Supporting Information for

Neurocomputational mechanisms underlying immoral decisions benefiting self or others

Chen Qu[1,†,*], Yang Hu[2,3,†,*], Zixuan Tang[2,4], Edmund Derrington[2,4], Jean-Claude Dreher[2,4*]

[1]Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, China

[2]Neuroeconomics Laboratory, Institut des Sciences Cognitives Marc Jeannerod, CNRS, Bron, France

[3]School of Psychological and Cognitive Sciences, Peking University, Beijing, China

[4] Université Claude Bernard Lyon 1, Lyon, France

[†]These authors equally contributed to this work

[*]Correspondence at fondest@163.com, huyang200606@gmail.com, or dreher@isc.cnrs.fr

This supporting information includes:

SI: Association selection

Procedure

An independent group of participants (N = 30) attended the rating task for selecting the associations used for the fMRI study. 16 different associations were used for the rating task. Besides those associations used in the fMRI study (i.e., 4 charities and 4 associations supporting the rights to own guns or hunting), there were another 4 charities and 4 associations supporting the legalization of soft drugs (see **vignettes and logos** below; also see **Fig. S7a**). Participants read vignettes (with the logo) which described the goal of each association. Vignettes were presented in a pseudo-random fashion.

Below each vignette, participants were asked to complete a series of questions by indicating degrees on a Likert rating scale, including 1) degree of familiarity (0 = “not at all”, 10 = “very much”), 2) degree of liking (-10 = “not at all” or “very negative”, 0 = “no preference”, 10 = “very much” or “very positive”), 3) degree of moral acceptance (-10 = “not at all”, 0 = “no preference”, 10 = “very much”), 4) degree of approval for its goal (-10 = “not at all”, 0 = “no preference”, 10 = “very much”), 5) amount of (hypothetical) donation (from 0 to 10; in CNY) and 6) degree of moral conflict while donating (0 = “not at all”, 10 = “very much”).

A rating summary of all associations in all questions by all participants is listed in **Table S5**. To induce stronger morally negative feelings we finally decided to use associations in support of gun ownership or hunting as the morally bad causes. Besides, since these morally bad causes were largely unknown to local participants, we chose to adopt charities with lower familiarity to participants for the current fMRI study to better control for familiarity difference between charities and bad causes (see **Table S5** and **Fig. S7b**).

To justify our selection, we compared the rating scores in all questions between the charities and morally bad causes used in the current fMRI study. Neither the charities nor the morally bad causes were familiar to participants, indicated by the low average scores of

familiarity (i.e., less than 2 on a 0-10 Likert scale; mean ± SD: charity: 1.3 ± 1.4; morally bad cause: 0.5 ± 0.8). As expected, participants donated much more (charity vs. bad cause [mean ± SD; same below]: 6.9 ± 2.2 vs. 0.6 ± 1.2; b (95% confidence interval (CI)) = 6.34 (5.76, 6.92), SE = 0.28, t(6) = 22.37, p < 0.001, Cohen'd = 18.3) and showed significantly higher levels of liking (7.1 ± 1.7 vs. -5.0 ± 2.9; b (95% CI) = 12.13 (10.83, 13.42), SE = 0.67, t(6) = 18.12, p < 0.001, Cohen'd = 14.8), moral acceptance (8.2 ± 1.6 vs. -3.7 ± 3.1; b (95% CI) = 11.84 (9.70, 13.98), SE = 1.11, t(6) = 10.70, p < 0.001, Cohen'd = 8.73), approval (8.0 ± 1.7 vs. -4.7 ± 3.1; b (95% CI) = 12.71 (11.21, 14.20), SE = 0.77, t(6) = 16.54, p < 0.001, Cohen'd = 13.50) and less moral conflict with the charities than bad causes (0.9 ± 1.7 vs. 6.6 ± 2.5; b (95% CI) = -5.70 (-6.41, -4.99), SE = 0.35, t(6) = -16.21, p < 0.001, Cohen'd = -13.23).

Vignettes and Logos

Charity

Charities familiar to local participants (CF)

International Committee of the Red Cross (ICRC; official website: https://www.icrc.org**)** was founded 1863 and is a humanitarian institution based in Geneva, Switzerland. It is one of the most widely recognized organizations in the world, having won three Nobel Peace Prizes in 1917, 1944, and 1963. ICRC aims to protect victims of international and internal armed conflicts. Such victims include war wounded, prisoners, refugees, civilians, and other non-combatants. (Information source: https://en.wikipedia.org/wiki/International_Committee_of_the_Red_Cross)

The United Nations Children's Fund (UNICEF; official website: https://www.unicef.org/**)** was founded in 1946 and is a United Nations (UN) program headquartered in New York City that provides humanitarian and developmental assistance to children and mothers in developing countries. UNICEF programs emphasize developing community-level services to promote the health and well-being of children. UNICEF was awarded the Nobel Peace Prize in 1965 and the Prince of Asturias Award of Concord in 2006.

(Information source: https://en.wikipedia.org/wiki/UNICEF)

SOS Children's Villages (official website: http://www.sos-childrensvillages.org/) was founded in 1949 and is an independent, non-governmental international development organization which has been working to meet the needs and protect the interests and rights of children. The organization's work focuses on abandoned, destitute and orphaned children requiring family-based child care. Children are supported to recover from being emotionally traumatized and to avoid the real danger of being isolated, abused, exploited and deprived of their rights.

(Information source: https://en.wikipedia.org/wiki/SOS_Children%27s_Villages)

World Wide Fund for Nature (WWF; official website: https://en.wikipedia.org/wiki/World_Wide_Fund_for_Nature**)** was founded in 1961 and is an international non-governmental organization, working in the field of the wilderness preservation, and the reduction of humanity's footprint on the environment. It is the world's largest conservation organization with over five million supporters worldwide. The group's mission is "to stop the degradation of the planet's natural environment and to build a future in which humans live in harmony with nature." Currently, much of its work concentrates on the conservation of oceans and coasts, forests, and freshwater ecosystems.

(Information source: https://en.wikipedia.org/wiki/World_Wide_Fund_for_Nature)

Charities unfamiliar to local participants (CUF; finally used in the pilot and the fMRI study)

Oxford Committee for Famine Relief (OXFAM; official website: https://www.oxfam.org/) was founded in 1942 in Oxford, and is an international confederation of charitable organizations focused on the alleviation of global poverty. OXFAM's programs address the structural causes of poverty and related injustice and work primarily through local accountable organizations, seeking to enhance their effectiveness. OXFAM's stated goal is to help people directly when local capacity is insufficient or inappropriate for OXFAM's purposes and to assist in the

development of structures which directly benefit people facing the realities of poverty and injustice. (Information source: https://en.wikipedia.org/wiki/Oxfam)

Save the Children (also known as the Save the Children Fund; official website: https://www.savethechildren.net/) was founded in 1919 in London, and is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries. Save the Children uses a holistic approach to help us achieve more for children, and to use resources in an efficient and sustainable way.

(Information source: https://en.wikipedia.org/wiki/Save_the_Children;
https://www.savethechildren.net/about-us)

First Aid Africa (FAA; official website: http://www.firstaidafrica.org/) was founded in 2008 in Edinburg and is a humanitarian charity that works in rural parts of southeastern Africa to provide sustainable equipment and education in first aid. FAA explains that a small amount of medical knowledge and equipment' can make a difference. Volunteers and students receive some training before traveling to Africa to teach first aid and survival skills in settings such as local communities, schools, orphanages, and villages.

(Information source: https://en.wikipedia.org/wiki/First_Aid_Africa)

Oceana (official website: http://oceana.org/) was found in 2001, Washington and is the largest international ocean conservation and advocacy organization. Oceana works to protect and restore the world's oceans through targeted policy campaigns. Oceana bases its policy campaign goals on science to achieve concrete and measurable results through targeted campaigns that combine policy, advocacy, science, law, media, and public pressure to prevent the collapse of fish populations, marine mammals and other sea life caused by industrial fishing and pollution.

(Information source: https://en.wikipedia.org/wiki/Oceana_(non-profit_group))

Morally Bad Cause

Drug Legalization Associations (D)

Law Enforcement Action Partnership (**LEAP**; formerly Law Enforcement Against Prohibition; official website : https://lawenforcementactionpartnership.org/) was founded in 2002 and is a non-profit, international, educational organization comprising former and current police officers, government agents and other law enforcement agents who oppose the current War on Drugs. In January 2017, while reaffirming their commitment to ending the War on Drugs, LEAP became the Law Enforcement Action Partnership in order to advocate for solutions across a broader range of drug policy and criminal justice issues.

(Information source: https://en.wikipedia.org/wiki/Law_Enforcement_Action_Partnership; https://www.dinafem.org/en/blog/cannabis-marijuana-legalization-groups/)

European Coalition for Just and Effective Drug Policies (ENCOD; official website: http://www.encod.org/info/-English-en-.html**)** was founded in 1993 and is a European non-governmental organization which brings European citizens together who believe that drug prohibition is an immoral and insane policy. Since 1994 they have been working to advocate more just and effective drugs control policies, which include an integrated solution for all problems related to the global drugs phenomenon.

(Information source: https://en.wikipedia.org/wiki/European_Coalition_for_Just_and_Effective_Drug_Policies; https://www.dinafem.org/en/blog/cannabis-marijuana-legalization-groups/)

Marijuana Policy Project (MPP; official website: https://www.mpp.org/**)** was founded in 1995 and is the largest organization working solely on marijuana policy reform in the United States. Its stated aims include: (1) increase public support for non-punitive, non-coercive marijuana policies and (2) change state laws to reduce or eliminate penalties for the medical and non-

medical use of marijuana. MPP believes the greatest harm associated with marijuana is a prison, so their focus is on removing criminal penalties for marijuana use.

(Information source: https://en.wikipedia.org/wiki/Marijuana_Policy_Project; https://www.dinafem.org/en/blog/cannabis-marijuana-legalization-groups/)

National Organization for the Reform of Marijuana Laws (NORML; official website: http://norml.org/**)** was founded in 1970 and is an American non-profit organization whose aim is to move public opinion sufficiently to achieve the legalization of non-medical marijuana in the United States so that the responsible use of cannabis by adults is no longer subject to penalty. NORML supports the removal of all criminal penalties for the private possession and responsible use of marijuana by adults, including the cultivation for personal use, and the casual nonprofit transfers of small amounts.

(Information source: https://en.wikipedia.org/wiki/National_Organization_for_the_Reform_of_Marijuana_Laws)

Morally Bad Cause

Gun/Hunting Rights Advocacy Associations (GH; finally used in the pilot and the fMRI study)

National Rifle Association of America (NRA; official website: https://home.nra.org/**)** is an American nonprofit organization which advocates for gun rights. Founded in 1871 to advance rifle marksmanship, the modern NRA continues to teach firearm competency and safety. NRA has been criticized by newspaper editorial boards, gun control and gun rights advocacy groups, political commentators, and politicians. For instance, a Washington Post/ABC News poll in January 2013 showed that only 36 percent of Americans had a favorable opinion of NRA leadership.

(Information source: https://en.wikipedia.org/wiki/National_Rifle_Association)

The European Federation of Associations for Hunting and Conservation of the EU (FACE; official website: http://www.face.eu/**)** was founded in 1977, and is a non-profit, non-

governmental organization which is pro-hunting and pro-gun. FACE protested a 2016 proposal by the European Union to revise the EU's firearm regulations, saying it would ban muzzle-loading weapons. In 2016 FACE opposed an EU proposal to ban the import of hunting trophies from certain African countries.

(Information source: https://en.wikipedia.org/wiki/Federation_of_Associations_for_Hunting_and_Conservation_of_the_EU)

The Society for Liberal Weapons Rights (ProTell; official website: https://www.protell.ch/de/**)** is a Swiss gun-rights advocacy group based in Bern, Switzerland. The association was founded in 1978 with the purpose of defending the right of law-abiding citizens to carry arms, and is opposed to any restrictions in this regard. ProTell was one of the principal opponents to the federal popular initiative "For the protection against gun violence", brought to a referendum on February 13, 2011. The initiative was broadly rejected by the voters.

(Information source: https://en.wikipedia.org/wiki/ProTell)

Safari Club International (SCI; official website: https://www.safariclub.org/**)** is an international organization composed of hunters dedicated to protecting the freedom to hunt and promoting wildlife conservation. SCI has been criticized for supporting the hunting of endangered African antelope species at fenced "game" ranches in Texas and Florida and for giving awards for hunting leopards, elephants, lions, rhinos and buffalo in Africa.

(Information source: https://en.wikipedia.org/wiki/Safari_Club_International)

SI: Pilot behavioral study

Methods

Thirty undergraduates or graduate students (15 females; mean age: 20.9 ± 1.7 years, ranging from 18 to 25 years) were recruited via online fliers for the pilot behavioral study.

The procedure and behavioral paradigm was same as the fMRI study, except the following: 1) besides the original payoff matrix (i.e., monetary gain: moral cost = 1:4), we adopted a balanced matrix with the payoffs ranging from 1 to 8 (in increments of 1; unit: CNY) for both parties involved in the dilemma (i.e., monetary gain: moral cost = 1:1), thereby producing 64 different payoff combinations as offers. As a consequence, the pilot study was extended to 256 trials in total (i.e., 64 trials for each context of moral dilemma displaying offers produced by different payoff matrices); 2) To maximize the differential effect induced by two types of payoff matrix, we adopted a mixed design so that offers from the same payoff matrix were presented in a block (i.e., one block for original and balanced matrix respectively) with the different dilemmas randomly intermixed in-between. The order of block was counterbalanced across participants; 3) To reduce the total duration of this experiment, we used a 500ms inter-trial interval (ITI) showing a cross fixation.

We analyzed the data with a similar approach as the fMRI study except that we added the predictors (i.e., main effect or interaction terms) relevant to matrix type (i.e., dummy variable; reference-level: original matrix) in the regression analyses.

Results

All these associations were selected by participants at least once (see **Fig. S8a**). Participants were familiar to none of the selected associations, indicated by the low average scores of familiarity (i.e., less than 3 on a 0-10 Likert scale; mean ± SD: charity: 2.80 ± 3.10; bad cause: 1.63 ± 2.27). Moreover, they rated the pre-selected charity positively (mean ± SD

(95% CI): 7.83 ± 2.00 (7.09, 8.58); t(29) = 21.44, p < 0.001, Cohen's d = 3.91) whereas they regarded the pre-selected bad cause negatively (mean ± SD (95% CI); -7.87 ± 2.54 (-8.82, -6.92); t(29) = -16.95, p < 0.001, Cohen's d = 3.09; see **Fig. S8b**).

We found a significant matrix × context interaction effect on the choice data (Odds Ratio = 4.33, b (95% CI) = 1.46 (1.19, 1.74), SE = 0.14, $\chi^2(1) = 106.63$, p < 0.001) after controlling the monetary gain and the moral cost. Splitting the data in terms of the matrix type, we found that participants were less likely to accept the offer in the *charity-bad_cause* dilemma (vs. *self-bad_cause* dilemma: accept rate: 26.4 ± 25.3 % vs. 40.6 ± 37.4 %; Odds Ratio = 0.24, b (95% CI) = -1.44 (-1.66, -1.22), SE = 0.11, $\chi^2(1) = 170.77$, p < 0.001) only when the original payoff matrix was adopted. No difference in acceptance rate between two contexts of moral dilemma was observed when the balanced payoff matrix was used (62.1 ± 21.8 % vs. 62.3 ± 28.5 %; Odds Ratio = 0.98, b (95% CI) = -0.02 (-0.25, 0.21), SE = 0.12, $\chi^2(1) = 0.03$, p = 0.861; see **Fig. S8c**; also see **Table S6** for details of model output).

For the decision time (DT), we first did the log-transformation due to its non-normal distribution (Anderson-Darling normality test: A = 219.3, p < 0.001). Regressions on log-transformed DT (logDT) revealed a three-way interaction (b (95% CI) = 0.10 (0.03, 0.16), SE = 0.03, t(7643) = 2.93, p = 0.003, Cohen's d = 0.07; see **Table S7** for descriptive summary of DT and logDT). To unpack the interaction effect, we ran similar regression analyses with matrix, context and their interaction as fixed-effect predictors for trials with acceptance and rejection decisions separately. For acceptance trials, we observed a significant two-way interaction on logDT (b (95% CI) = -0.12 (-0.16, -0.07), SE = 0.02, t(3644) = -4.80, p < 0.001, Cohen'd = -0.16). Again, we did the simple effect analyses by splitting the accept trials into two parts depending on matrix type, finding that such two-way interaction was mainly driven by the stronger effect of dilemma type in prolonging logDT in the *charity-bad_cause* dilemma (vs. *self-bad_cause* dilemma) when offers were chosen from the original matrix (b (95% CI) = 0.15 (0.11,

0.19), SE = 0.02, t(1265) = 7.41, p < 0.001, Cohen'd = 0.42) than the balanced matrix (b (95% CI) = 0.04 (0.01, 0.06), SE = 0.01, t(2359) = 2.74, p = 0.006, Cohen'd = 0.21). For reject trials, no two-way interaction effect was detected (b (95% CI) = 0.02 (-0.02, 0.06), SE = 0.02, t = 0.970, p = 0.332, Cohen'd = 0.03). We only found that people rejected more slowly in the block displaying offers from balanced (vs. original) payoff regardless of dilemmas (b (95% CI) = 0.07 (0.04, 0.11), SE = 0.02, t(3970) = 3.97, p < 0.001, Cohen'd = 0.13; see **Fig. S9**; also see **Table S8** for details of model output).

Behavioral Analyses

All behavioral analyses were conducted using R (http://www.r-project.org/) and relevant packages (R Core Team, 2014). All reported p values are two-tailed and $p < 0.05$ was considered statistically significant. Data visualization were performed via "ggplot2" package (Wickham H 2016).

Regarding the choice data, we performed a repeated mixed-effect logistic regression on the decision of choosing the "accept" option by the glmer function in "lme4" package (Bates D et al. 2013), with dilemma (dummy variable; reference level: *self-bad_cause* dilemma; same below) and payoffs for both parties involved in each dilemma (i.e., the monetary gain and the moral cost; mean-centered continuous variable; same below) as the fixed-effect predictors. In addition, we included the following random-effect factors allowing varying intercept across participants. For the statistical inference on each predictor, we performed the Type II Wald chi-square test on the model fits by using the Anova function in "car" package (Fox J et al. 2016), and reported the odds ratio as relevant effect size.

For decision time (DT), we first did a log-transformation due to its non-normal distribution (Anderson-Darling normality test: $A = 91.90$, $p < 0.001$) and then performed a mixed-effect linear regression on the log-transformed DT by the lmer function in "lme4" package, with decision (dummy variable; reference level: accept), dilemma, decision × dilemma, as well as payoffs for both beneficiaries as the fixed-effect predictors. Random-effect factors were specified in the same way as above. Similar analyses were also performed on the post-scanning rating except that dilemma was added as the only fixed-effect predictor. We followed the procedure recommended by Luke (2017) to obtain the statistics for each predictor by applying the Satterthwaite approximations on the restricted maximum likelihood model (REML) fit via the "lmerTest" package (Luke SG 2017). In addition, we computed the Cohen'd of each predictor via the "EMAtools" package (Kleiman E 2017), which provided the effect size measure

specially for the mixed-effect regressions. For likeness ratings of the selected associations, we compared whether the ratings significantly differed from 0 in each type of selected associations (i.e., charity or morally bad causes) respectively by the one-sample T-test, and computed the Cohen's d as effect size.

Supporting Figures

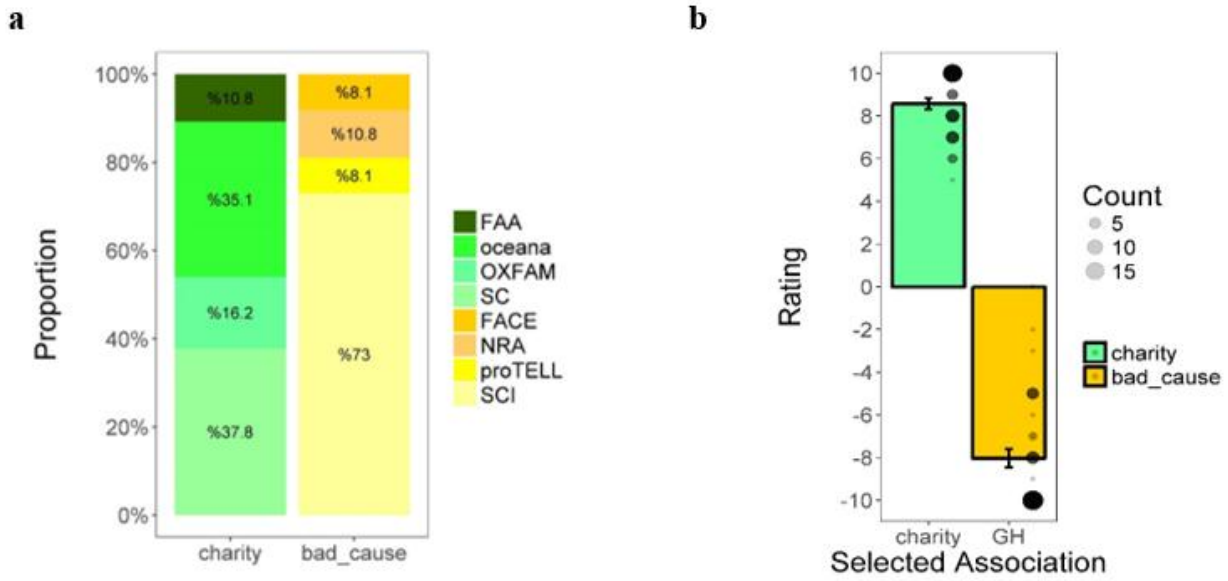


Fig. S1. (a) Summary for the charity and morally bad cause selected by all participants and (b) mean rating on liking (-10 – 10; -10 = dislike very much, 10 = like very much) for the selected charities and morally bad causes in the current fMRI study. The black dot refers to the individual ratings; the size of the dot indicates the numbers of participants with the same rating.

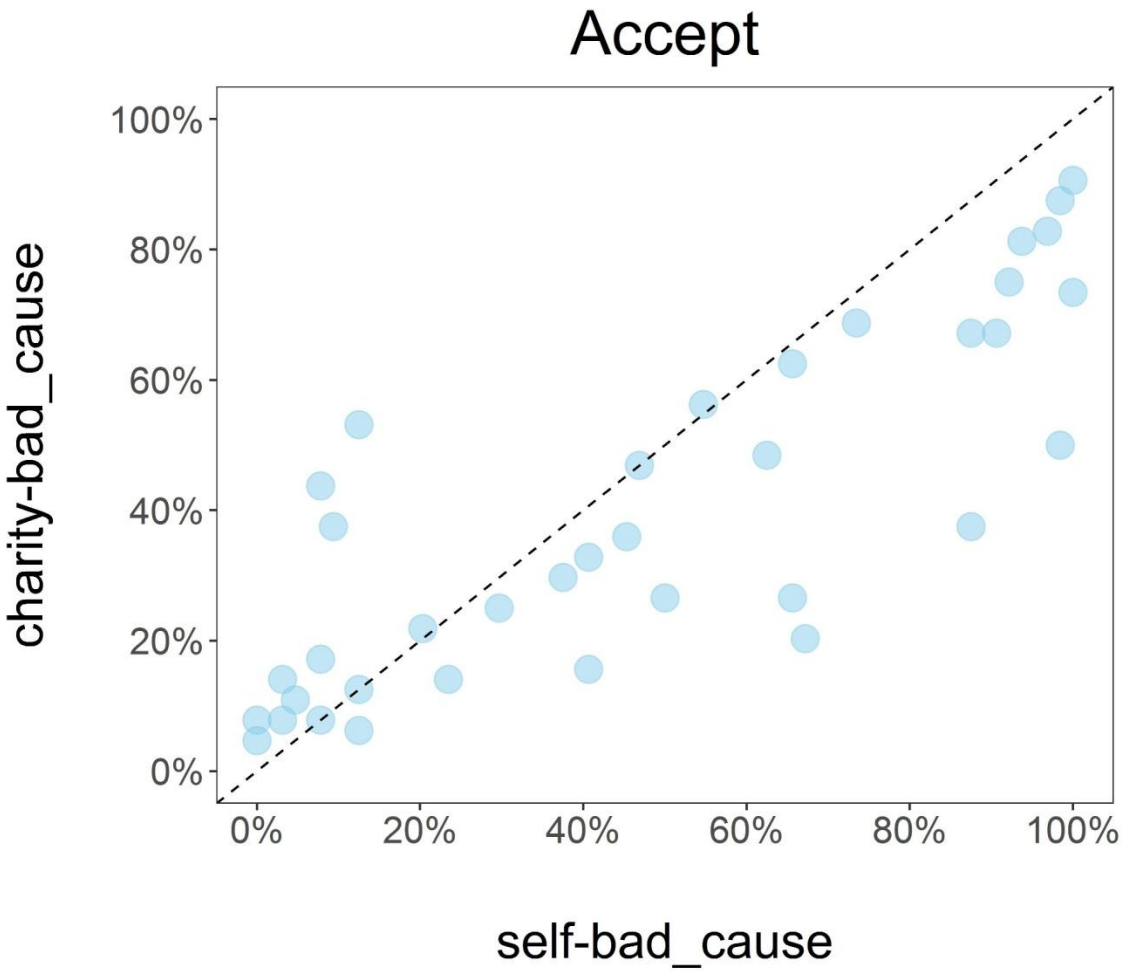
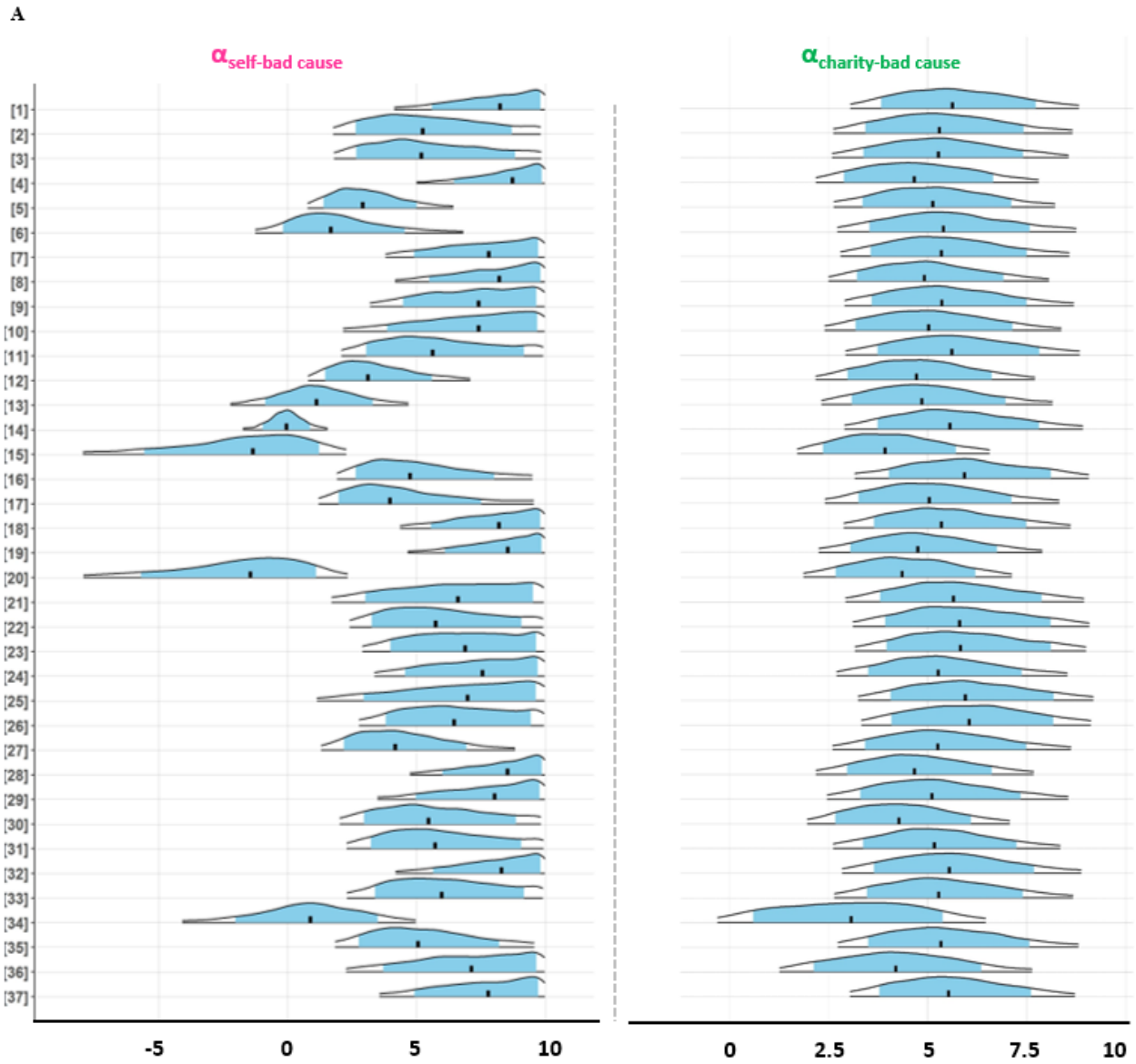


Fig. S2. Correlation of acceptance rate between two dilemmas. Each dot refers to the acceptance rate of a single participant. Dots below the dotted diagonal indicates participants who accepts more immoral offers in the *self-bad_cause* dilemma than in the *charity-bad_cause* dilemma.

297


A
αself-bad cause
αcharity-bad cause
[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]
[36]
[37]
-5
0
5
10
0
2.5
5
7.5
10

298
299

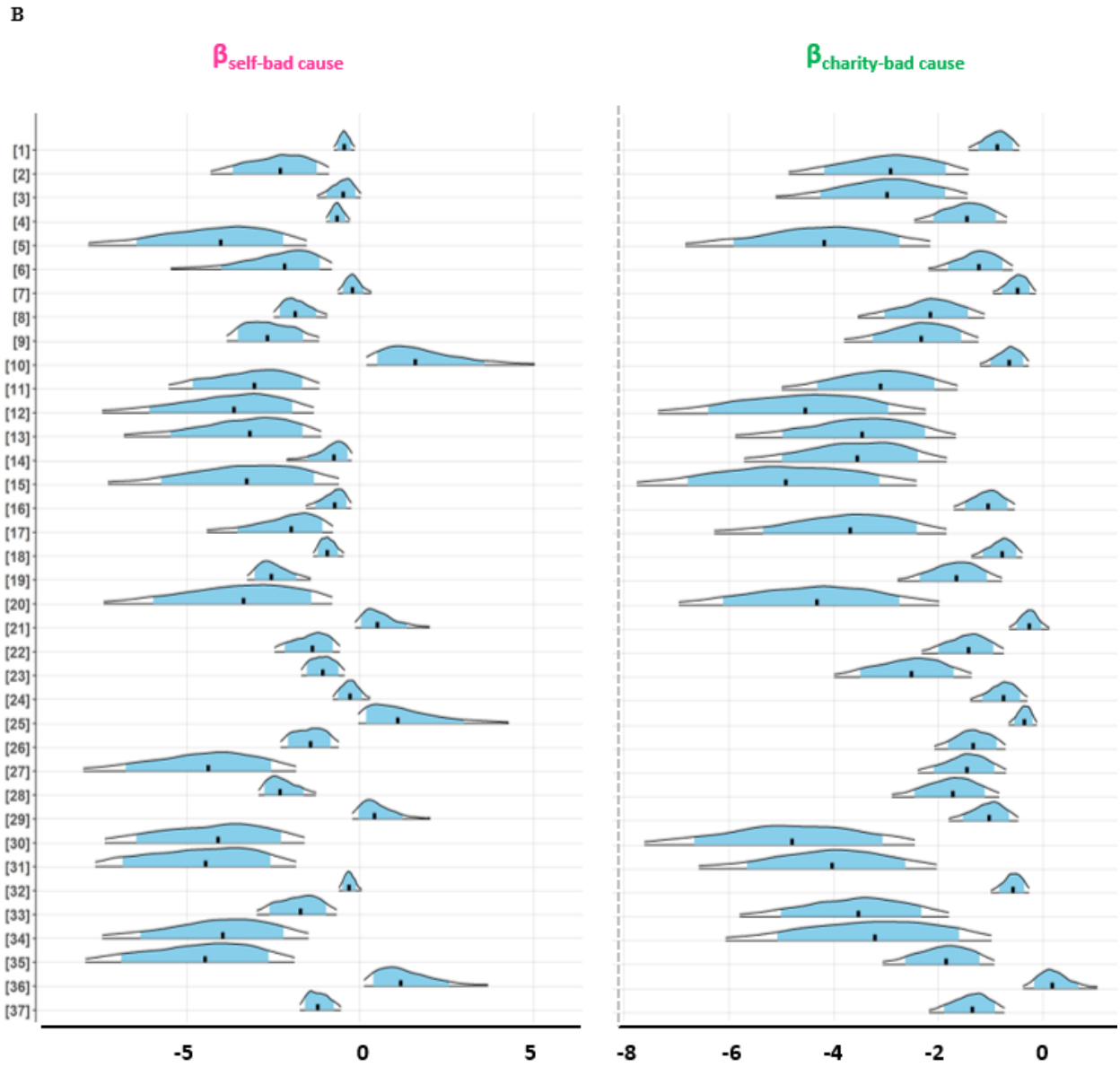


Figure S3 Posterior distributions of individual-level A) α and B) β. Each distribution indicates the posterior distribution of each parameter of a single participant in each condition. The sky blue shading and tailed white areas represent the 80% and 95% kernel density estimates respectively. The black dots represent the mean of the posterior distribution.

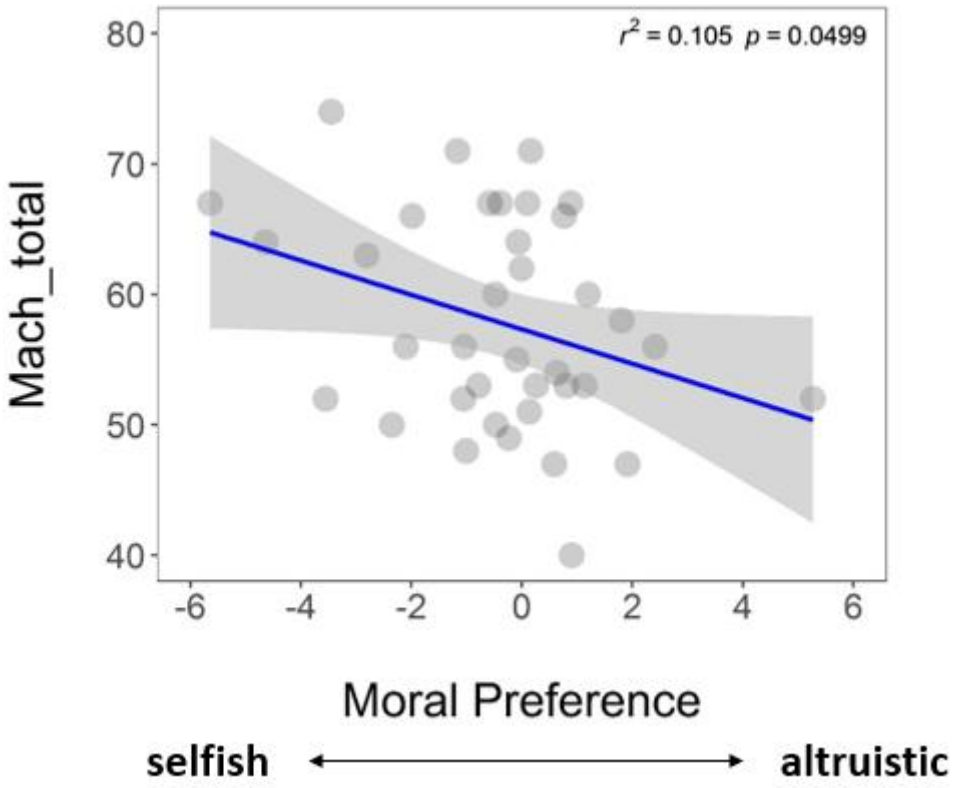


Fig. S4. Negative correlation between the moral preference and the Machiavellian score. The higher the Machiavellian score is, the higher degree that a person agrees with the idea of pursuing personal gain via immoral approaches. Each dot represents the data of a single participant.

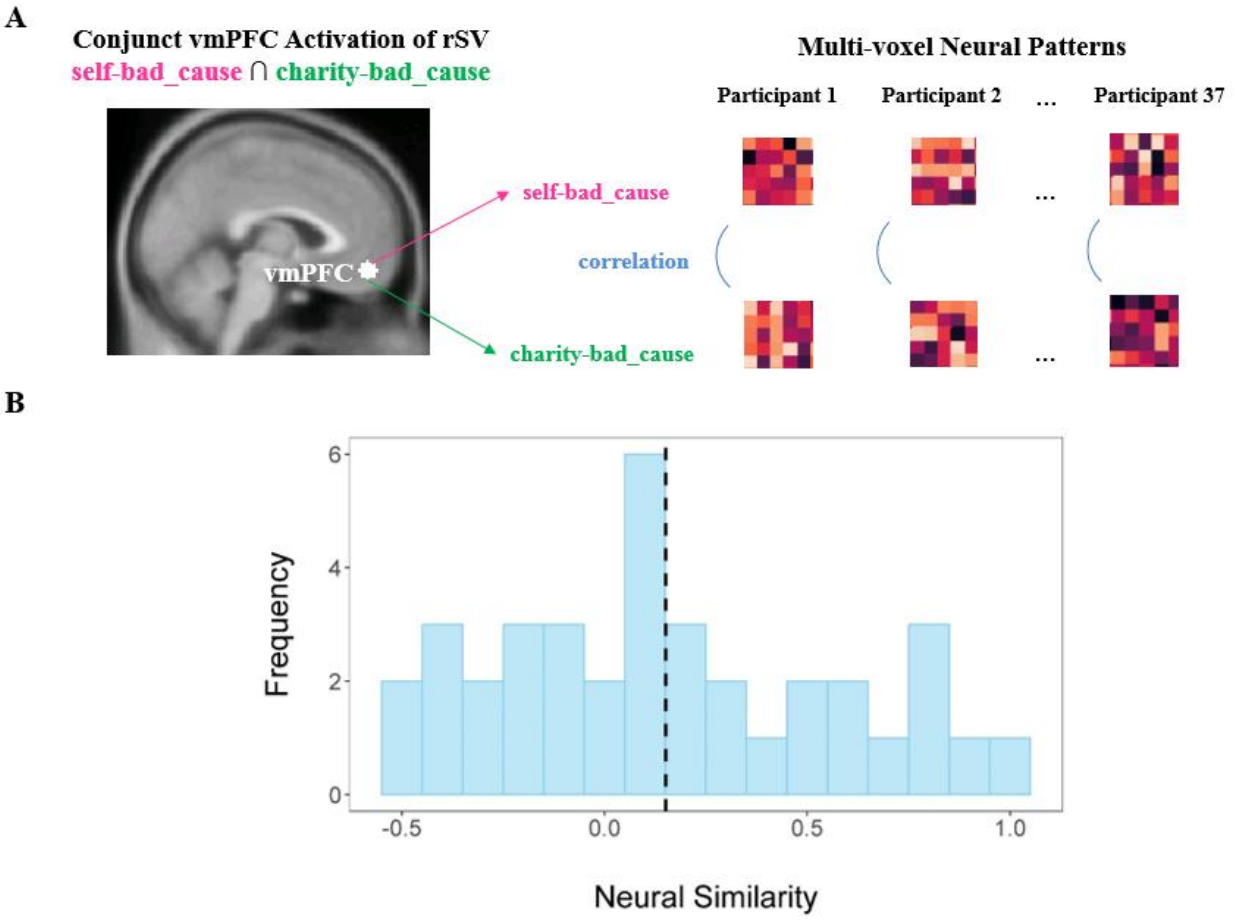


Fig. S5 Procedure and results of representational similarity analysis (RSA). (a) For each participant, we first defined the vmPFC mask based on the conjunction activation in GLM1 (peak MNI: -2/48/-14; a sphere with a radius of 6mm). Then we extracted the multi-voxel neural patterns (i.e., those heat maps; only for illustration) within the vmPFC mask from the contrast image characterizing the parametric effect of relative subjective value (SV) in each dilemma. Next we computed the dissimilarity between these neural patterns in two dilemmas and obtained the correlation coefficients (i.e., similarity) using one minus dissimilarity. For statistical analysis, all correlation coefficients were transformed to Fisher's z value. (b) Histogram of the distribution of neural similarity between across all participants. The vertical dashed line refers to the mean of the neural similarity.

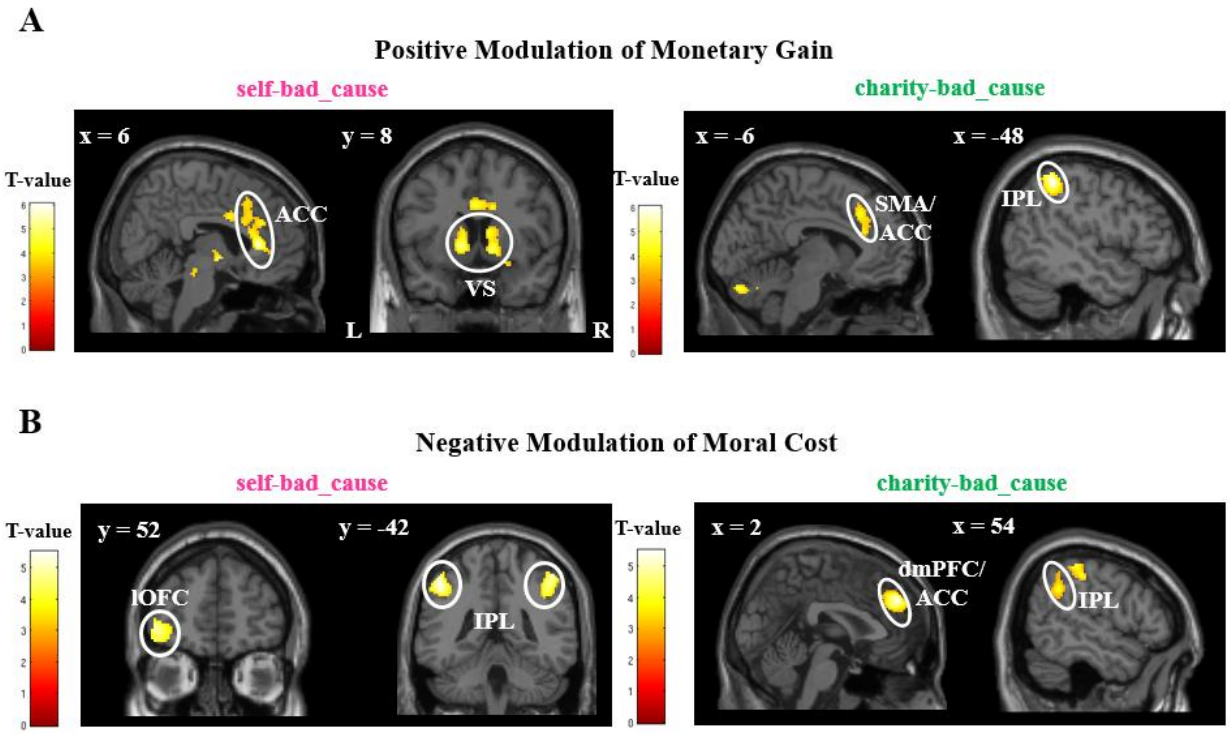


Fig. S6. Neural correlates of single attributes. (a) Positive modulation of the monetary gain (i.e., payoff for oneself and the charity) in each dilemma. (b) Negative modulation of the moral cost (i.e., benefits for the bad cause) in each dilemma. Abbreviation: ACC = anterior cingulate cortex; dmPFC = dorsomedial prefrontal cortex; SMA = supplementary motor area; IPL = inferior parietal lobule; lOFC = lateral orbitofrontal cortex; VS = ventral striatum; Display threshold: $p < 0.001$ uncorrected at the voxel-level with $k = 200$.

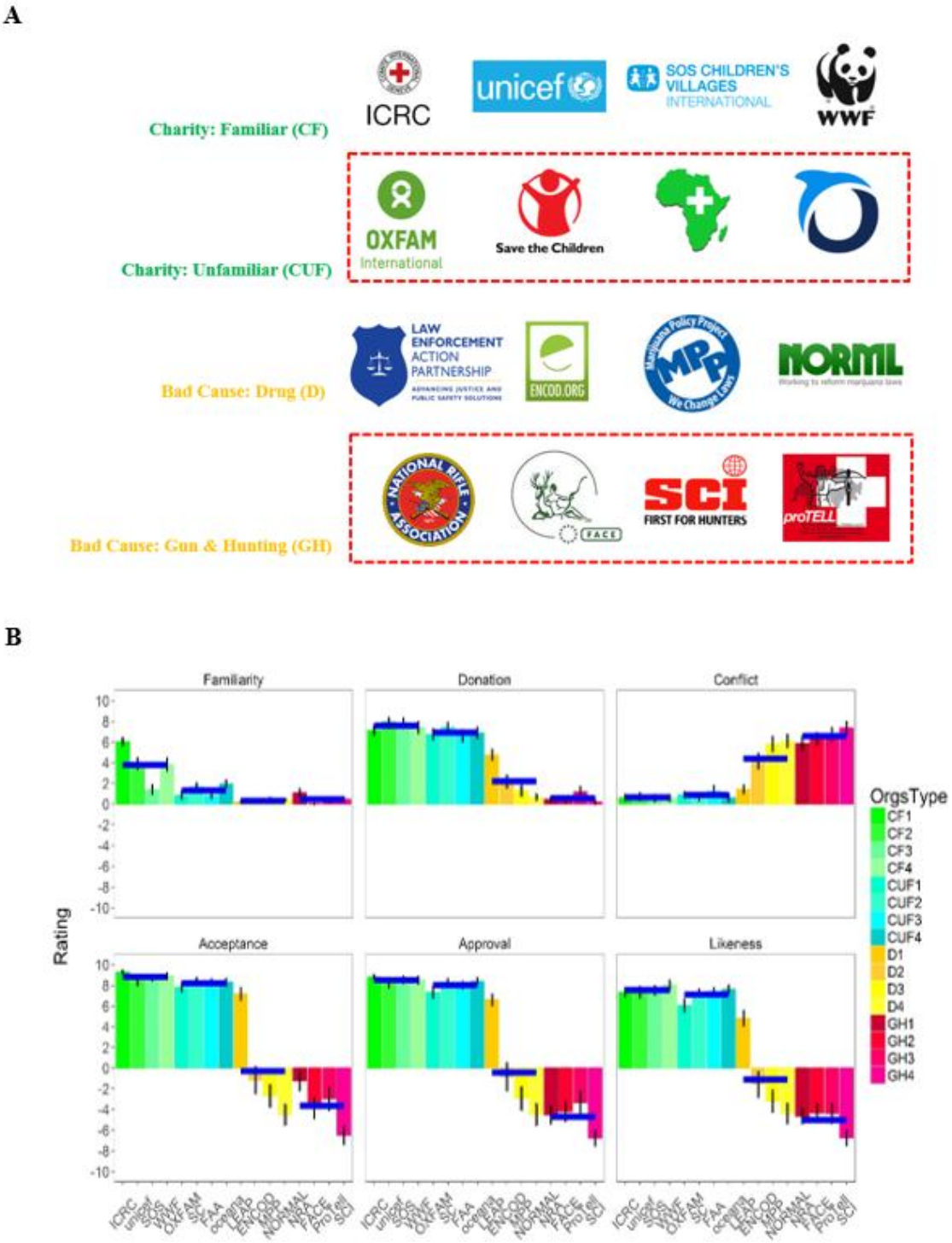


Fig. S7. (a) All logos of associations used for selection. Those charities and morally bad causes are marked by the red frame we finally used for the pilot. (b) Mean ratings on individual associations as well as categories according to six dimensions in the questionnaire.

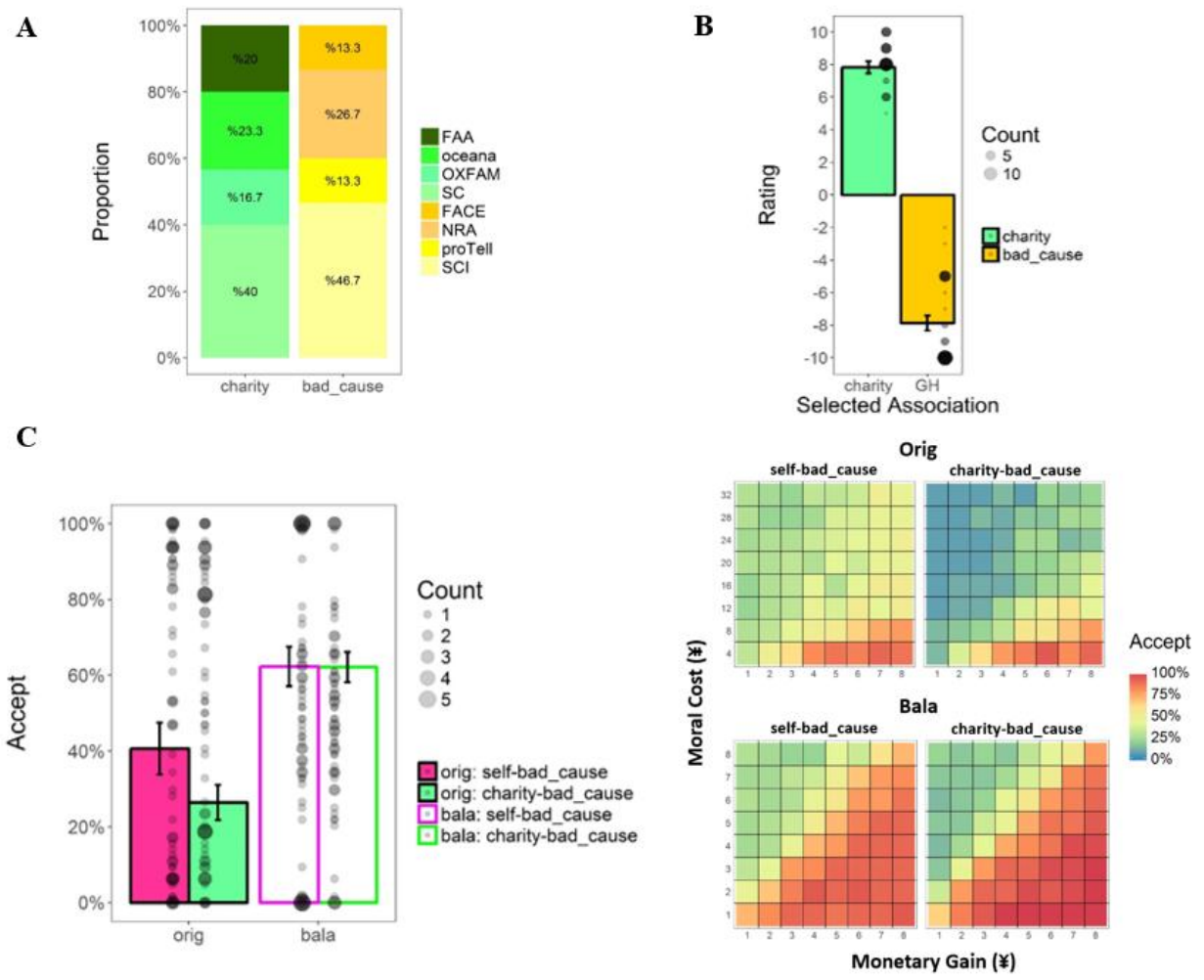


Fig. S8. Results of the pilot behavioral study. (a) Summary for the charity and the morally bad cause selected by all participants. (b) Mean rating on liking (-10 – 10; -10 = dislike very much, 10 = like very much) for the selected charity and morally bad cause. The black dot refers to the individual ratings; the size of the dot indicates the numbers of participants with the same rating. (c) Left panel: the mean acceptance rate in each dilemma with each payoff matrix; Right panel: the heat map of the mean acceptance rate (%) at each payoff amount for each beneficiary involved in both dilemmas. Abbreviations: orig= original payoff matrix (i.e., monetary gain: moral cost = 1:4); bala = balanced payoff matrix (i.e., monetary gain: moral cost = 1:1).

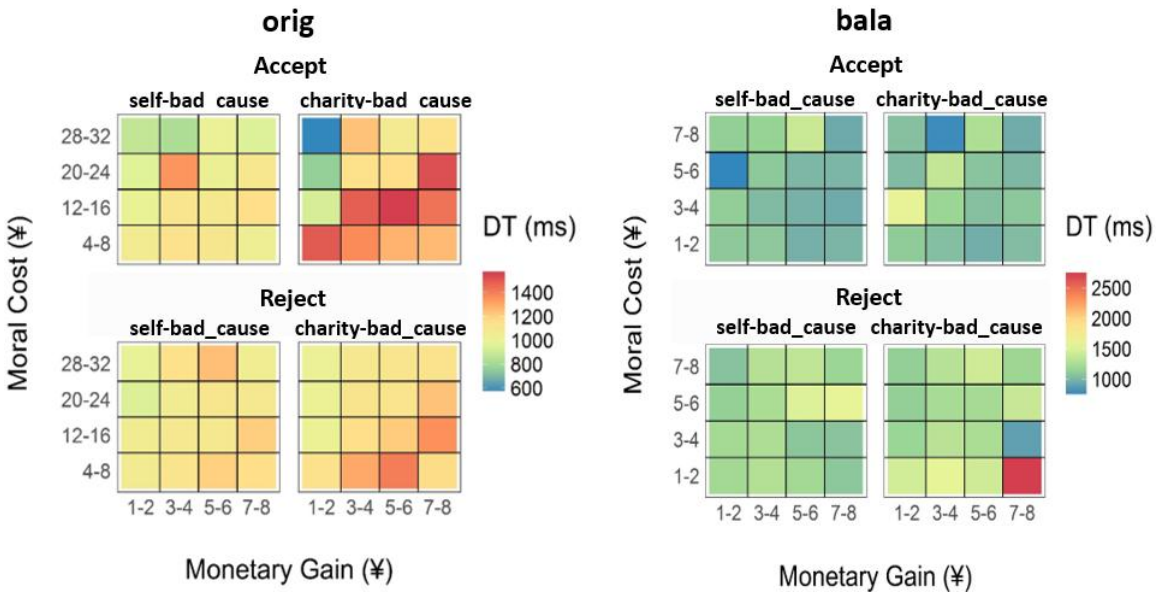


Fig. S9. Heat map of the mean decision time (DT) for specific choice as a function of the monetary gain and the moral cost in each dilemma, separated by different payoff matrix, in the pilot behavioral study. Data were collapsed into 4-by-4 matrices only for a better visualization. Abbreviations: orig= original payoff matrix (i.e., monetary gain: moral cost = 1:4); bala = balanced payoff matrix (i.e., monetary gain: moral cost = 1:1).

Supporting Tables

Table S1. Bayesian model evidence

Model	WAIC
M1	2892.081
M2	2418.411
M3	2885.438
M4*	2368.419
M5	2958.844
M6	2535.073
M7	3023.443
M8	2609.564

Note: Lower Watanabe-Akaike Information Criterion (WAIC) scores indicate better models.
*refers to the winning model (M4).

Table S2. Regions encoding the relative subjective value (rSV) during decision-making process in each dilemma (N = 37, GLM1)

Brain Region	Hemisphere	Cluster Size	MNI			BA	T-value	p(cl-FWE)
			x	y	z			
self-bad_cause								
rSV: positive modulation								
PCG/MFG	L	258	-28	-8	46	6	4.49	0.046
PCG/MFG	R	282	30	-4	42	6	4.67	0.034
FuG/LG/PHG/IOG	L	671	-36	-56	-12	19/37	5.42	0.001
FuG/LG/PHG/ MOG/IOG/MTG	R	1652	26	-70	2	18/19/20/ 36/37	5.47	< 0.001

Note: Regions shown here met the Family-Wise Error corrected cluster-level (cl-FWE) threshold of $p < 0.05$, with an uncorrected voxel-level threshold of $p < 0.001$ as the cluster-defining threshold. No significant region was found in positive or negative modulation of rSV in the *charity-bad_cause* dilemma under this threshold. Coordinates shown here were based on Montreal Neurological Institute (MNI) coordinate system. Abbreviations: L: left, R: right, B: bilateral, BA: Brodmann Area; MFG: middle frontal gyrus, PCG: pre-central gyrus, MTG: middle temporal gyrus, MOG: middle occipital gyrus, IOG: inferior occipital gyrus, FG: fusiform gyrus, LG: lingual gyrus, PHG: parahippocampal gyrus.

Table S3. Regions showing enhanced functional connectivity with vmPFC in different dilemmas (i.e., *PPI: self-bad_cause* vs. *implicit baseline*; *PPI: charity-bad_cause* vs. *implicit baseline*; N = 37, gPPI-GLM)

Regions	Hemisphere	Cluster Size	MNI			BA	T-value	p(cl-FWE)
			x	y	z			
PPI: *self-bad_cause* vs. *implicit baseline*								
vmPFC/MeFG/dmPFC/ ACC	B	2022	-2	38	-12	9/10/11/ 24/32	11.76	< 0.001
SFG	R	53	22	34	46	8	6.20	< 0.001
PCC/Prec	B	757	-4	-62	30	7/23/30/31	8.30	< 0.001
MTG/ITG	L	60	-56	-12	-20	21	7.06	< 0.001
ITG	R	150	60	-8	-20	21	7.77	< 0.001
IFG/STG	L	67	-32	16	-22	38/47	7.69	< 0.001
AG/SmG/STG/MTG	L	304	-46	-66	28	39	7.41	< 0.001
AG/SmG/STG/MTG	R	331	60	-52	28	39/40	7.86	< 0.001
Cerebellum	L	57	-10	-52	-36		6.53	< 0.001
PPI: *charity-bad_cause* vs. *implicit baseline*								
vmPFC/ACC	B	766	-4	50	-12	10/11/32	11.40	< 0.001
PCC/Prec		310	-8	-58	10	23/30	6.82	< 0.001

Note: A FWE-corrected voxel-level threshold of $p < 0.05$ with $k = 50$ was adopted. Coordinates shown here were based on Montreal Neurological Institute (MNI) coordinate system. Abbreviations: p(cl-FWE): cluster-level Family-Wise Error corrected threshold; L: left, R: right, B: bilateral, BA: Brodmann Area; ACC: anterior cingulate cortex, AG: angular gyrus, dmPFC: dorsomedial prefrontal cortex, MeFG: medial frontal gyrus, MTG, middle temporal gyrus, SFG: superior frontal gyrus, ITG: inferior temporal gyrus, SmG: supramarginal gyrus, PCC: posterior cingulate cortex, STG: superior temporal gyrus, vmPFC: ventromedial prefrontal cortext, Prec: precuneus.

Table S4. Regions encoding single attributes (i.e., monetary gain and moral cost) during decision-making process in each dilemma (N = 37, GLM2)

Brain Region	Hemisphere	Cluster Size	MNI			BA	T-value	p(cl-FWE)
			x	y	z			
self-bad_cause								
monetary gain: positive modulation								
VS (Caud/Put/Pal)/Tha	L	471	-12	0	-4		6.06	0.005
VS (Caud/Put/Pal)/IFG	R	574	14	4	-4	47	5.03	0.002
ACC/MCC/SMA/ MeFG/SFG	B	1277	6	32	8	8/24/32	5.55	< 0.001
moral cost: negative modulation								
IFG/MFG	L	413	-54	40	-4	10/11/47	5.09	0.004
MFG/IFG	R	451	36	26	40	8/9	4.95	0.003
IPL/AG/SmG/PoCG	L	607	-48	-42	46	40	5.51	< 0.001
IPL/SPL/AG/SmG/ STG/PoCG	R	1572	48	-46	56	7/39/40	5.04	< 0.001
charity-bad_cause								
monetary gain: positive modulation								
MeFG/SMA/ACC/MCC	B	445	-10	30	38	6/8/9/32	4.45	0.001
IPL	R	300	-48	-42	50	40	6.11	0.010
Cerebellum	L	349	-32	-68	-30		5.05	0.005
Cerebellum	R	675	36	-58	-32		5.00	< 0.001
moral cost: negative modulation								
MeFG/SFG/ACC/MCC	B	1037	2	46	34	6/8/9/32	6.45	< 0.001
SFG/MeFG/MFG	R	330	12	38	52	8	4.67	0.013
IPL/SmG/PoCG/AG	R	439	58	-26	44	40	5.46	0.003

Note: Regions shown here met the Family-Wise Error corrected cluster-level (cl-FWE) threshold of $p < 0.05$, with an uncorrected voxel-level threshold of $p < 0.001$ as the cluster-defining threshold. No significant region was found in the negative modulation of monetary gain and the positive modulation of moral cost, in each dilemma respectively. Coordinates shown here were based on Montreal Neurological Institute (MNI) coordinate system. Abbreviations: L: left, R: right, B: bilateral, BA: Brodmann Area; ACC: anterior cingulate cortex, MCC; mid-cingulate cortex, MFG: middle frontal gyrus, MeFG: medial frontal gyrus, IFG: inferior frontal gyrus, SFG: superior frontal gyrus, SMA: supplementary motor area, AG: angular gyrus, PoCG: postcentral gyrus, SmG: supramarginal gyrus, IPL: inferior parietal lobule, STG: superior temporal gyrus, VS: ventral striatum; Caud: caudate, Put: putamen, Pal: pallidum.

Table S5. Summary of average rating (mean ± SD) in the rating task for association selection (N = 30)

		Familiarity	Likeness	Acceptance	Approval	Donation	Conflict
CF	ICRC	6.0 ± 2.2	7.3 ± 2.3	9.3 ± 1.0	8.8 ± 1.2	7.1 ± 2.6	0.6 ± 1.4
	unicef	3.9 ± 3.0	7.3 ± 2.8	8.4 ± 2.4	8.2 ± 2.6	8.0 ± 2.1	0.7 ± 1.9
	SOS	1.4 ± 2.4	7.5 ± 2.4	8.7 ± 1.5	8.5 ± 1.8	7.9 ± 2.4	0.6 ± 1.9
	WWF	3.8 ± 3.3	8.0 ± 2.5	8.9 ± 1.5	8.5 ± 2.0	7.4 ± 3.1	0.6 ± 1.8
	pooled	3.8 ± 2.0	7.6 ± 1.8	8.8 ± 1.3	8.5 ± 1.5	7.6 ± 2.2	0.6 ± 1.6
CUF	**OXFAM**	**0.9 ± 1.9**	**6.1 ± 3.3**	**7.8 ± 2.4**	**7.3 ± 3.0**	**6.7 ± 2.8**	**0.9 ± 1.6**
	SC	**1.7 ± 2.0**	**7.3 ± 2.0**	**8.4 ± 2.0**	**8.3 ± 2.0**	**7.4 ± 2.4**	**0.7 ± 2.0**
	FAA	**0.8 ± 1.7**	**7.4 ± 1.6**	**8.3 ± 1.7**	**8.0 ± 1.9**	**6.6 ± 3.0**	**1.3 ± 2.8**
	Oceana	**2.0 ± 2.1**	**7.6 ± 2.2**	**8.3 ± 1.9**	**8.4 ± 1.8**	**6.9 ± 2.9**	**0.6 ± 2.0**
	pooled	**1.3 ± 1.4**	**7.1 ± 1.7**	**8.2 ± 1.6**	**8.0 ± 1.7**	**6.9 ± 2.2**	**0.9 ± 1.7**
D	LEAP	0.2 ± 0.6	4.8 ± 4.2	7.2 ± 3.1	6.6 ± 3.0	4.8 ± 2.7	1.4 ± 2.3
	ENCOD	0.3 ± 0.8	-1.6 ± 6.7	-1.2 ± 6.9	-0.8 ± 7.2	2.2 ± 3.3	4.2 ± 4.0
	MPP	0.4 ± 1.5	-3.2 ± 5.7	-2.7 ± 5.9	-2.9 ± 6.1	1.3 ± 2.9	5.8 ± 3.8
	NORML	0.4 ± 0.9	-4.4 ± 5.4	-4.5 ± 5.6	-4.5 ± 5.7	0.7 ± 1.7	6.1 ± 3.5
	pooled	0.3 ± 0.7	-1.1 ± 3.7	-0.3 ± 3.6	-0.4 ± 3.8	2.2 ± 2.0	4.4 ± 2.3
GH	**NRA**	**1.1 ± 2.1**	**-4.7 ± 4.0**	**-1.2 ± 5.0**	**-4.5 ± 4.7**	**0.5 ± 1.5**	**5.9 ± 3.5**
	FACE	**0.0 ± 0.2**	**-4.3 ± 4.6**	**-3.9 ± 5.2**	**-4.2 ± 5.2**	**0.4 ± 1.5**	**6.3 ± 3.2**
	proTELL	**0.3 ± 0.9**	**-4.4 ± 5.2**	**-3.0 ± 5.9**	**-3.3 ± 6.0**	**1.2 ± 2.7**	**6.8 ± 3.4**
	SCI	**0.5 ± 1.2**	**-6.7 ± 4.0**	**-6.5 ± 4.7**	**-6.8 ± 4.1**	**0.2 ± 1.1**	**7.4 ± 3.4**
	pooled	**0.5 ± 0.8**	**-5.0 ± 2.9**	**-3.7 ± 3.1**	**-4.7 ± 3.1**	**0.6 ± 1.2**	**6.6 ± 2.5**

Note: Pooled results refer to the average rating by pooling 4 associations within each type of associations (i.e., charity unfamiliar, CUF; charity familiar, CF; drug legalization group, D; gun/hunting rights advocacy group, GH). Bold texts indicate ratings for those associations finally selected for the pilot behavioral study and the current fMRI study. See the above section “vignettes used in the pilot rating task” for full names of all these associations.

Table S6. Output of fixed effects in mixed-effect logistic regressions predicting accept decisions in the pilot behavioral study

	All trials *b* (SE)	Trials in the block with original payoff matrix *b* (SE)	Trials in the block with balanced payoff matrix *b* (SE)
Intercept	0.59 (0.48)	0.29 (0.51)	-4.69 (0.69)***
Matrix	-0.54 (0.12)***		
Dilemma	-1.48 (0.11)***	-1.44 (0.11)***	-0.02 (0.12)
Matrix × Dilemma	1.46 (0.14)***		
Monetary Gain[a]	0.60 (0.02)***	0.49 (0.03)***	1.09 (0.04)***
Moral Cost[b]	-0.18 (0.01)***	-0.14 (0.01)***	-0.96 (0.04)***
AIC	5499.1	2532.6	2047.9
BIC	5547.7	2563.9	2079.1
Numbers of Observation	7680	3840	3840
N (Participant)	30	30	30

Note: [a]Monetary Gain: payoffs for participants (*self-bad_cause* dilemma) or the pre-selected charity (charity-bad_cause dilemma); grand mean-centered before putting into the regression model; [b]Moral Cost: benefits for the pre-selected bad cause (in both dilemma); grand mean-centered before putting into the regression model.

Reference levels were set as follows: Matrix = original payoff matrix (i.e., monetary gain: moral cost = 1:4), Dilemma = *self-bad_cause* dilemma. Table also shows goodness-of-fit statistics: AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: ***$p < 0.001$.

Table S7. Descriptive summary for the decision time (DT) and log-transformed DT (logDT; in ms) in the pilot behavioral study

		DT				logDT			
		Original		Balanced		Original		Balanced	
		self-bad_cause	charity-bad_cause	self-bad_cause	charity-bad_cause	self-bad_cause	charity-bad_cause	self-bad_cause	charity-bad_cause
Accept	Mean	1143.6	1351.7	1067.8	1094.7	6.96	7.12	6.86	6.89
	(SD)	(355.7)	(350.6)	(400.0)	(363.8)	(0.33)	(0.32)	(0.36)	(0.34)
	N	28	28	28	30	28	28	28	30
Reject	Mean	1143.3	1134.4	1248.5	1268.7	6.95	6.95	7.05	7.06
	(SD)	(327.8)	(295.2)	(311.3)	(351.3)	(0.27)	(0.27)	(0.25)	(0.27)
	N	27	28	25	27	27	28	25	27

Note: we first calculated the individual-level mean DT and logDT in terms of specific decisions for each context of moral dilemma in each type of payoff matrix, then we calculated the group-level mean (± SD) based on the individual mean; Due to individual difference in decisions, the sample size (i.e., N) in original payoff matrix (i.e., monetary gain: moral cost = 1:4) and balanced payoff matrix (i.e., monetary gain: moral cost = 1:1) for each dilemma context of specific decisions is different.

Table S8. Output of fixed effects in mixed-effect linear regressions predicting logDT in the pilot behavioral study

	All trials	Accept trials	Accept trials (original matrix)	Accept trials (balanced matrix)	Reject trials
	b (SE)	*b* (SE)	*b* (SE)	*b* (SE)	*b* (SE)
Intercept	6.87 (0.05)***	6.97 (0.06)***	6.95 (0.06)***	7.00 (0.07)***	7.01 (0.05)***
Decision	-0.003 (0.02)				
Matrix	-0.09 (0.02)***	-0.09 (0.02)***			0.07 (0.02)***
Dilemma	0.16 (0.02)***	0.15 (0.02)***	0.15 (0.02)***	0.04 (0.01)**	0.01 (0.01)
Decision × Matrix	0.18 (0.02)***				
Decision × Dilemma	-0.10 (0.02)***				
Matrix × Dilemma	-0.12 (0.02)***	-0.12 (0.02)***			0.02 (0.02)
Decision × Matrix × Dilemma	0.10 (0.03)**				
Monetary Gain[a]	0.007 (0.002)***	-0.02 (0.003)***	-0.01 (0.005)**	-0.02 (0.003)***	0.02 (0.002)***
Moral Cost[b]	-0.003 (0.001)***	0.001 (0.001)	0.001 (0.001)	0.02 (0.003)***	-0.005 (0.001)***
AIC	5534.5	2658.8	929.2	1419.4	2169.5
BIC	5617.9	2708.5	960.2	1454.0	2219.8
Numbers of Observation	7680	3676	1287	2389	4004
N (Participant)	30	30	29	30	29

Note: [a]Monetary Gain: payoff for participants (*self-bad_cause* dilemma) or the pre-selected charity (*charity-bad*_cause dilemma); grand mean-centered before putting into the regression model; [b]Moral Cost: benefits for the pre-selected bad cause (in both dilemmas); grand mean-centered before putting into the regression model.

Reference levels were set as follows: Decision = accept, Matrix = original payoff matrix (i.e., monetary gain: moral cost = 1:4), Dilemma = *self-bad_cause* dilemma. Table also shows goodness-of-fit statistics: AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: *$p < 0.05$, ***$p < 0.001$.

References

Bates D, Maechler M, Bolker B. 2013. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. 2012. URL: http://CRAN R-project org/package= lme4.

Fox J, Weisberg S, Adler D, Bates D, Baud-Bovy G, Ellison S, Firth D, Friendly M, Gorjanc G, Graves S. 2016. Package 'car'.

Kleiman E. 2017. EMAtools: data management tools for real-time monitoring/ecological momentary assessment data.

Luke SG. 2017. Evaluating significance in linear mixed-effects models in R. Behavior Research Methods. 49:1494-1502.

R Core Team. 2014. R: A language and environment for statistical computing.

Wickham H. 2016. ggplot2: elegant graphics for data analysis.