Neural circuits in the brain that are activated when mitigating criminal sentences

Makiko Yamada[1,2], Colin F. Camerer[3], Saori Fujie[1], Motoichiro Kato[4], Tetsuya Matsuda[5], Harumasa Takano[1], Hiroshi Ito[1], Tetsuya Suhara[1] & Hidehiko Takahashi[1,2,5,6]

In sentencing guilty defendants, jurors and judges weigh 'mitigating circumstances', which create sympathy for a defendant. Here we use functional magnetic resonance imaging to measure neural activity in ordinary citizens who are potential jurors, as they decide on mitigation of punishment for murder. We found that sympathy activated regions associated with mentalising and moral conflict (dorsomedial prefrontal cortex, precuneus and temporo-parietal junction). Sentencing also activated precuneus and anterior cingulate cortex, suggesting that mitigation is based on negative affective responses to murder, sympathy for mitigating circumstances and cognitive control to choose numerical punishments. Individual differences on the inclination to mitigate, the sentence reduction per unit of judged sympathy, correlated with activity in the right middle insula, an area known to represent interoception of visceral states. These results could help the legal system understand how potential jurors actually decide, and contribute to growing knowledge about whether emotion and cognition are integrated sensibly in difficult judgments.

[1] Department of Molecular Neuroimaging, Molecular Imaging Center, National Institute of Radiological Sciences, 4-9-1 Anagawa, Inage-ku, Chiba 263-8555, Japan. [2] Decoding and Controlling Brain Information, Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan. [3] Computation and Neural Systems and Humanities and Social Science Divisions, California Institute of Technology, Pasadena, California 91125-7700, USA. [4] Department of Neuropsychiatry, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. [5] Tamagawa University Brain Science Institute, 6-1-1 Tamagawa Gakuen, Machida, Tokyo 194-8610, Japan. [6] Department of Neuropsychiatry, Kyoto University School of Medicine, 54 Shogoin-kawara-cho, Sakyo-ku, Kyoto 606-8507, Japan. Correspondence and requests for materials should be addressed to M.Y. or H.T. (email: myamada@nirs.go.jp or hidehiko@nirs.go.jp).

Philosophers, psychologists and legal scholars have long debated whether mercy, sympathy and compassion should reduce moral culpability of legal defendants. People do have negative emotional responses to a wide range of situational factors that are not normatively justifiable because they are not considered evidence, or appeal to 'prejudices and sympathy', which jurors are typically instructed to ignore, for example, gruesome crime scene pictures lead to more mock jury convictions when they are in colour (compared with black and white[1]). Other studies show that evidence that provides negative emotions leads to more punitive judgments[2]. A mock trial study reported that jurors found a defendant less guilty when they heard a defense attorney urging sympathy for the defendant[3].

Studies of moral cognition in hypothetical non-legal scenarios have revealed the increased activity in emotion-related brain regions (that is, insula, amygdala and orbitofrontal cortex) and decreased activity in cognitive processes (that is, dorsolateral prefrontal and parietal areas) when participants contemplated morally charged actions, such as a trolley dilemma involving killing lives of some people to save others[4]. Other studies indicate that people do sometimes punish norm violators in non-legal or legal situations, driven by moral judgments in which negative emotional reactions have a critical role[5,6]. Prosocial emotions, such as sympathy, also influence decision making (for example, charitable giving)[3]; however, there is currently no direct cognitive and neural evidence for how sympathy is translated into legal outcomes.

All these studies raise an important question of whether moral judgments and accompanying emotions are used reasonably (as legal rules require) or not. Neuroscience is now entering this debate about defendants, but very little is known about whether mental activity of juries and judges conforms to legal principles. The case we explore here is 'mitigating circumstances', a rare case where emotions, such as sympathy, are allowed to influence juror judgment.

The legal domain is unusual because it may be especially challenging to map emotions into numerical legal outcomes. This difficulty is found in studies of hypothetical punitive damages in tort cases, which not only show that jurors typically agree on moral outrageousness of actions, but also show large disagreement about how outrage is mapped to punitive dollar awards[7,8].

Weighing mitigation puts an unusual burden on people (*qua* jurors) to have appropriate emotional sympathetic reactions, then encode their emotions into prison years. Uncovering the cognitive and neural mechanisms of sympathy that motivate mitigation will inform the role of emotion in jurors' decision process, and perhaps the ultimate policy issue of what role emotional evidence can and should have in trials. Understanding the neuroscientific basis of legal mitigation adds to a basic understanding of moral neuroscience. Neural evidence could also advance theory and practice of law, as so little is known about whether the mental activity of juries and judges conforms to normative legal principles[9].

In summary, our results revealed that sympathy activated brain regions associated with mentalising and moral conflict, including dorsomedial prefrontal cortex (DMPFC), precuneus and temporoparietal junction (TPJ). Sentence mitigation also recruited these sympathy regions, uncovering neural evidence for a close relationship between sympathy and mitigation. Furthermore, individual differences on the inclination to mitigate were reflected in differential middle insula activity. These findings do not just contribute to the field of neuroscience, but could help lawmakers to understand jurors' decision making and their individual differences in trials.

Results

Sympathy and no-sympathy scenarios. We measured brain activity using functional magnetic resonance imaging (fMRI) while subjects are making hypothetical sentence reduction decisions, in dramatic scenarios adapted from actual murder cases. Sympathy-related brain activity was collected during reading circumstances pertaining to defendants' crimes (Fig. 1a). Only actual Japanese murders were used, so the crime was serious, uniform across trials and lifelike. This simple design was chosen to generate engagement and limit nuisance brain activity due to subtle differences in crimes and plausibility of artificially created scenarios.

Mitigating circumstances were of two types: those that would induce sympathy and those that would not. The sympathy scenarios included desperate situations of defendants suffering from domestic violence, disease or poverty. Figure 1a gives one example of each type. The intentionality and severity of the murders were matched between conditions (see Supplementary Methods). After reading about the circumstances, subjects decided how much they would change the sentence given for the defendant (initially 20 years) if they were on a jury. After scanning, subjects were again presented with the same scenarios and asked to rate how much sympathy they felt for the defendant, using a visual analogue scale.

Sympathy and punishment ratings. The sympathy manipulation is internally valid because participants gave significantly higher sympathy ratings to those defendants with sympathy circumstances compared with those with no-sympathy circumstances ($n = 22$, paired *t*-test, $t_{21} = -18.94$, $P < 0.001$, Fig. 1b). They also reduced sentences much more in sympathy circumstances ($n = 22$, paired *t*-test, $t_{21} = 11.82$, $P < 0.001$, Fig. 1c). Unsurprisingly, sympathy and punishment (sentence length) were highly negatively correlated ($n = 32$ stories, linear regression analysis, $P < 0.001$, $R^2 = 0.97$, Fig. 1d).

fMRI results were analysed using standard generalized linear model regression techniques (see Methods). A short block design was used where regressors were included for the various events of the trials (Fig. 1a). Interaction terms corresponding to punishment and sympathy ratings interacted with trial onsets that were added as parametric regressors.

Brain regions associated with sympathy scenarios. We first analysed brain areas exhibiting a stronger response in sympathy scenarios than no-sympathy scenarios. Precuneus, left TPJ and DMPFC showed larger activities for sympathy than for no-sympathy scenarios ($n = 22$, one-sample *t*-test, Table 1). These regions are related to mentalization and sympathy as discussed below in detail, which confirms that the experimental manipulation of sympathy produced results consistent with the sympathy ratings data.

Brain regions associated with punishment and sympathy. We then searched for brain regions that responded, during the description, to the subjects' trial-by-trial ratings of sympathy and their amounts of punishment reduction. Activity in precuneus, DMPFC and left TPJ were correlated with sympathy ($P < 0.05$, small-volume-corrected, Fig. 2, Supplementary Table S1). Signal increase in precuneus and DMPFC were also associated with the reduction of punishment ($P < 0.05$, small-volume-corrected, Fig. 2: note that a small TPJ region was also activated in sentence reduction, but only with $k = 6$ voxels). Thus, precuneus and DMPFC were commonly activated by both sympathy and reduction of punishment. Sentence reduction was also associated with activity in anterior cingulate cortex (ACC). All regions showing a whole-brain correlation at $P < 0.001$ are listed in Supplementary Tables S1 and S2.

Brain regions associated with an inclination to mitigate. Next, we constructed an individual-specific measure of an inclination to mitigate, by reducing sentences, as a function of sympathy. This measure comes from a simple linear regression on each individual's decisions: punishment $= b_0 + b_1 *$sympathy$+$ error[10]. A measure of an inclination to mitigate, the reduction in sentence per unit of sympathy, was given by the b_1 coefficient of the regression. This number represents a complex mapping from an emotional response to a number representing prison time for a defendant (a years-per-emotion coefficient).
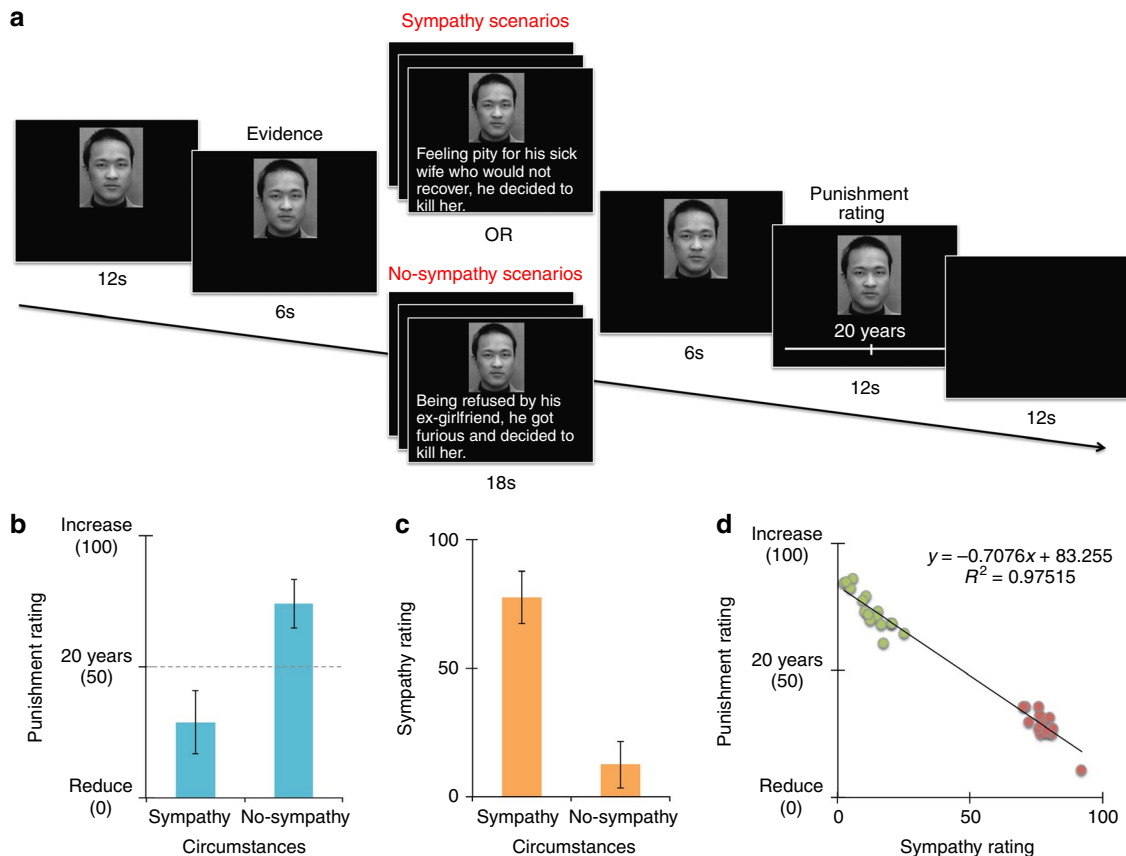
Figure 1 | Task design and behavioural performance. (**a**) Study paradigm. (**b**) Mean punishment ratings for sympathy and no-sympathy trials ($n = 22$, paired t-test, $t_{21} = -18.94$, $P < 0.001$). (**c**) Mean sympathy ratings for sympathy and no-sympathy trials ($n = 22$, paired t-test, $t_{21} = 11.82$, $P < 0.001$). (**d**) Correlation between sympathy and punishment ratings for sympathy stories (red circles) and no-sympathy stories (green circles). Error bars indicate s.d.

These b_1 coefficients were negative for all subjects (mean $= -6.5 \pm 0.2$, linear regression analysis, all $P < 0.001$), indicating that the feeling of sympathy did correlate with reduction of punishment, but to different degrees across subjects. A negative linear regression between the individual-specific b_1 coefficient and blood oxygenation level-dependent responses in sympathy minus no-sympathy trials found activity in the right middle insula ($P < 0.05$, small-volume-corrected, Fig. 3). Individuals who had larger activities in the insula when reading circumstances showed higher tendencies to mitigate, reducing sentencing years more as their sympathy increased.

Discussion

Comparison of activity during judgments of sympathy, and sentence reduction, suggest that activity in DMPFC (also known as paracingulate), precuneus (also known as posterior cingulate) and TPJ reflect a judgment-action circuit, which is illustrated in Fig. 2. Strength of sympathy judgments is associated with activity in DMPFC, precuneus and TPJ. DMPFC is involved in general mentalising[11] and is active when empathizing[12] or sympathizing[13] with others in pain. Precuneus has been linked to subjective perspective taking[14–16]. TPJ is also commonly identified as a part of a theory-of-mind circuit[17], including mentalising about intentions[18], and was activated in one study on judging innocence of intentions of people who caused harm[19]. This suggests that the sympathy judgment is an engagement with a reasoned simulation of what the defendant was thinking when committing the crime or how most people would judge the normative basis for mitigation. Note that although precuneus and TPJ can each be activated by non-social demands as well (for example, attention reorienting[20]), they are rarely co-activated as a group unless social cognitive demands are present.

Regions activated by the punishment reduction judgment include a large region of precuneus and smaller regions in DMPFC and ACC. As noted, precuneus has been linked to simulation on the self to understand others, and is also active when compassionating[21] and forgiving others[22]. The activation in this area is correlated with more iterated steps[23] and higher-value[24] strategic thinking in game theory tasks. Note that both feelings of sympathy and judged mitigation of punishment were encoded in activity in precuneus. This overlapping activity suggests that the precuneus may be a region that accepts emotional judgment input and maps it into concrete punishment actions.

ACC is a region now thought to be activated by negative affect[25,26], positive reward[27] and cognitive control[26,28]. In our context, mapping emotional sympathy into numerical sentencing requires high-level executive function by weighing negative affective reaction to murder, positive sympathy for the murder's mitigating circumstances and exerting cognitive control to choose numerical punishments that weigh these emotions consistently.

Smaller regions in caudate also showed differential activation in response to punishment reduction (Supplementary Table S2). Caudate activity is consistent with the hypothesis that sentence reduction is encoded as a special kind of prosocial 'charity' (as mitigation is like giving to charity[29]), as other prosocial choices activate the caudate too[30].

One fMRI study elicited punishment judgments in artificial scenarios varying offender culpability[6] and reports right DLPFC activity associated with responsibility judgments. We speculate that the absence of right DLPFC activation in our study is because there is no doubt about the defendants' guilt, so the most morally burdensome question of guilt versus innocence is resolved (right DLPFC is discharged from jury duty, so to speak). Activity then shifts to

Table 1 | Areas exhibiting a stronger response in sympathy scenarios.

Region	Side	MNI coordinates			Z	k	SVC
Posterior cingulate cortex	**R**	**2**	**−50**	**28**	**4.05**	**2946**	**0.001**
Precuneus	**L**	**−2**	**−68**	**42**	**3.85**		
Precuneus	**R**	**0**	**−18**	**28**	**3.73**		
Middle frontal gyrus	R	36	16	60	4.11	1501	
Superior frontal gyrus	R	20	32	58	4.08		
Superior frontal gyrus	L	−18	24	64	4.06		
Temporo-parietal junction (middle temporal gyrus)	**L**	**−50**	**−72**	**20**	**4.15**	**535**	**0.001**
Temporo-parietal junction (angular gyrus)	**L**	**−46**	**−64**	**34**	**3.58**		
Inferior frontal gyrus (p. orbitalis)	L	−40	34	−14	4.37	296	
Inferior frontal gyrus (p. triangularis)	L	−50	36	18	3.62		
Inferior frontal gyrus (p. triangularis)	L	−42	34	0	3.54		
Dorsomedial prefrontal cortex (superior medial gyrus)	**L**	**−4**	**44**	**22**	**3.50**	**217**	**0.008**
Dorsomedial prefrontal cortex (superior medial gyrus)	**L**	**−4**	**54**	**28**	**3.49**		
Superior frontal gyrus	**L**	**−12**	**52**	**34**	**3.33**		
Caudate nucleus	R	12	18	8	3.82	179	
Hippocampus	R	26	−32	−8	3.47	151	
Parahippocampal gyrus	R	28	−24	−22	3.43		
Caudate nucleus	L	−8	18	2	3.53	101	
Caudate nucleus	L	−12	14	10	3.37		
Middle temporal gyrus	L	−52	4	−32	3.72	81	
Lingual gyrus	R	16	−44	−8	3.50	59	
Parahippocampal gyrus	R	26	10	−30	3.68	48	
Dorsomedial prefrontal cortex (superior medial gyrus)	**R**	**10**	**44**	**50**	**3.37**	**23**	**0.011**
Superior frontal gyrus	**R**	**16**	**16**	**46**	**44**		
Inferior frontal gyrus (p. orbitalis)	L	−26	26	−26	3.56	22	
Temporal pole	L	−36	22	−32	3.19		
Medial temporal pole	R	44	16	−38	3.37	13	
Inferior frontal gyrus (p. orbitalis)	R	38	28	−12	3.22	13	
Temporal pole	L	−34	6	−20	3.17	12	

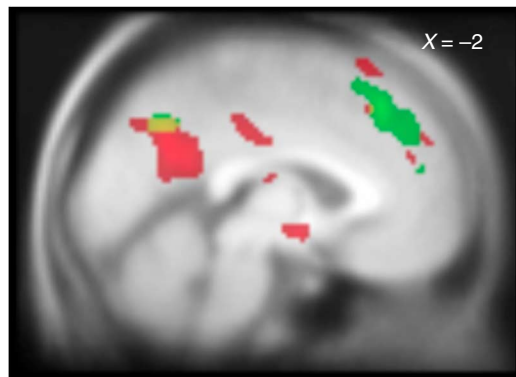Abbreviation: SVC, small volume correction.
A priori regions of interest are in bold ($P < 0.001$, $k > 10$).

Figure 2 | Brain regions activated during trial-by-trial sympathy and punishment reduction. Regions in which activity correlated with parametric regressors of increasing sympathy (green) and reduced punishment (red). Common areas were found in precuneus (yellow). The image is shown at $P < 0.001$ (uncorrected; $n = 22$, one-sample t-tests).

Figure 3 | Individual differences in inclination towards mitigation. Activation in right middle insula in the contrast sympathy minus no-sympathy trials (**a**, MNI 34/−8/12; $n = 22$, one-sample t-test, $P = 0.023$, small-volume-corrected) correlated with participant-wise coefficient of mitigation (**b**, $n = 22$, Pearson $r = −0.62$, $P = 0.002$).

areas generally associated with moral and cognitive conflicts instead (DMPFC, precuneus and ACC).

The individual difference analysis showed that activity in the right mid-insula was related to the individual differences on the inclination to mitigate (mapping sympathy to sentences). Middle/posterior insula has been suggested to encode economic inequity[31], norm violation[32], somatosensory representation[33], monitoring internal states[34] and heart rate during a compassionate state[35]. These types of interoceptive processing in middle insula suggest that it is sensitive to emotions linked to sociality. Our study provides unusual evidence of this processing associated with a unique high-impact social judgment that affects others.
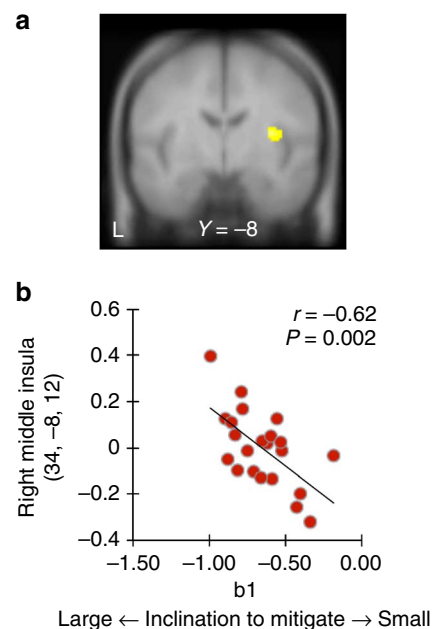
The identified brain activity provides new insight into the capacity of the average brain to translate sympathetic feelings into appropriate legal action. A plausible neuro-legal standard is that the influence

of sympathetic reaction on difficult sentencing should recruit brain areas that process sympathy judgments in general, as well as areas that encode difficult decision conflict. Activity in these 'sympathy' regions is evident in our study when judging sympathy alone, and in choosing sentence mitigation. However, not every brain maps sympathy to prison sentences in the same numerical way (as reflected in differential mid-insula activity). Differences in these brain circuits between individuals, suggest that differential juror responses might need to be considered unequally. There is also mixed evidence about the normative basis of legal judgment, including a recent finding that judges' decisions are affected by timing of meals[36].

The current finding would also contribute to the attribution literature on situational correction, as there has not been any fMRI work on this. People often attribute behaviour to the corresponding personal disposition, which is, yet, corrected based on situational inducements too[37]. This is an apt psychological function behind mitigating circumstances that is to make someone less culpable, and the revealed brain activity might be associated with this type of situational correction.

Finally, we note that many legal principles treat emotional responses as likely to be prejudicial and prone to inflammatory manipulation (that is, an ideal juror would suppress them and legal rules limit their influence). Weighing mitigating circumstances during sentencing (*after* a verdict) represents an unusual case in which emotional sympathy judgment is actually required. Japanese criminal law, for example, requires that the decision among these sentences be based on mitigating circumstances[38].

Ironically, the fact that sympathy is clearly evident in brain activity, and influences sentence mitigation (as it should), raises interest in the opposite question: can people also suspend emotions when the law instructs them to? More generally, a deeper understanding of the brain could help figure out how highly evolved brain structures, which were sculpted to maintain order in small-scale ancestral societies, can be put to work under modern legal rules in much more challenging cases to create modern justice.

Methods

Participants. A total of 26 right-handed healthy subjects (12 males; mean age = 21.5 ± 1.8 (s.d.) years) participated in the study. All subjects had normal or corrected-to-normal vision, no history of neurological or psychiatric disorder and were not taking medications that interfere with the performance of fMRI. The study was approved by the Ethics and Radiation Safety Committee of the National Institute of Radiological Sciences (Chiba, Japan). Data from four participants were discarded owing to excessive movement (three subjects) and sleep (one subject). The data from 22 subjects (11 males; mean age = 21.5 ± 1.9 (s.d.) years) were included and analysed.

Stimuli and the task. A total of 32 faces of defendants and criminal acts were prepared in the study. Faces were drawn from a set of Asian face stimuli developed by CAS-PEAL face database[39]. Pilot participants ($n = 30$) rated attractiveness for each faces using an 11-point scale of −5 (not attractive at all) to +5 (very attractive). The average attractiveness for 32 faces was 0.15 ± 0.70.

All criminal acts were the pure murder cases, which were modified from real stories taken from a precedent search engine provided by Supreme Court of Japan (www.courts.go.jp) and from news articles on the web. We prepared 32 scenarios and consulted a judicial expert for advice on how externally valid the scenarios were. Each scenario composed of the fact (when, where and to whom) and the circumstance (why the defendant committed the murder). Half of the scenarios would induce sympathy for the defendants who suffer from domestic violence, death to disease or poverty (sympathy scenarios), and the other half that would not (no-sympathy scenarios). Responsibility, intentionality and severity of crime were matched between scenarios (see Supplementary Methods for all scenarios used in the study).

Subjects were instructed to judge crimes as if they were jurors (lay judges). Before entering the scanner, subjects read each criminal case and were asked to indicate if they have known them through media. The aim of this session was to ensure and ease comprehension in the scanner that constrain duration of presenting each scenario. During the fMRI task, subjects read each scenario and rated how much they would like to reduce or increment the sentence (all 20 years) given for the defendant using a VAS (with scores ranging from 0 = reduce as much as possible to 100 = increment as much as possible, and 50 = given sentence of 20 years; Fig. 1a). We chose a sentence of 20 years as a reference point, based on a sentencing

guideline in Japan that gives 20 years in prison to a person who commits murder. Defendants' faces were presented together with the scenarios in order to make the experimental setting more realistic and to help participants to dedicate to the study. They completed two sessions, and each session contained 8 sympathy and 8 no-sympathy scenarios, 16 trials in total. The presentation of each scenario within a session was pseudorandomized. The order of sessions and the combination of faces and circumstance types were counterbalanced between subjects.

After the scan, subjects were shown the same scenarios again to rate how much they felt sympathy for the defendants using a VAS (with scores ranging from 0 = having no sympathy to 100 = having sympathy very much). They were also asked to fill out Interpersonal Reactivity Index (IRI), a self-report measure of dispositional empathy.

fMRI data acquisition. The functional imaging was conducted using a GE 3.0 Tesla Excite system to acquire gradient echo T2*-weighted echoplanar images with blood oxygenation level-dependent contrast. Each volume comprised 35 transaxial contiguous slices with a slice thickness of 3.8 mm to cover almost the whole brain (flip angle, 75; echo time, 25 ms; repetition time, 2000 ms; matrix, 64*64; and field of view, 24×24 cm²).

fMRI data preprocessing. Image analysis was performed using SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK), except for the overlay image shown in Fig. 2 that was created using the MRIcron software (http://www.sph.sc.edu/comd/rorden/mricron/). All volumes were realigned to the first volume of each session to correct for subject motion and were spatially normalised to the standard space defined by the Montreal Neurological Institute (MNI) template. After normalisation, images were spatially smoothed using a Gaussian kernel with a full-width-at-half-maximum of 8 mm. Intensity normalisation and high-pass temporal filtering (using a filter width of 128 s) were also applied to the data.

Generalized linear model. For the primary whole-brain analyses, two modelling approaches were used. In the first approach, trials from each of the sympathy and no-sympathy circumstances were modelled as separate conditions. This allowed separate estimation of the evoked response for each of the conditions at each voxel. The model also included regressors for face, criminal evidence, punishment, and fixation, as well as six head-motion parameters as regressors of no interest. We first calculated a first-level single-subject contrast for reading sympathy minus no-sympathy circumstances. Then, we calculated a second-level group contrasts using a one-sample *t*-test. In the second approach, all trials from reading circumstances modelled using a single condition (that is, overall task-related activation), and two additional orthogonal parametric regressors were included representing (a) the amount of reducing punishment and (b) the level of sympathy rating. We first calculated first-level single-subject contrasts for reading circumstances modulated by reducing punishment and for reading circumstances modulated by sympathy rating. Then, we calculated second-level group contrasts using one-sample *t* tests on the single-subject contrasts to identify brain regions whose activation correlated with the amount of reducing punishment and the sympathy rating.

A statistical threshold of $P < 0.05$ corrected for multiple spatial comparisons across the whole brain at cluster level was used, except for a priori hypothesized regions that were voxel thresholded at $P < 0.001$ uncorrected (only clusters involving $k > 10$ or more contiguous voxels were reported[40]). Small volume correction (a 10-mm radius sphere) was used on a priori regions of interest: DMPFC/ACC, precuneus/posterior cingulate cortex, TPJ and the insula.

References

1. Whalen, D. H. & Blanchard, F. A. Effects of photographic evidence on mock juror judgment. *J. Appl. Soc. Psychol.* **12,** 30–41 (1982).
2. Salerno, J. M. & Bottoms, B. L. Emotional evidence and jurors' judgments: the promise of neuroscience for informing psychology and law. *Behav. Sci. Law* **27,** 273–296 (2009).
3. Haegerich, T. M. & Bottoms, B. L. Empathy and jurors' decisions in patricide trials involving child sexual assault allegations. *Law Hum. Behav.* **24,** 421–448 (2000).
4. Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. An fMRI investigation of emotional engagement in moral judgment. *Science* **293,** 2105–2108 (2001).
5. Haidt, J. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* **108,** 814–834 (2001).
6. Buckholtz, J. W. *et al.* The neural correlates of third-party punishment. *Neuron* **60,** 930–940 (2008).
7. Kahneman, D., Schkade, D. & Sunstein, C. Shared outrage and erratic awards: the psychology of punitive damages. *J. Risk Uncertain.* **16,** 49–86 (1998).
8. Schkade, D., Sunstein, C. R. & Kahneman, D. Deliberating about dollars: the severity shift. *Columbia Law Rev.* **100,** 1139–1175 (2000).
9. Jones, O. D. Law, evolution and the brain: applications and open questions. *Philos. Trans. R Soc. Lond. B. Biol. Sci.* **359,** 1697–1707 (2004).
10. Hare, T. A., Camerer, C. F., Knoepfle, D. T. & Rangel, A. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* **30,** 583–590 (2010).

11. Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7,** 268–277 (2006).

12. Lamm, C., Decety, J. & Singer, T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* **54,** 2492–2502 (2011).

13. Decety, J. & Michalska, K. J. Neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood. *Dev. Sci.* **13,** 886–899 (2010).

14. Vogeley, K. *et al.* Neural correlates of first-person perspective as one constituent of human self-consciousness. *J. Cogn. Neurosci.* **16,** 817–827 (2004).

15. Ruby, P. & Decety, J. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* **4,** 546–550 (2001).

16. Farrer, C. & Frith, C. D. Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *Neuroimage* **15,** 596–603 (2002).

17. Greene, J. & Haidt, J. How (and where) does moral judgment work? *Trends Cogn. Sci.* **6,** 517–523 (2002).

18. Atique, B., Erb, M., Gharabaghi, A., Grodd, W. & Anders, S. Task-specific activity and connectivity within the mentalizing network during emotion and intention mentalizing. *Neuroimage* **55,** 1899–1911 (2011).

19. Young, L. & Saxe, R. Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia* **47,** 2065–2072 (2009).

20. Corbetta, M., Patel, G. & Shulman, G. L. The reorienting system of the human brain: from environment to theory of mind. *Neuron* **58,** 306–324 (2008).

21. Immordino-Yang, M. H., McColl, A., Damasio, H. & Damasio, A. Neural correlates of admiration and compassion. *Proc. Natl Acad. Sci. USA* **106,** 8021–8026 (2009).

22. Farrow, T. F. *et al.* Investigating the functional anatomy of empathy and forgiveness. *Neuroreport* **12,** 2433–2438 (2001).

23. Kuo, W. J., Sjostrom, T., Chen, Y. P., Wang, Y. H. & Huang, C. Y. Intuition and deliberation: two systems for strategizing in the brain. *Science* **324,** 519–522 (2009).

24. Bhatt, M. & Camerer, C. F. Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games Econ. Behav.* **52,** 424–459 (2005).

25. Etkin, A., Egner, T. & Kalisch, R. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* **15,** 85–93 (2011).

26. Shackman, A. J. *et al.* The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* **12,** 154–167 (2011).

27. Liu, X., Hairston, J., Schrier, M. & Fan, J. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* **35,** 1219–1236 (2011).

28. Botvinick, M. M., Cohen, J. D. & Carter, C. S. Conflict monitoring and anterior cingulate cortex: an update. *Trends. Cogn. Sci.* **8,** 539–546 (2004).

29. Harbaugh, W. T., Mayr, U. & Burghart, D. R. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* **316,** 1622–1625 (2007).

30. Fehr, E. & Camerer, C. F. Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* **11,** 419–427 (2007).

31. Hsu, M., Anen, C. & Quartz, S. R. The right and the good: distributive justice and neural encoding of equity and efficiency. *Science* **320,** 1092–1095 (2008).

32. Montague, P. R. & Lohrenz, T. To detect and correct: norm violations and their enforcement. *Neuron* **56,** 14–18 (2007).

33. Danziger, N., Faillenot, I. & Peyron, R. Can we share a pain we never felt? Neural correlates of empathy in patients with congenital insensitivity to pain. *Neuron* **61,** 203–212 (2009).

34. Damasio, A. R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (Harcourt Brace, New York, 1999).

35. Lutz, A., Greischar, L. L., Perlman, D. M. & Davidson, R. J. BOLD signal in insula is differentially related to cardiac function during compassion meditation in experts vs. novices. *Neuroimage* **47,** 1038–1046 (2009).

36. Danziger, S., Levav, J. & Avnaim-Pesso, L. Extraneous factors in judicial decisions. *Proc. Natl Acad. Sci. USA* **108,** 6889–6892 (2011).

37. Gilbert, D. T., Pelham, B. W. & Krull, D. S. On cognitive busyness: when person perceivers meet persons perceived. *J. Pers. Soc. Psychol.* **54,** 733–740 (1988).

38. Egashira, K., Kobayakawa, K., Nishida, N., Takahashi, H. & Nomi, Y. *Roppo Zensho [Comdemium of Laws] (Japanese)* (Yuhikaku, 2011).

39. Gao, W. *et al.* The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Trans. Syst., Man and Cy- bernetics, Part A* **38,** 149–161 (2008).

40. Forman, S. D. *et al.* Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* **33,** 636–647 (1995).

Acknowledgements

Author contributions

M.Y. and H.T. designed the experiment. M.Y. and F.S. prepared and conducted the experiment. M.Y. analysed the data and wrote the paper. C.F.C. edited the manuscript. H.T. coordinated subject recruitment. All authors discussed the results.

Additional information