosci 9:1289–1302.