◆ Human Brain Mapping 38:1281–1298 (2017) ◆

Mortality Salience Reduces the Discrimination Between In-Group and Out-Group Interactions: A Functional MRI Investigation Using Multi-Voxel Pattern Analysis

Chunliang Feng,[1,3] Bobby Azarian,[4] Yina Ma,[1] Xue Feng,[1] Lili Wang,[5] Yue-Jia Luo,[2,6]* and Frank Krueger[7]

[1]*State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, China*
[2]*Institute of Affective and Social Neuroscience Shenzhen University, China*
[3]*College of Information Science and Technology, Beijing Normal University, China*
[4]*The Krasnow Institute for Advanced Study George Mason University, Fairfax, Virginia*
[5]*School of Educational Science, Huaiyin Normal University, Huaian, China*
[6]*Shenzhen Institute of Neuroscience, China*
[7]*Department of Psychology, George Mason University, Fairfax, Virginia*

Abstract: As a fundamental concern of human beings, mortality salience impacts various human social behaviors including intergroup interactions; however, the underlying neural signature remains obscure. Here, we examined the neural signatures underlying the impact of mortality reminders on in-group bias in costly punishment combining a second-party punishment task with multivariate pattern analysis of fMRI data. After mortality salience (MS) priming or general negative affect priming, participants received offers from racial in-group and out-group proposers and decided how to punish proposers by reducing their payoffs. We revealed that MS priming attenuated in-group bias and dampened the discriminated activation patterns pertaining to group identities in regions previously implicated in costly punishment, including dorsomedial prefrontal cortex, temporo-parietal junction, anterior cingulate cortex, and dorsolateral prefrontal cortex. The group identity represented in multivariate patterns of activity of these regions predicted in-group bias for the control condition, i.e., the stronger discriminative representations of group identities in these regions; the larger was the in-group bias. Furthermore, the in-group bias was reliably

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: National Natural Science Foundation of China; Contract grant numbers: 31530031, 81471376 (to Y.J.L.), 31300869 (to L.W.); Contract grant sponsor: 973 Program; Contract grant number: 2011CB711000 (to Y.J.L.); Contract grant sponsor: Natural Science Foundation of Jiangsu Province of China; Contract grant number: BK20130415 (to L.W.); Contract grant sponsor: Chinese postdoctoral innovation talent support program (to C.F.); Contract grant sponsor: German Federal Ministry of Education and Research; Contract grant number: P-57191936 (to F.K.)

*Correspondence to: Yue-Jia Luo, Institute of Affective and Social Neuroscience, Shenzhen University, China. E-mail: luoyj@szu.edu.cn

Received for publication 22 May 2016; Revised 21 October 2016; Accepted 23 October 2016.

DOI: 10.1002/hbm.23454

Published online 10 November 2016 in Wiley Online Library (wileyonlinelibrary.com).

decoded by distributed activation patterns in the punishment-related networks but only in the control condition and not in the MS condition. These findings elucidate the neural underpinnings of the effects of mortality reminders on intergroup interaction. *Hum Brain Mapp 38:1281–1298, 2017.*

Key words: costly punishment; fMRI; in-group bias; mortality salience; multivoxel pattern analysis; pattern regression analysis

INTRODUCTION

Awareness of the inevitability of death represents one of principal existential concerns for humans [Koole et al., 2006]. The need to defend against the consciousness of mortality plays a critical role in various aspects of human behaviors [Becker, 1973; Burke et al., 2010; Pyszczynski et al., 1997]. For instance, people reported increased tension in response to the inappropriate use of cherished cultural symbols when reminded of mortality [Greenberg et al., 1995a). When confronted with existential threat, people were motivated to live up to social norms that are prescribed by cultural worldviews; showing increased compassionate and generous behaviors towards others [Hirschberger et al., 2005; Jonas et al., 2002]. Moreover, reminders of mortality could result in harsher reactions to out-group members, attitudinally dissimilar others, and social-norm violators [Greenberg et al., 1990; Rosenblatt et al., 1989]. Those effects of mortality salience (MS) were especially robust when death concerns were out of consciousness, but were attenuated or even eliminated when death concerns remained in the active memory [Greenberg et al., 1994]. Accordingly, it is unlikely that these effects could be attributed to the conscious experience of negative affect (e.g., stress) that might be induced by the mortality reminders. Indeed, the modulations of death-related thoughts on social behaviors remained robust when generally aversive thoughts or specific stressful events (e.g., an upcoming important exam) were employed as control conditions [Greenberg et al., 1994; Greenberg et al., 1995b; Greenberg et al., 1992].

Terror Management Theory attempts to explain these effects of existential threat on human motivational behaviors, positing that maintaining one's cultural worldviews offers ways to attain at least symbolic immortality; and, thereby helps to manage the potential for paralyzing terror aroused by mortality reminders [Greenberg et al., 1986; Greenberg et al., 1997]. In other words, when reminded of death, social norms (e.g., fairness) and group membership that constitute one's worldviews become more salient and people defend these values either in an aggressive or a benevolent manner [Burke et al., 2010; Gailliot et al., 2008; Jonas et al., 2008]. The impact of mortality reminders on people's responses to social norms and group membership has been extensively documented in the social psychology literature [Burke et al., 2010]; however, the underlying neural signature remains underspecified.

The current work aims to reveal how reminders of mortality modulated behavioral and neural signatures of fairness-related decision-making during racial intergroup interactions. Fairness and justice are key social norms that constitute fundamental aspects of cultural worldviews across human societies [Buckholtz and Marois, 2012; Henrich et al., 2006]. People respond to unfair treatments with negative emotions [Xiao and Houser, 2005; Yamagishi et al., 2009] and are willing to punish transgressors at a personal loss (i.e., costly punishment) [Fehr and Fischbacher, 2003; Fehr and Gächter, 2002]. Furthermore, the in-group bias of costly punishment has been identified, such that people punish in-group members less severely than out-group members for the same norm violations [Baumgartner et al., 2012; Bernhard et al., 2006; Kubota et al., 2013].

Previous neuroimaging studies have revealed several large-scale brain networks associated with costly punishment. First, the salience network, anchored in the anterior insula (AI) and anterior cingulate cortex (ACC) with extensive connectivity to subcortical structures (e.g., caudate), is generally involved in generating an aversive feeling to induce punishment [Corradi-Dell'Acqua et al., 2016; Harlé et al., 2012]. The salience network modulates the engagement of a

Abbreviations

ACC,	Anterior cingulate cortex;
AI,	Anterior insula;
ANOVA,	Analyses of variances;
ATPS,	Attitudes-toward-Punishment Scale;
BIS,	Barratt impulsiveness scale;
BNU,	Beijing Normal University;
DG,	Dictator game;
dmPFC,	Dorsomedial PFC;
EPI,	Echo-planar imaging;
IRI,	Interpersonal reactivity index;
LOSOCV,	Leave-one-subject-out cross-validation;
MKL,	Multiple Kernel Learning;
mPFC,	Medial prefrontal cortex;
MPRAGE,	Magnetization prepared rapid acquisition with gradient-echo;
MS,	Mortality salience;
MVPA,	Multivariate pattern analysis;
PANAS,	Positive and negative affective schedule;
SVM,	Support vector machine;
vlPFC,	Ventrolateral PFC;
vmPFC,	Ventromedial PFC

second network, the default-mode network, which is anchored in the medial prefrontal cortex (mPFC) [Bressler and Menon, 2010]. This network is involved in integrating emotional processes related to outcome of the norm violation and social cognitive processes related to the goal behind the norm violation [Buckholtz and Marois, 2012; Krueger and Hoffman, 2016]. The outcome-integrating portion of the network presumably works through the ventromedial PFC (vmPFC), with its inter-network connections to regions within the salience network [Krueger et al., 2009; Uddin, 2015]. The goal-integration portion of the network works through the dorsomedial PFC (dmPFC), with its intra-network connections to regions associated with mentalizing (e.g., temporo-parietal junction, TPJ), including inferring intentions or desires of others [Frith and Frith, 2003; Krueger et al., 2009]. The in-group bias in social decision-making is thought to be mediated by the default-mode network [Baumgartner et al., 2012; Baumgartner et al., 2013b; Rilling et al., 2008]. Finally, deciding a specific punishment, involves yet a third network, the central-executive network (e.g., lateral PFC), which is associated with converting the blame signal emanating from the default network into an actual punishment decision [Buckholtz and Marois, 2012].

Taken together, costly punishment engages a reflexive system in the form of the salience network (e.g., AI) which represents a motivation to punish the violators and a deliberate system including the default-mode (e.g., mPFC) and central-executive (e.g., lateral PFC) networks that integrate different sources of information (e.g., membership) to optimize decision-making. Notably, recent neuroimaging studies on MS have suggested the modulations of death reminders on those punishment-related networks [Han et al., 2010; Quirin et al., 2012; Yanagisawa et al., 2016; Yanagisawa et al., 2013]. For instance, compared to other aversive stimuli, death-related stimuli evoked decreased neural responses in bilateral AI [Han et al., 2010] and increased neural responses in amygdala, caudate, ACC and lateral PFC [Quirin et al., 2012; Yanagisawa et al., 2013]. Moreover, a recent study has identified the involvement of amygdala and ventrolateral PFC (vlPFC) and their connectivity in the processing of death-related stimuli, which were further modulated by self-esteem [Yanagisawa et al., 2016]. However, previous neuroimaging findings only indicate that punishment-related networks are involved in the processing of death-related stimuli, and it remains unclear whether the effects of mortality reminders on costly punishment to racial in-group and out-group members are modulated by similar neural circuits.

Here we combined functional MRI and a multivariate pattern analysis (MVPA) approach with a second-party punishment task (i.e., ultimatum game, UG) [Güth et al., 1982] to examine the neural signatures underlying the influence of mortality reminders on discriminated reactions to in-group and out-group members (i.e., in-group bias). First, forty Chinese female participants were randomly assigned to undergo one of the two different priming conditions: mortality-salience (MS) priming (MS condition) and the priming of generally aversive thoughts that are not associated with death (control condition) [see also Li et al., 2015; Luo et al., 2014]. Then, participants completed two delay tasks which presumably allowed thoughts of death to recede from consciousness but yet remain highly accessible in the following punishment task [Greenberg et al., 1994]. Afterwards, participants acted as second-party decision-makers (i.e., responders) receiving offers from both racial in-group and out-group members (i.e., proposers), and decided how much money they were willing to spend to punish the proposers for their offers [Strobel et al., 2011]. After the fMRI experiment, to examine the influence of MS on in-group bias in altruistic giving, participants completed a dictator game (DG) in which they decided to distribute money between themselves (i.e., dictators) and passive in-group or out-group recipients [Kahneman et al., 1986].

In light of previous findings, there were two competing hypotheses concerning the influence of MS priming on in-group bias of costly punishment [Jonas and Fritsche, 2013]. On the one hand, people might exhibit more distinct reactions to in-group and out-group members when confronted with potential death anxiety [Rosenblatt et al., 1989], thus manifested higher in-group bias of costly punishment in the MS than the control condition. Alternatively, death reminders could foster benevolent and affiliating responses to others (i.e., peaceful effects) [Greenberg et al., 1992; Hirschberger et al., 2005], especially among Eastern cultural groups who exhibited more positive attitudes to dissimilar others in MS condition than control condition [Ma-Kellams and Blascovich, 2011] and among females who were prone to showing concern and care for others in the context of existential threat [Hirschberger et al., 2005]. Therefore, it was possible that reminding Chinese female participants of their mortality would attenuate discriminations against out-group members. Furthermore, group identity encoded in multivariate activation patterns were examined using MVPA, a sensitive and increasingly popular technique for discriminating fine-grained activation patterns pertaining to cognitive states (e.g., membership) [Haxby et al., 2014; Kriegeskorte et al., 2006]. In particular, we predicted that higher in-group bias in behavioral responses would be associated with more reliable and distinct representations of group identity in the multi-voxel patterns of activation among punishment-related networks (e.g., AI, ACC, lateral PFC, mPFC, TPJ).

MATERIALS AND METHODS

Subjects

Chinese female students ($n = 40$) participated for monetary compensation in the current study. Participants were right-handed, had normal or corrected-to-normal vision,

and had no neurological or psychiatric history. There were 20 participants (mean age ± s.d.: 21.2 ± 2.7 years) in the MS condition, and another 20 participants (mean age ± s.d.: 22.1 ± 2.6 years) in the control condition. The study was approved by the Institutional Review Board at Beijing Normal University (BNU) and written informed consent was collected from all participants before the experiment.

Experimental Procedure and Task

Participants underwent three sessions for this study. First, they were invited to the lab for a *screening session* a week prior to the fMRI scanning and completed psychological surveys/questionnaires that were later used to assign them to one of the two matched groups undergoing different priming conditions. Psychological surveys/questionnaires included: Machiavelli (Mach IV test of Machiavellian) [Christie and Geis, 1970], attitudes toward punishment (Attitudes-toward-Punishment Scale, ATPS) [Viney et al., 1982], empathy (interpersonal reactivity index, IRI) [Davis and Association, 1980], impulsiveness (Barratt impulsiveness scale, BIS) [Patton and Stanford, 1995], personality style (NEO five-factor inventory, NEO-FFI) [Costa and MacCrae, 1992], attachment style (Relationship Scale Questionnaire, RSQ) [Griffin and Bartholomew, 1994], alexithymia (Toronto Alexithymia Scale-20 Items, TAS-20) [Bagby et al., 1994], self-esteem (Rosenberg Self-Esteem Scale, RSES) [Rosenberg, 1965], and narcissism (Narcissistic Personality Inventory, NPI) [Raskin and Hall, 1979].

Participants then had the chance to jointly earn monetary rewards with other players by estimating the duration of one second in a time-estimation task [Miltner et al., 1997]. They were instructed that their performance for each task trial would be paired with the performance of another player from either a racial in-group (Chinese players) or out-group (Korean players). On each trial both players could earn a joint reward (i.e., 12 monetary units [MUs]), only when they both responded correctly. Reponses were considered correct when they were within a certain critical time interval, which was adapted according to participants' performance to maintain an average accuracy of about 50% [Boksem et al., 2011]. Finally, participants were informed the money that they had earned with different partners would be used in the follow-up fMRI experiment which implemented a variant of UG.

Second, participants returned the next week for the *fMRI scanning session* to complete the UG. As responders, per round of the game Chinese participants received an offer (i.e., a split of 12 MUs) from other putative players (i.e., proposers) with whom they had previously earned the money during the time-estimation task in the screening session. Previous research has demonstrated an in-group bias among Chinese participants when Korean participants were employed as out-group members [Ma et al., 2014; Stoddard and Leibbrandt, 2014]. Thus participants were instructed that some putative proposers were Chinese (i.e., racial in-group members) and others Korean (i.e., racial out-group members) participants studying at the BNU campus. Participants were told that responders and proposers were given another 6 MUs per round of the game [Strobel et al., 2011]. In response to each proposer's allocation, they had to decide how many MUs they were willing to spend to punish proposers by reducing their payoffs: each MU spent by the responder reduced the payoff of the proposer by 3 MUs [Bernhard et al., 2006; Fehr and Fischbacher, 2004]. For instance, if the proposer split the 12 MUs into 6 MUs for herself and 6 MUs for the responder and the responder decided not to punish the proposer for this offer, then both players ended up with 12 MUs for that round. Likewise, if the proposer split the 12 MUs into 12 MUs for herself and 0 MUs for the responder and the responder decided to use all her 6 MUs to punish the proposer, then both players ended up with 0 MUs for that round. Importantly, participants were told that proposers might end up with a loss in the case of serve sanctions, which would be compensated by their show-up fee [Fehr and Fischbacher, 2004]. Importantly, terms such as "fairness," "punish," or "sanction" were not used in the instructions. Instead, participants were told that they had the chance to assign "deduction points" to the proposers [Fehr and Fischbacher, 2004].

Prior to the fMRI scanning (while already in the scanner), participants were asked to decide whether they agree or not with statements (within 7 s) that were used to prime them either for MS or general aversive thoughts (for similar procedures, see also Luo et al., 2014]. The MS priming statements were related to death (e.g., "I feel suffering that I cannot escape from death."), whereas statements in the control condition were related to negative emotions such as sadness, but not death (e.g., "I am often unhappy about trivial matters in life."). Since an aversive baseline condition controls for the effects of generally negative affect unspecific to MS [Greenberg et al., 1994], it has been often employed in the MS literature [Goldenberg et al., 2006; Li et al., 2015; Luo et al., 2014; Van den Bos and Miedema, 2000].

Following previous studies on MS [Greenberg et al., 1995b; Van den Bos and Miedema, 2000], participants rated their state feelings using the positive and negative affective schedule (PANAS) [Watson et al., 1988] after the priming procedure. Afterwards, participants were asked to perform 20 calculations in which they needed to judge whether calculations (e.g., "4573–2649") would result in an odd or even number by pressing corresponding buttons [Li et al., 2015; Luo et al., 2014]. Both procedures (PANAS and calculation task) were employed to create a delay between MS induction and decisions as well as neural responses in the UG as the core dependent measures. After the delay, thoughts of death were likely receded from consciousness but yet remained highly accessible when participants completed the UG [cf. Greenberg et al., 1994]. Note that producing one or more delays between

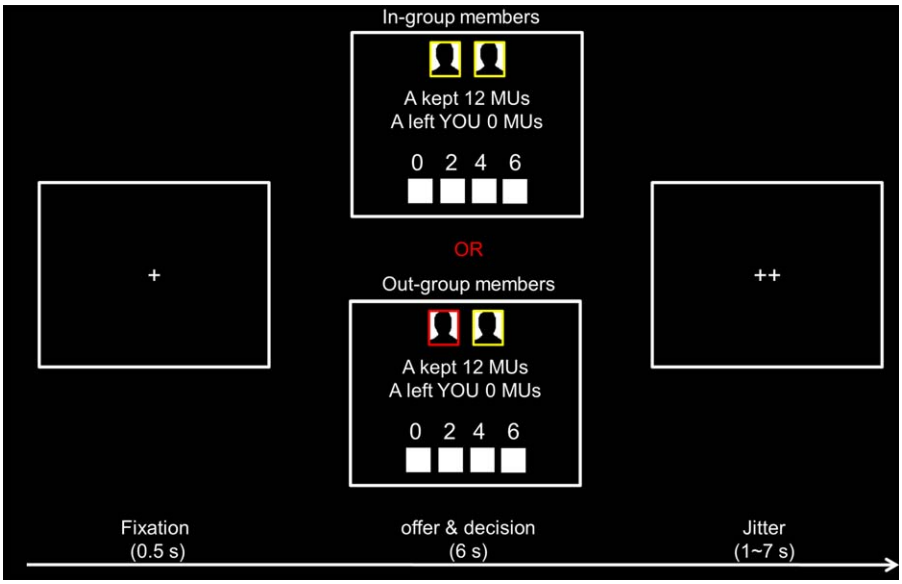

Figure 1.
Time line of a single round of the second-party punishment experiment. MUs, monetary units. [Color figure can be viewed at wileyonlinelibrary.com]

MS induction and dependent measures is a key experimental procedure, since it has been consistently demonstrated that the effects of MS on social behaviors are eliminated if death-related thoughts remain in the current focal attention [Burke et al., 2010; Greenberg et al., 1994]. Participants were randomly assigned either to the MS or control condition based on a predefined sequence unknown to the experimenter who implemented the economic games (UG, DG). The priming procedures and delay tasks were implemented by another experimenter.

Participants received instructions about the UG and played four rounds of the game to get familiar with the task. In each round of UG, a fixation cross was presented (0.5 s) and followed by an offer from either in-group or out-group members (6 s) (Fig. 1). Two images at the top of the screen, easily identified by the color (red or yellow) of the frame, represented group membership: one image (right) representing the participant herself and the other one (left) representing the proposer of the current round. To distinguish membership, Chinese participants were randomly assigned to either a red or a yellow group and told that Korean participants were in the other group. Using a response box, participants decided how many MUs (0, 2, 4, or 6 MUs) they were willing to spend to reduce the proposer's payoff. The associations between buttons and decisions were counterbalanced across participants. Finally, an optimized jitter (range of 1 to 7 s, average of 4 s) generated by an fMRI simulator software (http://www.cabi.gatech.edu/CABI/archives/resources/guide/fmrisim/) was presented, resulting in an average duration of 10.5 s for each round. Stimulus presentation and behavioral data collection were implemented using Psychtoolbox (http://psychtoolbox.org/) [Brainard, 1997; Pelli, 1997].

Participants completed two runs, each lasting 12 minutes and 36 seconds (378 scans, 2 s/scan). Furthermore, three additional scans (6 s) were automatically added to the beginning of each run to allow the MR signal to reach equilibrium but were discarded later from the data analysis. Each run consisted of 72 rounds (with an average of 10.5 s per round) from both in-group (36 rounds) and out-group members (36 rounds): nine rounds of 12:0 offers, nine rounds of 9:3 offers, nine rounds of 7:5 offers, and nine rounds of 6:6 offers. Offers of 6:6 and 7:5 were clustered as fair offers and offers of 9:3 and 12:0 as unfair offers. Clustering was based on a recent meta-analysis, which indicated that dictators, on average, generously give about 30% of their endowment to the recipients [Engel, 2011]. The within-subjects factors (2 [Offer: fair, unfair] $\times$ 2 [Membership: in-group, out-group]) consisted of 36 rounds per condition, including fair offers from in-group members, unfair offers from in-group members, fair offers from out-group members, and unfair offers from out-group members.

Finally, in the *post-scan session* participants were reminded of statements from the priming procedure and asked to report their subjective feelings of closeness to death (i.e., "How close did you feel to death after reading all the statements and making your judgments?") and fear of death (i.e., "How fearful do you feel about death after reading all the statements and making your judgments?") on a Likert scale (0 = "not at all," 10 = "very much"). In addition, participants rated their state feelings using the PANAS. Afterwards, participants played ten rounds of the DG [Kahneman et al., 1986]. They acted as dictators and decided how to allocate 10 MUs between themselves and passive recipients from either the in-

group or out-group (five rounds per group membership). Note that only data from 39 participants were collected for this part of the experiment, since one participant undergoing the control condition failed to complete the game.

To encourage real decisions from participants, it was emphasized that they would be paid according to their choices in the games in addition to fixed show-up compensation. However, each participant was paid privately with the same amount of money (¥150 RMB, about $25) at the end of the experiment [Civai et al., 2014; Corradi-Dell'Acqua et al., 2013; Grecucci et al., 2013]. Before leaving the laboratory, participants were debriefed and completed questionnaires designed to examine their beliefs about the experimental setup. No participants expressed doubts as to whether the received offers were really proposed by other players and the payoffs were dependent on their decisions in the game.

Data Acquisition

Imaging was performed on a 3 T Siemens Trio scanner equipped with a 12-channel transmit/receive gradient head coil at BNU's Imaging Center for Brain Research. A T2-weighted gradient-echo-planar imaging (EPI) sequence was used to acquire functional images: TR/TE = 2000 ms/30 ms, flip angle = 90°, number of axial slices = 33, slices thickness = 3.5 mm, gap between slices = 0.7 mm, matrix size = 64 × 64, and FOV = 224 mm × 224 mm. High-resolution anatomical images covering the entire brain were obtained by applying a magnetization prepared rapid acquisition with gradient-echo (MPRAGE) sequence: TR/TE = 2530 ms/3.39 ms, flip angle = 7°, number of slices = 144, slices thickness = 1.33 mm, matrix size = 256 × 256, FOV = 256 mm × 256 mm.

Statistical Analysis

Behavioral data

Behavioral data analyses were carried out using SPSS 21.0 (IBM, Somers, USA) with a threshold of $P < 0.05$ (two-tailed). Mixed 2 × 2 × 2 analyses of variances (ANOVA) on performances in UG (i.e., rates of punishment, response times) were applied with Offer (fair, unfair) and Membership (in-group, out-group) as within-subjects factors and Priming (MS, control) as a between-subjects factor. Furthermore, a mixed 2 × 2 ANOVA on allocations in the DG was applied with Membership (in-group, out-group) as a within-subjects factor and Priming (MS, control) as a between-subjects factor. Finally, two-sample *t* tests were performed to investigate differences in age, psychological surveys/questionnaires (e.g., empathy traits), and state feelings (e.g., subjective feelings of closeness to death, fear of death, and state positive and negative emotions) between MS and control conditions.

fMRI data: preprocessing

Neuroimaging data analyses were performed with SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Preprocessing of functional data included slice-timing correction, realignment through rigid-body registration to correct for head motion, normalization to MNI space, interpolation of voxel sizes to 2 × 2 × 2 mm^3, smoothing (8-mm full-width/half-maximum kernel) and temporal high-pass filtering (removal of low frequency drift of $T > 80$ s).

fMRI data: univariate activation analysis

At the first level, we estimated a GLM for each subject with separate experimental regressors for the factors Offer (fair, unfair) and Membership (in-group, out-group). Onsets of the experimental regressors were set to the beginning of the offer phase, and they were modeled as stick functions and convolved with canonical hemodynamic response function [Büchel et al., 1998]. To control for the differences in response times across conditions, response times were added as an additional parametric regressor for each regressor as suggested by previous works [Mumford and Poldrack, 2014; Poldrack et al., 2011]. The six movement parameters of the realignment (three translations, three rotations) were also included as nuisance regressors. The resulting GLM was corrected for temporal autocorrelations using a first-order autoregressive model. We defined the contrast of 'in-group vs. out-group' for further analyses at the second level. Our interest in this contrast was based on behavioral findings, in which the priming effects were identified only with the Membership × Priming interaction (see Results section). At the group level, the 'in-group vs. out-group' contrast images were fed into a two-sample *t*-tests (MS condition vs. control condition) to examine the Membership × Priming interaction.

Statistical interferences were conducted with a region-of-interest (ROI) analysis approach, focusing on those brain regions consistently implicated in costly punishment based on recent meta-analyses and reviews [Feng et al., 2015; Gabay et al., 2014; Rilling and Sanfey, 2011], including salience network (insula, ACC, caudate, putamen), default-mode network (mPFC, PCC, TPJ), and central-executive network (dlPFC). A small volume of interest was created to include these brain regions in both hemispheres according to the automated anatomic labeling (*aal*) atlas [Tzourio-Mazoyer et al., 2002]: *aal* Insula and adjacent Frontal_Inf_Orb ROIs were used for insula; *aal* ROIs Cingulum_Ant, Cingulum_Mid, and Cingulum_Post were used for cingulate cortex; mPFC was constructed as combination of the separate *aal* ROIs Frontal_Sup_Medial and Frontal_Med_Orb; TPJ was constructed as *aal* ROIs Parietal_Inf and SupraMarginal; dlPFC was generated with the separate *aal* ROIs Frontal_Mid and Frontal_Sup; *aal* Caudate ROIs were employed for caudate; and *aal* Putamen ROIs were used for putamen [for similar ROI constructions, see also Corradi-Dell'Acqua et al., 2013; Strobel et al., 2011]. Clusters as defined with a threshold of $t_{(38)} = 3.32$ (corresponding to $P < 0.001$, uncorrected), were significant if larger than the 95th percentile of the distribution of the largest clusters across the small volume

obtained through 5,000 replications of the same analysis on permuted datasets. The permutation test was implemented with the SnPM toolbox of SPM (http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/Nichols/software/snpm/) [see also Corradi-Dell'Acqua et al., 2011; Corradi-Dell'Acqua et al., 2014].

fMRI data: univariate correlation analysis

At the group level, we employed the contrast of interest (i.e., "in-group vs. out-group") to conduct voxel-wise correlation analyses with the purpose of identifying brain regions that correlated with the in-group bias of punishment (i.e., rates of punishment to in-groups vs. rates of punishment to out-groups) separately for the MS condition and control condition. Multiple comparisons were corrected with the same non-parametrical permutation routines used in the univariate activation analysis.

fMRI data: multi-voxel pattern analysis

To identify the effects of priming on distributed patterns of brain activation discriminating membership, we used MVPA to detect distributed neural signatures of experimental conditions (i.e., in-group vs. out-group), even if the pattern changes occurred in the absence of regional-average activation changes [Haxby et al., 2014; Haxby et al., 2001]. Compared to conventional univariate analyses, the increased sensitivity of MVPA is due to the following reasons: (i) MVPA employs information distributed across multiple voxels and (ii) MVPA is blind to the direction of activation changes across individuals [Woolgar et al., 2014].

MVPA was implemented using unnormalized and unsmoothed functional images. For each subject, we estimated a GLM that was identical to that for the univariate analysis, with the exception that each trial was separately modeled. The estimated beta images of the GLM was then fed to a support vector machine (SVM) classifier, implemented in *The Decoding Toolbox* [Hebart et al., 2014]. Using a fixed cost parameter ($c = 1$), we performed a searchlight decoding analysis as described in previous studies [Corradi-Dell'Acqua et al., 2011; Kriegeskorte et al., 2006; Wisniewski et al., 2016]. For each voxel in the individual native brain image, a sphere with a radius of four voxels was defined. For both types of membership (in-group, out-group) in each run, the parameter estimates for each of the N voxels in a given sphere was then extracted to represent an N-dimensional pattern vector. Pattern vectors in one run (i.e., training set) were employed to train the SVM to discriminate between the two types of membership, and the performance of the classifier was assessed on the other independent run (i.e., test set). We computed d′ [Green and Swets, 1966] as a measure of the sensitivity of the classifier to discriminate experimental conditions in the test set [Corradi-Dell'Acqua et al., 2011, 2016]. This procedure was repeated twice, with each run being the test set once. The average d′ across two folds was calculated and assigned to the central voxel of the sphere. The classification was conducted for each voxel, resulting in a 3-D d′ map for each subject. These d′ maps were then normalized and smoothed using the same parameters as those for the univariate analysis. At the group level, the effects of priming on the discriminations between memberships were examined with the same non-parametrical permutation routines used in the univariate activation analysis.

fMRI data: pattern regression analysis

The analysis aimed to test whether participants' in-group bias of punishment (i.e., rates of punishment to in-groups vs. rates of punishment to out-groups) could be decoded from patterns of differences in brain responses between memberships (Fernandes et al., 2017]. This complemented the univariate correlation analysis by providing two primary advantages: (i) the multivariate nature of the analysis enabled the detection of subtle and spatially distributed effects and (ii) the analysis allowed for predicting unseen participants, offering information at the individual rather than at the group level.

The pattern regression analysis was implemented in PRoNTo (http://www.mlnl.cs.ucl.ac.uk/pronto/) and included the following steps [Fernandes et al., 2017; Schrouff et al., 2013]: (1) The "in-group vs. out-group" contrast image for each participant was derived from the first-level univariate activation analysis. (2) The ROIs defined according to the *aal* atlas (see "*fMRI data: univariate activation analysis*") were used to reduce the number of anatomical regions in the brain, resulting in 26 regions in total. (3) For each ROI, a linear kernel or "similarity matrix" was computed according to the activation patterns of all voxels within the region. Therefore, a 20×20 (i.e., 20 participants in each condition) kernel matrix was generated for each region in the MS and control conditions. (4) Each kernel was normalized and mean centered to address the difference in number of voxels among brain regions. (5) The kernels from all the selected ROIs were hierarchically combined using Multiple Kernel Learning (MKL) [Schrouff et al., 2014], which aims at simultaneously learning the kernel weights in supervised learning settings. In particular, MKL determines the relative contribution of each region (kernel weights) for the final decision function, as well as the relative contribution of each voxel (voxel weights) within each region. In other words, MKL can be considered a hierarchical model, in which the models corresponding to individual brain regions are assembled to form the final brain model. Given the sparse nature of the MKL implemented in PRoNTo, only a subset of the regions would be selected in the regression analysis. (6) A nested cross-validation procedure was employed to train the model and optimize the hyperparameters of the model, with the external loop for examining the performance of the model and the internal loop for optimizing the hyperparameters. The leave-one-subject-out cross-validation (LOSOCV) was adopted for both external and internal loops [Cui et al., 2016; Fernandes et al., 2017]. (7) The performance was assessed by measuring the consistency

between the predicted and actual values, using Pearson's correlation coefficient (r). (8) The permutation test was applied to determine the significance of the model's performance. That is, the differences in rates of punishment to in-group members compared to out-group members (i.e., in-group bias) was permutated across the sample ($n = 20$ in each condition) 1,000 times without replacement within the MS or control condition, and the entire regression procedure was reapplied each time. The P value for the r was calculated by dividing the number of permutations that showed a higher value than the actual value for the real sample by the total number of permutation (i.e., 1,000).

RESULTS

Behavioral Results

Priming manipulation check

Participants felt closer to death ($t_{38} = 2.33$, $P < 0.05$, Cohen's $d = 0.74$) and reported heightened fear of death ($t_{38} = 2.29$, $P < 0.05$, Cohen's $d = 0.72$) in the MS condition than in the control condition (Supporting Information Table S1). In line with previous studies [Greenberg et al., 1994; Van den Bos and Miedema, 2000], death reminders did not modulate participants' general affective feelings after priming (negative affect [MS vs. control]: $t_{38} = 0.49$, $P > 0.05$, Cohen's $d = 0.16$; positive affect [MS vs. control]: $t_{38} = -1.00$, $P > 0.05$, Cohen's $d = -0.32$) and after fMRI scanning (negative affect [MS vs. control]: $t_{38} = -1.41$, $P > 0.05$, Cohen's $d = 0.32$; positive affect [MS vs. control]: $t_{38} = -1.44$, $P > 0.05$, Cohen's $d = -0.46$) (Supporting Information Table S1).

Decisions

The ANOVA on rates of punishment in UG showed significant main effects of Offer ($F_{1,\ 38} = 68.38$, $P < 0.0005$) and Membership ($F_{1,\ 38} = 17.05$, $P < 0.0005$), revealing that participants punished more frequently in response to unfair offers than to fair offers and to offers from out-group members than from in-group members (Supporting Information Fig. S1). In addition, a significant interaction effect of Membership × Priming was observed ($F_{1,\ 38} = 5.41$, $P < 0.05$) (Fig. 2a), such that out-group members were punished more frequently than in-group members independently for type of offer but only after negative-affect priming ($P < 0.0005$) and not after mortality-salience priming ($P > 0.05$), indicating that MS priming reduced in-group bias in punishment. Further, out-group members were more frequently punished after negative-affect compared to mortality-salience ($P < 0.05$) priming. Other effects were not significant: Priming ($F_{1,\ 38} = 2.14$, $P > 0.05$), Offer × Priming ($F_{1,\ 38} = 0.001$, $P > 0.05$), Offer × Membership ($F_{1,\ 38} = 1.00$, $P > 0.05$), and Offer × Membership × Priming ($F_{1,\ 38} = 3.30$, $P > 0.05$). Note that the statistical analysis for response time in UG is described in the supplementary material (Supporting Information Fig. S2).

Finally, to further examine differences in the in-group bias between conditions, we assessed the correlations between punishment rates of in-group and out-group members separately for the MS condition and control condition (Fig. 2b). A significant positive correlation was identified for the MS condition ($n = 20$, $r = 0.94$, $P < 0.0005$) but not for the control condition ($n = 20$, $r = -0.15$, $P > 0.05$), echoing the notion that participants treated in-group and out-group members similarly in the MS condition but not in the controls condition.

The ANOVA on amounts of DG allocation after the fMRI scanning demonstrated a main effect of Membership ($F_{1,\ 37} = 7.90$, $P < 0.01$, $\eta^2_p = 0.18$), indicating that participants showed greater generosity for in-group than for out-group members. Moreover, the main effect of Priming was significant ($F_{1,\ 37} = 12.25$, $P < 0.005$, $\eta^2_p = 0.25$), revealing that people were more generous in the MS condition than the control condition (Fig. 2c). The interaction effect of Membership × Priming was not significant ($F_{1,\ 37} = 0.46$, $P > 0.05$, $\eta^2_p = 0.01$). In addition, a significant positive correlation between amounts of giving to in-group and out-group members was only identified for the MS condition ($n = 20$, $r = 0.73$, $P < 0.0005$) but not for the control condition ($n = 19$, $r = -0.08$, $P > 0.05$), indicating that participants treated in-group and out-group members similarly in the MS condition but not in the controls condition (Fig. 2d).

Control measures

No significant differences were observed in age and psychological surveys/questionnaires between priming groups (all $P > 0.05$) (Supporting Information Table S2).

Neuroimaging Results

Univariate activation analysis

The interaction between Priming and Membership (i.e., MS [out-group – in-group] vs. control [out-group – in-group]) did not reveal any significant changes in brain activations.

Univariate correlation analysis

Only for the control condition but not the MS condition, differences in rates of punishment showed significant positive correlations with the neural responses to the contrast of interest (in-group vs. out-group) in the following brain regions: bilateral TPJ, AI, rostral ACC, and middle frontal gyrus (Table I and Fig. 3). That is, the stronger the activations in these regions in response to out-group compared to in-group members, the more frequently participants punished out-group than in-group members.

Multi-voxel pattern analysis

The analysis aimed to identify regions in which spatial patterns of brain activity differentiating memberships (i.e.,

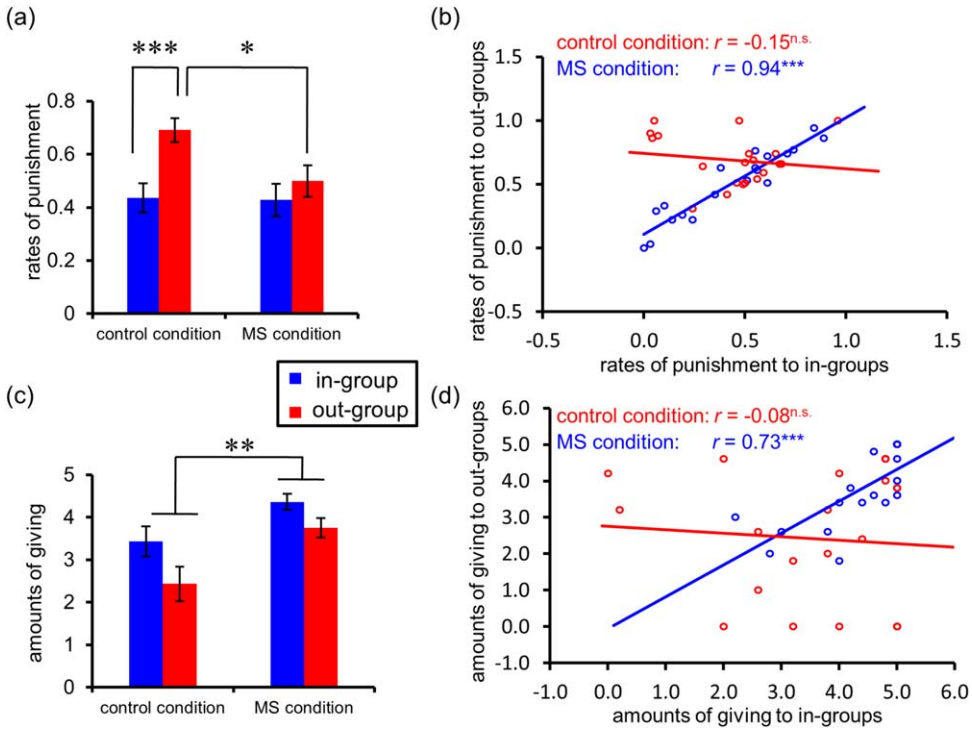

Figure 2.

Behavioral data. **(a)** Average rates of punishment as a function of membership and priming. **(b)** Correlations between rates of punishment to in-group and out-group members as a function of priming. **(c)** Average amounts of allocations in the Dictator Game as a function of membership and priming. **(d)** Correlations between amounts of altruistic giving to in-group and out-group members as a function of priming. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0005$. Error bars indicate standard error. MS, mortality salience; n.s., not significant. [Color figure can be viewed at wileyonlinelibrary.com]

in-group vs. out-group) were more robust and distinct in the control condition than the MS condition based on the patterns observed at the behavioral level. The MVPA revealed the following brain regions (q(FDR) < 0.01, corrected at the voxel level in conjunction with cluster size > 20 voxels): bilateral TPJ, dmPFC, caudal ACC, right dlPFC, and left lateral orbitofrontal cortex (lOFC) (Table II and Fig. 4). Among these regions, discriminated activation patterns pertaining to membership were more robust and distinct in the control condition than in the MS condition.

TABLE I. Brain regions showing positive correlations between in-group bias of costly punishment and neural response to in-groups compared to out-groups in the control condition

	MNI coordination of local maxima (mm)			Local maxima	
Brain regions	x	y	z	T	Cluster size (voxel)
LTemporo-parietal junction	−30	−54	50	7.26	1059
RTemporo-parietal junction	30	−52	46	7.59	406
Anterior cingulate cortex	−12	8	52	5.98	205
RMiddle frontal gyrus	24	4	54	4.34	197
LMiddle frontal gyrus	−24	6	56	4.44	189
LAnterior insula	−28	30	6	5.91	173
RAnterior insula	30	26	6	5.44	128

$P < 0.05$ permutation-based correction for multiple comparisons at the cluster level with a cluster-defining threshold of uncorrected $P < 0.001$, voxel size = 2 × 2 × 2 mm. R, right; L, left.

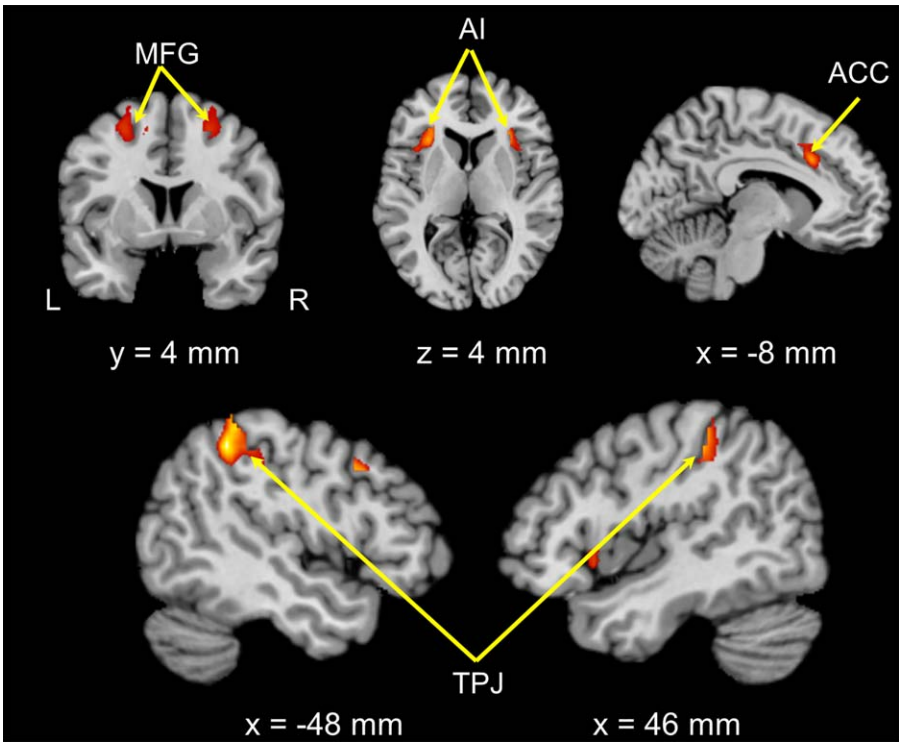

Figure 3.

Brain regions showing positive correlations between in-group bias of costly punishment and neural response to in-groups compared to out-groups in the control condition. Images were thresholded at $P < 0.05$ corrected for multiple comparisons at the cluster level. L, left; R, right; MFG, middle frontal gyrus; AI, anterior insula; ACC, anterior cingulate cortex; TPJ, temporo-parietal junction. [Color figure can be viewed at wileyonlinelibrary.com]

To explore whether the group identity represented in these regions predicted behavioral discriminations against out-group members (i.e., in-group bias), we performed exploratory correlation analyses between d' values extracted from these regions and participants' in-group bias of costly punishment. For the control condition, significant positive correlations were found between behavioral in-group bias and membership-discrimination information in all of these regions, whereas no significant correlations were identified for the MS condition: dmPFC (control: $n = 20$, $r = 0.61$, $P < 0.01$; MS: $n = 20$, $r = -0.10$, $P > 0.05$) (Fig. 5a), caudal ACC (control: $n = 20$, $r = 0.59$, $P < 0.01$; MS: $n = 20$, $r = 0.24$, $P > 0.05$) (Fig. 5b), left lOFC (control: $n = 20$, $r = 0.56$, $P < 0.05$; MS: $n = 20$, $r = 0.27$, $P > 0.05$) (Fig. 5c), right dlPFC (control: $n = 20$, $r = 0.64$, $P < 0.005$; MS: $n = 20$, $r = -0.12$, $P > 0.05$) (Fig. 5d), left TPJ (control: $n = 20$, $r = 0.77$, $P < 0.0005$; MS: $n = 20$, $r = 0.32$, $P > 0.05$) (Fig. 5e), and right TPJ (control: $n = 20$, $r = 0.70$, $P < 0.005$; MS: $n = 20$, $r = 0.33$, $P > 0.05$) (Fig. 5f).

Pattern regression analysis

The analysis aimed to complement the univariate correlation analysis by examining whether patterns of differences in activity between out-groups and in-groups in the punishment-related networks could be used to decode individual in-group bias of costly punishment. Based on patterns of differences in neural response between memberships, the correlation coefficient (r) between actual and predicted in-group bias of costly punishment was significant for the model in the control condition ($r = 0.78$, $P = 0.001$) (Figs. 6a and 7a). In the MS condition, however, no significant results were identified ($r = -0.03$, $P > 0.05$) (Fig. 6b). In other words, the MKL model could decode behavioral in-group bias of costly punishment from activation patterns of differences in neural response between memberships in the control condition but not in the MS condition.

In the control condition, a total of 13 regions had a non-null contribution to the final decision model in the MKL, including bilateral TPJ, insula, vmPFC, PCC, and dlPFC among others (Table III). Voxel weights in the whole small volume and the six top regions ranked by the region weights for the MKL model are illustrated in Figure 7b,c, respectively.

TABLE II. Brain regions exhibiting higher discriminations of membership information in the control condition than in the mortality-salience condition

Brain regions	MNI coordination of local maxima (mm)			Local maxima	Cluster size
	x	*y*	*z*	*T*	(voxel)
Dorsomedial prefrontal cortex	4	46	32	4.66	638
Caudal anterior cingulate cortex	2	−10	30	4.16	517
LLateral orbitofrontal cortex	−50	34	−4	4.23	64
RTemporo-parietal junction	34	−46	46	4.00	61
RDorsolateral prefrontal cortex	18	22	56	4.13	48
LTemporo-parietal junction	−54	−28	46	3.82	22

Voxel-wise $q(\text{FDR}) < 0.01$ in conjunction with cluster size >= 20 voxels (voxel size = 2 × 2 × 2 mm), permutation-based correction for multiple comparisons. R, right; L, left.

DISCUSSION

Our study employed multi-voxel decoding of functional MRI to examine the neural signatures underlying the effects of mortality reminders on the in-group bias in costly punishment. We showed that participants punished out-group members more frequently than in-group members in the control condition, which were consistent with previous observations on the in-group bias in costly punishment [Baumgartner et al., 2011; Baumgartner et al., 2013a). Those exaggerated aggressive reactions to out-group members might reflect the motivations of revenge or the attempts to establish dominance by expressing aggression, given that participants in the control condition punished out-group members regardless of

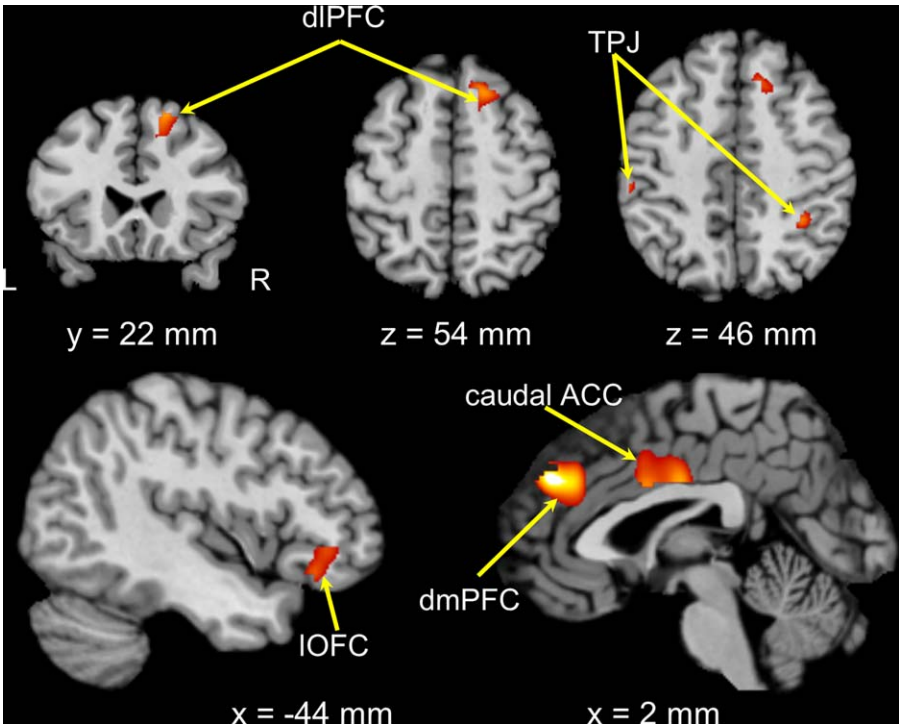

Figure 4.

Brain regions exhibiting higher discriminations of membership information in the control condition than in the mortality-salience condition. Images were thresholded at $P < 0.01$ FDR-corrected for multiple comparisons at the voxel level. L, left; R, right; dlPFC, dorsolateral prefrontal cortex; TPJ, temporo-parietal junction; lOFC, lateral orbital frontal cortex; dmPFC, dorsomedial prefrontal cortex; ACC, anterior cingulate cortex. [Color figure can be viewed at wileyonlinelibrary.com]

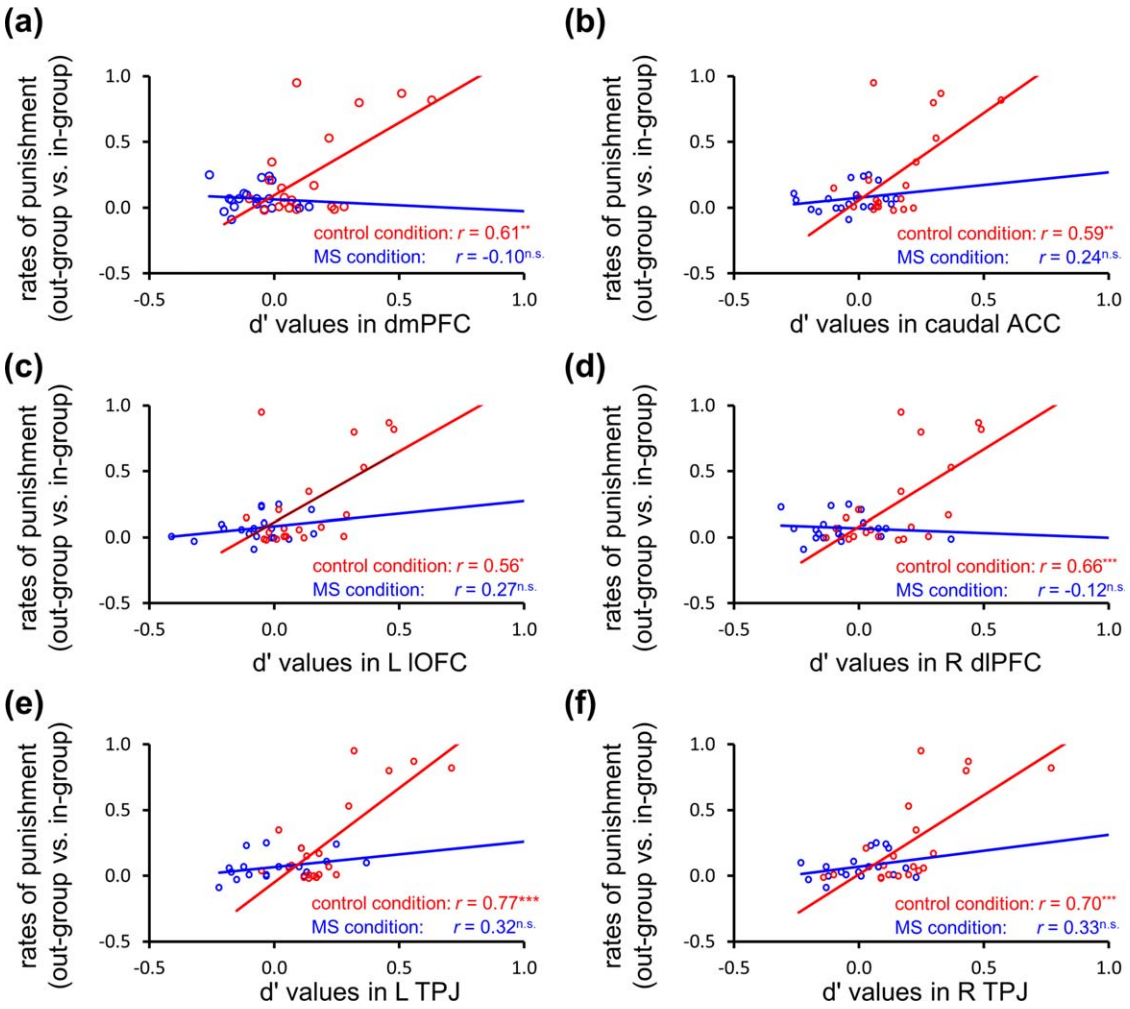

Figure 5.

Correlations between in-group bias of costly punishment and membership-discrimination information represented in punishment-related regions. In the control condition but not in the MS condition, in-group bias of punishment (out-group vs. in-group) showed positive correlations with discriminated representations of membership in dmPFC **(a)**, caudal ACC **(b)**, left lOFC **(c)**, right dlPFC **(d),** left TPJ **(e)**, and right TPJ **(f)**. L, left; R, right; dmPFC, dorsomedial prefrontal cortex; ACC, anterior cingulate cortex; lOFC, lateral orbital frontal cortex; dlPFC, dorsolateral prefrontal cortex; TPJ, temporo-parietal junction; n.s., not significant. $^{*}P < 0.05$, $^{**}P < 0.005$. [Color figure can be viewed at wileyonlinelibrary.com]

fairness [see also Gächter and Herrmann, 2009; Sylwester et al., 2013]. Importantly, the in-group bias of costly punishment was diminished by MS priming and the awareness of death instigated higher generosity towards both in-group and out-group members. Furthermore, participants' behavioral reactions to out-group and in-group members were barely correlated with each other in the control condition, further implicating the decimations against to out-group members. In contrast, MS priming resulted in highly coordinated reactions between memberships, such that participants' reactions to in-group members were positively correlated with those to out-group members. Taken together, our behavioral findings provided converging evidence on the MS-induced decreases in discriminations against out-group members.

Underlying the behavioral effects were more robust and distinct neural representations of membership in the bilateral TPJ, dmPFC, ACC, right dlPFC, and left lOFC for the control condition compared to the MS condition. Furthermore, the membership-discrimination information in these regions predicted in-group bias of costly punishment in the control condition, whereas mortality reminders dampened these brain-behavior associations. In addition, activations of AI, ACC, dlPFC, and TPJ to in-groups compared to out-groups served as neural predictors of the in-group bias in the control condition, i.e., the stronger the

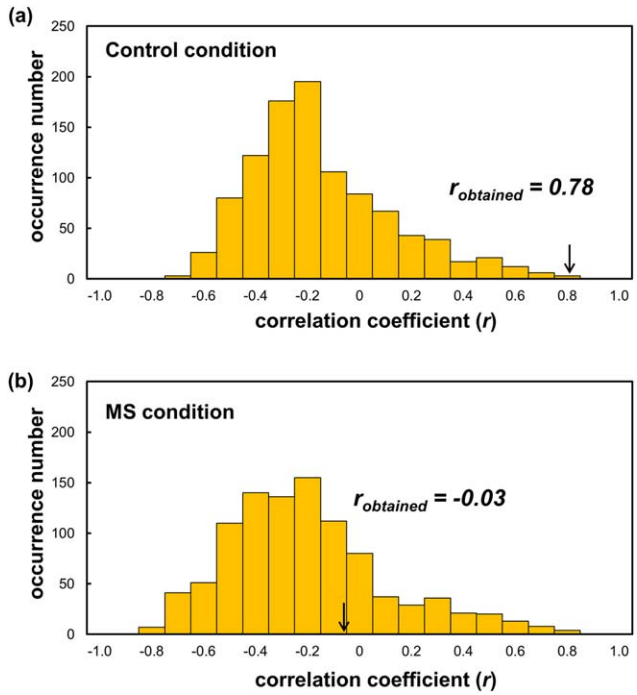

Figure 6.

Histograms of the permutation distribution of the correlation coefficient observed in pattern regression analysis. **(a)** Permutation distribution of the correlation coefficient in the control condition. **(b)** Permutation distribution of the correlation coefficient in the MS condition. The values obtained using the real data are indicated by the arrows. MS, mortality salience. [Color figure can be viewed at wileyonlinelibrary.com]

activation of these regions in response to out-group members, the higher in-group bias. Further, it was revealed that patterns of differences in neural responses between memberships were sufficient to decode the behavioral in-group bias at the individual level in the control condition. Again, those brain-behavior associations regarding in-group bias were diminished by MS priming. We interpret these findings as that reminding people of their mortality attenuates discriminations between in-group and out-group members.

Importantly, it is unlikely that those behavioral and neural effects are attributed to the participants' aversive feelings such as stress responses, which also instigate pro-social behaviors [von Dawans et al., 2012]. First, MS effects are stronger when death-related thoughts are out of consciousness compared to when these thoughts remain in current active memory [Burke et al., 2010; Greenberg et al., 1994], underscoring the unconscious nature of death concerns on social behaviors. Second, the MS manipulations have consistently failed to evoke enhanced general negative affect or physiological arousal compared to the control condition [Rosenblatt et al., 1989; Van den Bos and Miedema, 2000], whereas these measures are often employed as indicators of stress responses [van Marle et al., 2009; von Dawans et al., 2012]. Notably, one study did identify the increased negative affect induced by MS, but only after delay tasks and not immediately after MS priming [Routledge et al., 2010]. However, the current findings cannot be attributed to the negative affect raised after the delay tasks, since there were no significant differences in negative affect collected in the *post-scan session* between MS and control conditions. Third, MS effects remain robust when stressful life events are employed as control topics [Greenberg et al., 1994, 1995b; Rosenblatt et al., 1989]. Together, it is more likely the accessibility to death concerns than potential affective impact that induces the MS effects on social behaviors. In line with this idea, our manipulation check indicated that MS priming compared to the control condition induced stronger fear of death and feelings of closeness to death, implicating an increased accessibility to death-related thoughts in the MS condition.

From the perspective of Terror Management Theory, social values such as fairness and membership protect people from existential threat [Greenberg et al., 1990, 1997]. Therefore, reminding people of their own mortality increases adherence and defense of these social values. On the one hand, this could be manifested as harsh and hostile reactions to moral transgressors [Florian et al., 2001; Greenberg et al., 1990; Schindler et al., 2012] and out-group members [Harmon-Jones et al., 1996; Rosenblatt et al., 1989]. On the other hand, existential threat induced by death reminders can also be buffered by benevolent and tolerant responses to others (i.e., peaceful effects), including out-group members or norm violators in certain contexts [Greenberg et al., 1992; Hirschberger et al., 2005; Jonas and Fritsche, 2013; Lieberman et al., 2001; Niesta et al., 2008; Schimel et al., 2006; Wisman and Koole, 2003]. Indeed, mortality reminders does not necessarily lead to intergroup conflict and intolerance, but also fosters forgiveness toward an antagonistic out-group member [Schimel et al., 2006]. Similarly, an awareness of death reduces prejudice toward out-group members when people are reminded of the importance of pro-social norms [Gailliot et al., 2008].

This idea is further reinforced by our findings that mortality reminders attenuated the in-group bias during intergroup decision-makings by (i) reducing exaggerated harsh reactions to out-group members, (ii) increasing coordinated behavioral reactions between memberships, and (iii) augmenting altruistic giving to both in-group and out-group members. Taken together, our findings in line with the Terror Management Theory complement previous observations on the benevolent behaviors induced by mortality reminders. There are two possible accounts for current observations that MS priming evoked peaceful rather than aggressive responses to out-group members. First, an Eastern cultural sample was recruited in the current study. It has been previously demonstrated that MS priming induced benevolent rather than harsh responses to

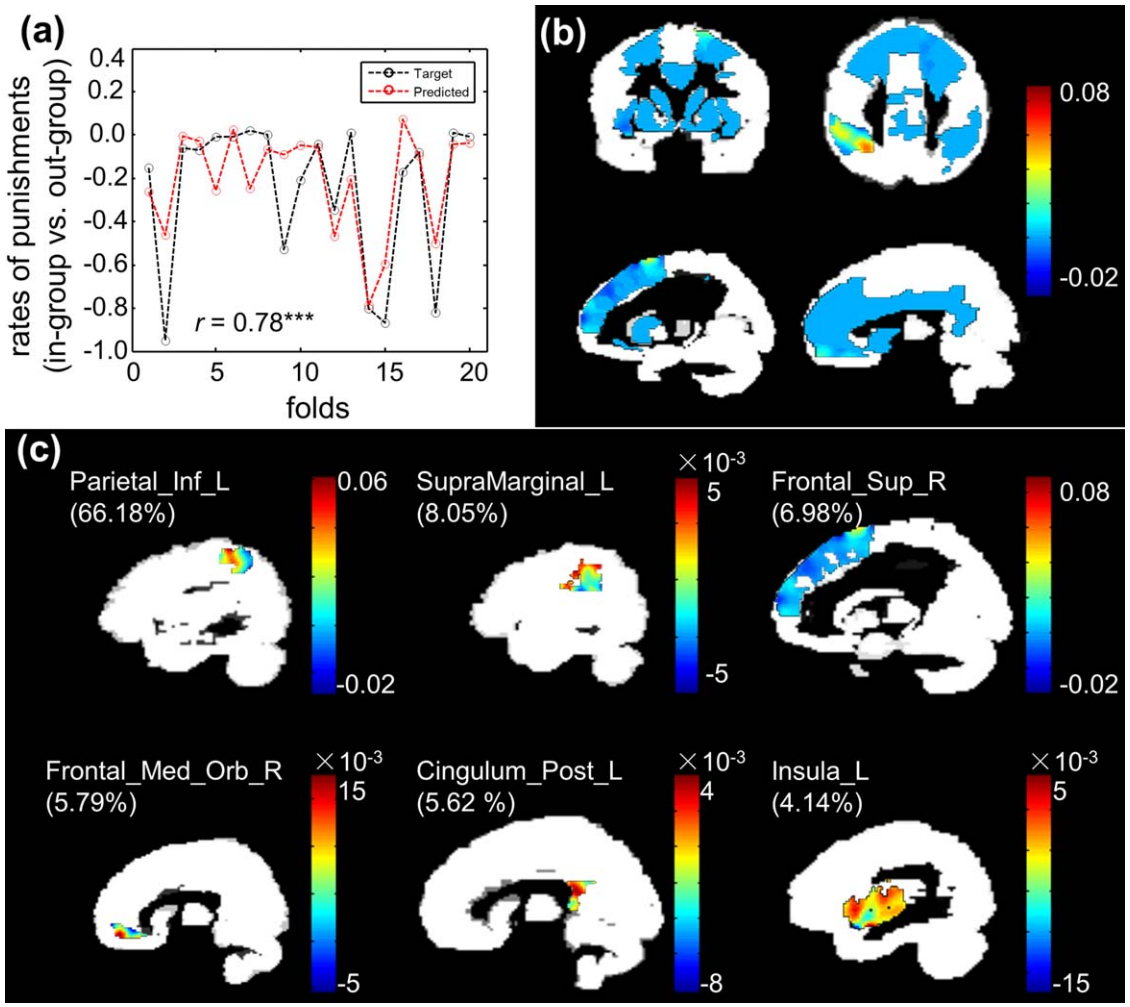

Figure 7.

MKL findings revealed in the pattern regression analysis in the control condition. **(a)** Line plot showing consistency between actual and predicted in-group bias of punishment. **(b)** Weight maps in the whole small volume. **(c)** Top six regions ranked by the MKL model used for predicting in-group bias of costly punishment. [Color figure can be viewed at wileyonlinelibrary.com]

dissimilar others among participants from Eastern cultures [Ma-Kellams and Blascovich, 2011]. This is presumably due to the reason that high values were put to social ties (i.e., interdependent self) in Eastern cultures; therefore, it is likely that Eastern cultural groups defend themselves from existential threat by affirming rather than derogating others [see also Ma-Kellams and Blascovich, 2011]. Second, only female participants were recruited. Males and females defend themselves against death differently, such that males are motived to display strength and independence, whereas females focus on showing concern and care for others [Hirschberger et al., 2005], which might be another explanation for why we found decreased discriminations between in-group and out-group members under MS priming. In either case, our findings indicate that socially constructive behaviors, such as showing tolerance and care towards others, might be an important way to defend against existential threat [Jonas and Fritsche, 2013; Niesta et al., 2008].

Our neuroimaging findings provided additional evidence for the assumption that people show less distinct reactions to in-group and out-group members under MS priming. Specifically, the strength of representations in the dmPFC, bilateral TPJ, caudal ACC, and right dlPFC differentiating between memberships were significantly attenuated in the MS condition compared to the control condition. The dmPFC and TPJ as core regions of implementing theory-of-mind reasoning have been consistently associated with in-group favoritism and out-group derogation during the intergroup interactions [Baumgartner et al., 2012, 2013b; Rilling et al., 2008]. For instance, Baumgartner et al. (2013b) has demonstrated a causal role of TPJ in in-group bias of costly punishment by

TABLE III. Brain regions ranked according to their importance to the decision function for the model trained to predict in-group bias of costly punishment from patterns of differences in neural response to out-groups compared to in-groups

Rank	Region label	Region weight (%)	Region size (voxel)
1	Parietal_Inf_L	66.18	1075
2	SupraMarginal_L	8.05	1614
3	Frontal_Sup_R	6.98	3046
4	Frontal_Med_Orb_R	5.79	436
5	Cingulum_Post_L	5.62	335
6	Insula_L	4.14	1769
7	Frontal_Inf_Orb_L	2.02	1072
8	Parietal_Inf_R	0.88	1254
9	Caudate_R	0.14	943
10	Cingulum_Post_R	0.10	463
11	Putamen_R	0.04	1009
12	Insula_R	0.03	1856
13	Frontal_Inf_Orb_R	0.03	1296

showing that disruption of the TPJ led to decreased in-group bias. The ACC as a region of the salience network has been implicated in encoding aversive feelings and/or norm violations during social interactions and are predictive of harsh reactions to norm violations [Chang and Sanfey, 2013; Xiang et al., 2013]. For instance, a recent neuroimaging study employing MVPA identified domain-general affective processing in the ACC, pointing to a common coding of affective unpleasantness, arousal, or the salience of the experience [Corradi-Dell'Acqua et al., 2016]. Finally, the dlPFC as a region of the central-executive network is associated with integrating context-dependent information (e.g., membership) from the default and salience networks and converting them into an actual punishment decision [Feng et al., 2015; Krueger and Hoffman, 2016]. Taken together, these regions provide a potential neural substrate for the human tendency to discriminate against out-group members during social interactions, a notion supported by our observations on positive correlations between membership-discrimination information represented in these regions and behavioral in-group bias of costly punishment in the control condition.

Finally, our correlation and pattern regression findings revealed neural predictors of in-group bias of costly punishment. We first demonstrated that augmented neural responses of AI, ACC, dlPFC, and TPJ to out-groups compared to in-groups were associated with a higher in-group bias in the control condition. Furthermore, these regions showed discriminative spatial patterns of neural responses that were sufficient to decode individual in-group bias. Notably, both of those brain-behavior associations regarding in-group bias were diminished by MS priming. Considering the critical roles of these regions in the costly punishment [Feng et al., 2015; Gabay et al., 2014], these findings provided additional evidence for the robust discriminations against out-group members at both neural and behavioral levels in the control condition, both of which were diminished by death reminders.

Several limitations should be noted as they relate to this study. First, only female participants were recruited, given that death-related thoughts are more accessible to females than males [Lester, 1972], suggesting that female participants would be more sensitive to the manipulation of MS. However, future studies are needed to replicate our findings and contrast those with findings for male participants to clarify potential gender differences. Second, our design did not allow the direct assessment of potential modulators that could have determined either hash or peaceful reactions after MS priming [see also Jonas and Fritsche, 2013], including social contexts and cultures [Hirschberger et al., 2005; Jonas et al., 2013; Schimel et al., 2006]. Third, our study did not identify the reliable associations between self-esteem and MS-induced effects, which would have been predicted by Terror Management Theory [see also Yanagisawa et al., 2016]. However, it is noteworthy that the associations between self-esteem and MS-induced effects remain inconclusive in the literature, partly due to the possibility that self-reported self-esteem may only provide coarse estimates of the construct [Burke et al., 2010].

In summary, our results shed light on the behavioral and neural signatures underlying the effects of death reminders on in-group bias of costly punishment. We demonstrated that behavioral discriminations against out-group members were attenuated by the reminders of mortality. Underlying these behavioral effects were less distinct neural representations of membership in brain regions implicated in mentalizing (dmPFC, TPJ), encoding aversive feelings (ACC), and decision selection (dlPFC) under MS priming. Furthermore, distributed activation patterns in the punishment-related networks were found to reliably encode behavioral in-group bias of costly punishment in the control condition but not in the MS condition. Our results complement previous behavioral observations that socially constructive behaviors, such as affiliating with others, can help defend against the possibility of feeling annihilation anxiety associated with increased existential threat [Hirschberger et al., 2008; Jonas et al., 2002]. These findings might have significant implications for understanding real-life intergroup interactions (e.g., peace negotiation) in the context of existential threat and provide a neurocognitive mechanism for socially constructive behaviors that can be instigated by death reminders.

ACKNOWLEDGMENTS

The authors thank Zhichao Xia and Jing Jiang for providing advice on data analyses and two anonymous reviewers for their helpful comments on an earlier draft of this manuscript.

REFERENCES

Büchel C, Holmes A, Rees G, Friston K (1998): Characterizing stimulus–response functions using nonlinear regressors in parametric fMRI experiments. Neuroimage 8:140–148.

Bagby RM, Parker JD, Taylor GJ (1994): The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. J Psychosom Res 38:23–32.

Baumgartner T, Knoch D, Hotz P, Eisenegger C, Fehr E (2011): Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. Nat Neurosci 14:1468–1474.

Baumgartner T, Götte L, Gügler R, Fehr E (2012): The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. Hum Brain Mapp 33:1452–1469.

Baumgartner T, Schiller B, Hill C, Knoch D (2013a): Impartiality in humans is predicted by brain structure of dorsomedial prefrontal cortex. Neuroimage 81:317–324.

Baumgartner T, Schiller B, Rieskamp J, Gianotti LR, Knoch D (2013b): Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. Soc Cogn Affect Neurosci nst023.

Becker E (1973): The Denial of Death. New York, NY. Free Press.

Bernhard H, Fischbacher U, Fehr E (2006): Parochial altruism in humans. Nature 442:912–915.

Boksem MA, Kostermans E, De Cremer D (2011): Failing where others have succeeded: Medial frontal negativity tracks failure in a social context. Psychophysiology 48:973–979.

Brainard DH (1997): The psychophysics toolbox. Spatial Vis 10: 433–436.

Bressler SL, Menon V (2010): Large-scale brain networks in cognition: Emerging methods and principles. Trend Cogn Sci 14: 277–290.

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.

Burke BL, Martens A, Faucher EH (2010): Two decades of terror management theory: A meta-analysis of mortality salience research. Person Soc Psychol Rev 14:155–195.

Chang LJ, Sanfey AG (2013): Great expectations: Neural computations underlying the use of social norms in decision-making. Soc Cogn Affect Neurosci 8:277–284.

Christie R, Geis FL (1970): Studies in Machiavellianism. New York, NY: Academic Press.

Civai C, Miniussi C, Rumiati RI (2014): Medial prefrontal cortex reacts to unfairness if this damages the self: A tDCS study. Soc Cogn Affect Neurosci nsu154.

Corradi-Dell'Acqua C, Hofstetter C, Vuilleumier P (2011): Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. J Neurosci 31:17996–18006.

Corradi-Dell'Acqua C, Civai C, Rumiati RI, Fink GR (2013): Disentangling self-and fairness-related neural mechanisms involved in the ultimatum game: An fMRI study. Soc Cogn Affect Neurosci 8:424–431.

Corradi-Dell'Acqua C, Hofstetter C, Vuilleumier P (2014): Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. Soc Cogn Affect Neurosci 9:1175–1184.

Corradi-Dell'Acqua C, Tusche A, Vuilleumier P, Singer T (2016): Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. Nat Commun 7:

Costa PT, MacCrae RR (1992): Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): Professional manual. Odessa, FL: Psychol Assess Resource.

Cui Z, Xia Z, Su M, Shu H, Gong G (2016): Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach. Hum Brain Mapp 37:1443–1458.

Davis MH (1980): A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology 10:85.

Engel C (2011): Dictator games: A meta study. Exp Econom 14: 583–610.

Fehr E, Gächter S (2002): Altruistic punishment in humans. Nature 415:137–140.

Fehr E, Fischbacher U (2003): The nature of human altruism. Nature 425:785–791.

Fehr E, Fischbacher U (2004): Third-party punishment and social norms. Evol Hum Behav 25:63–87.

Feng C, Luo YJ, Krueger F (2015): Neural signatures of fairness-related normative decision making in the ultimatum game: A coordinate-based meta-analysis. Hum Brain Mapp 36:591–602.

Fernandes O Jr, Portugal LC, Rita de Cássia SA, Arruda-Sanchez T, Rao A, Volchan E, Pereira M, Oliveira L, Mourao-Miranda J. (2017): Decoding negative affect personality trait from patterns of brain activation to threat stimuli. NeuroImage 145:337–345.

Florian V, Mikulincer M, Hirschberger G (2001): An existentialist view on mortality salience effects: Personal hardiness, death-thought accessibility, and cultural worldview defence. Br J Soc Psychol 40:437–453.

Frith U, Frith CD (2003): Development and neurophysiology of mentalizing. Phil Trans R Soc Lond B: Biol Sci 358:459–473.

Gächter S, Herrmann B (2009): Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. Phil Trans Roy Soc Lond B: Biol Sci 364:791–806.

Güth W, Schmittberger R, Schwarze B (1982): An experimental analysis of ultimatum bargaining. J Econom Behav Org 3:367–388.

Gabay AS, Radua J, Kempton MJ, Mehta MA (2014): The Ultimatum Game and the brain: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 47:549–558.

Gailliot MT, Stillman TF, Schmeichel BJ, Maner JK, Plant EA (2008): Mortality salience increases adherence to salient norms and values. Person Soc Psychol Bull 34:993–1003.

Goldenberg JL, Hart J, Pyszczynski T, Warnica GM, Landau M, Thomas L (2006): Ambivalence toward the body: Death, neuroticism, and the flight from physical sensation. Person Soc Psychol Bull 32:1264–1277.

Grecucci A, Giorgetta C, van't Wout M, Bonini N, Sanfey AG (2013): Reappraising the ultimatum: An fMRI study of emotion regulation and decision making. Cereb Cortex 23:399–410.

Green D, Swets J (1966): Signal Detection Theory and Psychophysics, Wiley, New York, 889.

Greenberg J, Pyszczynski T, Solomon S (1986): The causes and consequences of a need for self-esteem: A terror management theory. In: Baumeister RR, editor. Public Self and Private Self. New York, NY: Springer. pp 189–212.

Greenberg J, Pyszczynski T, Solomon S, Rosenblatt A, Veeder M, Kirkland S, Lyon D (1990): Evidence for terror management theory II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview. J Person Soc Psychol 58:308.

Greenberg J, Simon L, Pyszczynski T, Solomon S, Chatel D (1992): Terror management and tolerance: Does mortality salience

always intensify negative reactions to others who threaten one's worldview? J Person Soc Psychol 63:212.
Greenberg J, Pyszczynski T, Solomon S, Simon L, Breus M (1994): Role of consciousness and accessibility of death-related thoughts in mortality salience effects. J Person Soc Psychol 67:627.
Greenberg J, Porteus J, Simon L, Pyszczynski T, Solomon S (1995a): Evidence of a terror management function of cultural icons: The effects of mortality salience on the inappropriate use of cherished cultural symbols. Person Soc Psychol Bull 21: 1221–1228.
Greenberg J, Simon L, Harmon-Jones E, Solomon S, Pyszczynski T, Lyon D (1995b): Testing alternative explanations for mortality salience effects: Terror management, value accessibility, or worrisome thoughts? Eur J Soc Psychol 25:417–433.
Greenberg J, Solomon S, Pyszczynski T (1997): Terror management theory of self-esteem and cultural worldviews: Empirical assessments and conceptual refinements. Adv Exp Soc Psychol 29:61–139.
Griffin DW, Bartholomew K (1994): Models of the self and other: Fundamental dimensions underlying measures of adult attachment. J Person Soc Psychol 67:430.
Han S, Qin J, Ma Y (2010): Neurocognitive processes of linguistic cues related to death. Neuropsychologia 48:3436–3442.
Harlé KM, Chang LJ, van't Wout M, Sanfey AG (2012): The neural mechanisms of affect infusion in social economic decision-making: A mediating role of the anterior insula. Neuroimage 61:32–40.
Harmon-Jones E, Greenberg J, Solomon S, Simon L (1996): The effects of mortality salience on intergroup bias between minimal groups. Eur J Soc Psychol 26:677–681.
Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001): Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.
Haxby JV, Connolly AC, Guntupalli JS (2014): Decoding neural representational spaces using multivariate pattern analysis. Ann Rev Neurosci 37:435–456.
Hebart MN, Görgen K, Haynes JD (2014): The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. Front Neuroinform 8:88.
Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas JC, Gurven M, Gwako E, Henrich N (2006): Costly punishment across human societies. Science 312:1767–1770.
Hirschberger G, Florian V, Mikulincer M (2005): Fear and compassion: A terror management analysis of emotional reactions to physical disability. Rehab Psychol 50:246.
Hirschberger G, Ein-Dor T, Almakias S (2008): The self-protective altruist: Terror management and the ambivalent nature of prosocial behavior. Person Soc Psychol Bull 34:666–678.
Jonas E, Fritsche I (2013): Destined to die but not to wage war: How existential threat can contribute to escalation or de-escalation of violent intergroup conflict. Am Psychol 68:543.
Jonas E, Schimel J, Greenberg J, Pyszczynski T (2002): The Scrooge effect: Evidence that mortality salience increases prosocial attitudes and behavior. Person Soc Psychol Bull 28:1342–1353.
Jonas E, Martens A, Niesta Kayser D, Fritsche I, Sullivan D, Greenberg J (2008): Focus theory of normative conduct and terror-management theory: The interactive impact of mortality salience and norm salience on social judgment. J Person Soc Psychol 95:1239.
Jonas E, Sullivan D, Greenberg J (2013): Generosity, greed, norms, and death–Differential effects of mortality salience on charitable behavior. J Econom Psychol 35:47–57.
Kahneman D, Knetsch JL, Thaler RH (1986): Fairness and the assumptions of economics. J Business 59:S285–S300.
Koole SL, Greenberg J, Pyszczynski T (2006): Introducing science to the psychology of the soul experimental existential psychology. Curr Direct Psychol Sci 15:212–216.
Kriegeskorte N, Goebel R, Bandettini P (2006): Information-based functional brain mapping. Proc Natl Acad Sci U S A 103: 3863–3868.
Krueger F, Hoffman M (2016): The emerging neuroscience of third-party punishment. Trend Neurosci 39:499–501.
Krueger F, Barbey AK, Grafman J (2009): The medial prefrontal cortex mediates social event knowledge. Trend Cogn Sci 13:103–109.
Kubota JT, Li J, Bar-David E, Banaji MR, Phelps EA (2013): The price of racial bias intergroup negotiations in the ultimatum game. Psychol Sci 24:2498–2504.
Lester D (1972): Studies in death attitudes: Part two. Psychol Rep 30:440.
Li X, Liu Y, Luo S, Wu B, Wu X, Han S (2015): Mortality salience enhances racial in-group bias in empathic neural responses to others' suffering. NeuroImage 118:376–385.
Lieberman JD, Arndt J, Personius J, Cook A (2001): Vicarious annihilation: The effect of mortality salience on perceptions of hate crimes. Law Hum Behav 25:547.
Luo S, Shi Z, Yang X, Wang X, Han S (2014): Reminders of mortality decrease midcingulate activity in response to others' suffering. Soc Cogn Affect Neurosci 9:477–486.
Ma-Kellams C, Blascovich J (2011): Culturally divergent responses to mortality salience. Psychol Sci 22:1019–1024.
Ma X, Luo L, Geng Y, Zhao W, Zhang Q, Kendrick KM (2014): Oxytocin increases liking for a country's people and national flag but not for other cultural symbols or consumer products. Front Behav Neurosci 8:
Miltner WH, Braun CH, Coles MG (1997): Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. J Cogn Neurosci 9:788–798.
Mumford JA, Poldrack RA (2014): Adjusting mean activation for reaction time effects in BOLD fMRI. OHBM poster, Available at: https://ww4.aievolution.com/hbm1401/files/content/abstracts/43589/2053_Mumford.pdf. Last accessed: 29 October 2016.
Niesta D, Fritsche I, Jonas E (2008): Mortality salience and its effects on peace processes: A review. Soc Psychol 39:48–58.
Patton JH, Stanford MS (1995): Factor structure of the Barratt impulsiveness scale. J Clin Psychol 51:768–774.
Pelli DG (1997): The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vis 10: 437–442.
Poldrack RA, Mumford JA, Nichols TE (2011): Handbook of Functional MRI Data Analysis. New York, NY: Cambridge University Press.
Pyszczynski T, Greenberg J, Solomon S (1997): Why do we need what we need? A terror management perspective on the roots of human social motivation. Psychol Inquiry 8:1–20.
Quirin M, Loktyushin A, Amdt J, Kustermann E, Lo YY, Kuhl J, Eggert L (2012): Existential neuroscience: A functional magnetic resonance imaging investigation of neural responses to reminders of ones mortality. Soc Cogn Affect Neurosci 7:193–198.
Raskin RN, Hall CS (1979): A narcissistic personality inventory. Psychol Rep 45:590–590.
Rilling JK, Sanfey AG (2011): The neuroscience of social decision-making. Ann Rev Psychol 62:23–48.

Rilling JK, Dagenais JE, Goldsmith DR, Glenn AL, Pagnoni G (2008): Social cognitive neural networks during in-group and out-group interactions. Neuroimage 41:1447–1461.

Rosenberg M. (1965): Society and the adolescent self-image. Princeton, N. J.: Princeton University Press.

Rosenblatt A, Greenberg J, Solomon S, Pyszczynski T, Lyon D (1989): Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. J Person Soc Psychol 57:681.

Routledge C, Ostafin B, Jhl J, Sedikides C, Cathey C, Liao J (2010): Adjusting to death: The effects of mortality salience and self-esteem on psychological well-being, growth motivation, and maladaptive behavior. J Person Soc Psychol 99:897–916.

Schimel J, Wohl MJ, Williams T (2006): Terror management and trait empathy: Evidence that mortality salience promotes reactions of forgiveness among people with high (vs. low) trait empathy. Motiv Emotion 30:214–224.

Schindler S, Reinhard MA, Stahlberg D (2012): Mortality salience increases personal relevance of the norm of reciprocity 1. Psychol Rep 111:565–574.

Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, Phillips C, Richiardi J, Mourão-Miranda J (2013): PRoNTo: Pattern recognition for neuroimaging toolbox. Neuroinformatics 11:319–337.

Schrouff J, Monteiro J, Joao Rosa M, Portugal L, Phillips C, Mourao-Miranda J (2014): Can we interpret linear kernel machine learning models using anatomically labelled regions? OHBM poster, Available at: http://orbi.ulg.ac.be/handle/2268/170848. Last accessed: 29 October 2016.

Stoddard O, Leibbrandt A (2014): An experimental study on the relevance and scope of nationality as a coordination device. Econom Inquiry 52:1392–1407.

Strobel A, Zimmermann J, Schmitz A, Reuter M, Lis S, Windmann S, Kirsch P (2011): Beyond revenge: Neural and genetic bases of altruistic punishment. Neuroimage 54:671–680.

Sylwester K, Herrmann B, Bryson JJ (2013): Homo homini lupus? Explaining antisocial punishment. J Neurosci Psychol Econom 6:167.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002): Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

Uddin LQ (2015): Salience processing and insular cortical function and dysfunction. Nat Rev Neurosci 16:55–61.

Van den Bos K, Miedema J (2000): Toward understanding why fairness matters: The influence of mortality salience on reactions to procedural fairness. J Person Soc Psychol 79:355.

van Marle HJ, Hermans EJ, Qin S, Fernandez G (2009): From specificity to sensitivity: How acute stress affects amygdala processing of biologically salient stimuli. Biol Psychiatry 66: 649–655.

Viney W, Waldman DA, Barchilon J (1982): Attitudes toward punishment in relation to beliefs in free will and determinism. Hum Relat 35:939–949.

von Dawans B, Fischbacher U, Krischbaum C, Fehr E, Heinrichs M (2012): The social dimension of stress reactivity acute stress increases prosocial behavior in humans. Psychol Sci 23: 651–660.

Watson D, Clark LA, Tellegen A (1988): Development and validation of brief measures of positive and negative affect: The PANAS scales. J Person Soc Psychol 54:1063.

Wisman A, Koole SL (2003): Hiding in the crowd: Can mortality salience promote affiliation with others who oppose one's worldviews? J Person Soc Psychol 84:511–526.

Wisniewski D, Goschke T, Haynes JD (2016): Similar coding of freely chosen and externally cued intentions in a frontoparietal network. NeuroImage 134:450–458.

Woolgar A, Golland P, Bode S (2014): Coping with confounds in multivoxel pattern analysis: What should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. Neuroimage 98:506–512.

Xiang T, Lohrenz T, Montague PR (2013): Computational substrates of norms and their violations during social exchange. J Neurosci 33:1099–1108.

Xiao E, Houser D (2005): Emotion expression in human punishment behavior. Proc Natl Acad Sci U S A 102:7398–7401.

Yamagishi T, Horita Y, Takagishi H, Shinada M, Tanida S, Cook KS (2009): The private rejection of unfair offers and emotional commitment. Proc Natl Acad Sci 106:11520–11523.

Yanagisawa K, Kashima ES, Moriya H, Masui K, Furutani K, Nomura M, Yoshida H, Ura M (2013): Non-conscious neural regulation against mortality concerns. Neurosci Lett 552:35–39.

Yanagisawa K, Abe N, Kashima ES, Nomura M (2016): Self-esteem modulates amygdala-ventrolateral prefrontal cortex connectivity in response to mortality threats. J Exp Psychol Gen 145:273–283.