Psychological Science OnlineFirst, published on May 21, 2013 as doi:10.1177/0956797612469537

Research Article

Psychological Science
XX(X) 1–10

DOI: 10.1177/0956797612469537
pss.sagepub.com

Compassion Training Alters Altruism and Neural Responses to Suffering

Helen Y. Weng[1,2,3], Andrew S. Fox[1,2,3,4], Alexander J. Shackman[4,5], Diane E. Stodola[2], Jessica Z. K. Caldwell[1,2,6,7], Matthew C. Olson[2], Gregory M. Rogers[5], and Richard J. Davidson[1,2,3,4,5]

[1]Department of Psychology, University of Wisconsin–Madison; [2]Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison; [3]Center for Investigating Healthy Minds at the Waisman Center, University of Wisconsin–Madison; [4]HealthEmotions Research Institute, University of Wisconsin–Madison; [5]Department of Psychiatry, University of Wisconsin–Madison; [6]Department of Psychiatry and Human Behavior, Brown University; and [7]Miriam Hospital, Brown University

Abstract

Compassion is a key motivator of altruistic behavior, but little is known about individuals' capacity to cultivate compassion through training. We examined whether compassion may be systematically trained by testing whether (a) short-term compassion training increases altruistic behavior and (b) individual differences in altruism are associated with training-induced changes in neural responses to suffering. In healthy adults, we found that compassion training increased altruistic redistribution of funds to a victim encountered outside of the training context. Furthermore, increased altruistic behavior after compassion training was associated with altered activation in brain regions implicated in social cognition and emotion regulation, including the inferior parietal cortex and dorsolateral prefrontal cortex (DLPFC), and in DLPFC connectivity with the nucleus accumbens. These results suggest that compassion can be cultivated with training and that greater altruistic behavior may emerge from increased engagement of neural systems implicated in understanding the suffering of other people, executive and emotional control, and reward processing.

Keywords

compassion, meditation, altruism, emotion regulation, fMRI, social behavior, neuroimaging, decision making, emotional control, individual differences

Received 3/21/12; Revision accepted 11/1/12

Compassion and altruism are of great interest to philosophical and scientific inquiry because of their central role in successful societies (Darwin, 1871/2004; Fehr & Fischbacher, 2003; Smith, 1759/2010; Sober, Wilson, & Wilson, 1999). Compassion is the emotional response of caring for and wanting to help those who are suffering (Batson, 1991; Eisenberg, Fabes, & Spinrad, 2006; Goetz, Keltner, & Simon-Thomas, 2010) and may have evolved in humans to foster altruistic acts that increase survival of kin as well as nonkin (Darwin, 1871/2004; Goetz et al., 2010; Sober et al., 1999). Such acts include enhancing the welfare of vulnerable offspring, promoting intimate bonds between partners, and facilitating cooperation among genetically unrelated strangers (Batson, 1991; Darwin, 1871/2004; Goetz et al., 2010; Sober et al., 1999). Despite the clear societal benefits of cultivating compassion, little is known about whether compassion and altruism can be trained and about the neural mechanisms that might underlie such effects.

Contemplative traditions claim that compassion can be enhanced with meditation training and that this results in greater real-world altruistic behavior (Lutz, Brefczynski-Lewis, Johnstone, & Davidson, 2008). In compassion training, compassion is cultivated toward different people, including loved ones, strangers, and difficult persons, as well as toward the self (Salzberg, 1997). Studies indicate

Corresponding Author:
Richard J. Davidson, Center for Investigating Healthy Minds, Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin–Madison, 1500 Highland Ave., Madison, WI 53705-2274
E-mail: rjdavids@wisc.edu

that compassion training can improve personal well-being, including stress-related immune responses (Pace et al., 2009), positive affect (Fredrickson, Cohn, Coffey, Pek, & Finkel, 2008; Hutcherson, Seppala, & Gross, 2008), and psychological and physical health (Fredrickson et al., 2008). Compassion training also enhances responses toward other people. Expert meditation practitioners show greater empathic neural responses when listening to sounds of other people's suffering during compassion meditation practice than control subjects do (Lutz, Brefczynski-Lewis, et al., 2008). Recent work suggests that compassion training can increase prosocial behavior (Leiberg, Klimecki, & Singer, 2011; Condon, Desbordes, Miller, & DeSteno, in press), positive emotions toward people who are suffering (Klimecki, Leiberg, Lamm, & Singer, 2012), and empathic accuracy (Mascaro, Rilling, Tenzin Negi, & Raison, 2013).

The neural mechanisms by which compassion training alters altruistic responses to suffering remain unknown. In the study reported here, we investigated whether short-term compassion training would enhance altruistic behavior toward a victim encountered outside of the training context. Altruistic behavior was assessed using the redistribution game, a novel economic decision-making task that models both unfair treatment of a victim and costly redistribution of funds to the victim. Furthermore, we measured brain activation before and after 2 weeks of training using functional MRI (fMRI) and investigated whether increased altruism could be explained by training-induced changes in the neural response to human suffering.

To rigorously test these hypotheses, we compared altruistic responses of participants given compassion training with responses of participants given an active control intervention of reappraisal training. Compassion trainees cultivated compassion for different targets, and reappraisal trainees practiced reinterpreting personally stressful events to decrease negative affect. Both interventions trained emotion-regulation strategies that promote well-being, but they differed in that the goal of compassion training was to increase empathic concern and the desire to relieve suffering (Lutz, Brefczynski-Lewis, et al., 2008), whereas the goal of reappraisal training was to decrease one's personal distress (Ochsner & Gross, 2005). Reappraisal training provided an ideal control for compassion training because although the combination of decreased distress and increased empathic concern predicts helping behavior (Batson, 1991; Eisenberg et al., 2006), reappraisal training only decreases distress without enhancing concern.

We hypothesized that compassion training would increase altruistic behavior by enhancing neural systems involved in (a) the recognition and understanding of another person's suffering and (b) emotion regulation of responses to suffering that support affiliation and helping behavior. The neuroscience of empathy highlights two systems for understanding the states of other people: experience sharing, which involves vicariously sharing the states of others, and mentalizing, which involves explicitly considering and understanding others' mental states through social inferences as well as through self-referential processes (Lamm, Decety, & Singer, 2011; Zaki & Ochsner, 2012). If the neural representation of suffering is increased by compassion training, then regulatory systems are needed to respond to this suffering with an approach rather than an avoidance response.

Prior theoretical and empirical work suggests that altruistic responses toward another person's suffering can be strengthened through either of two regulatory pathways (Decety & Jackson, 2006): (a) decreasing personal distress, which reduces negative arousal and avoidance behavior, or (b) increasing empathic concern, which strengthens the motivation to approach and relieve another person's suffering (Batson, 1991). In response to suffering, we predicted that greater altruism in compassion trainees would be associated with increased activation in prefrontal cortex (PFC), given its role in controlled processing (Miller & Cohen, 2001), emotion regulation (Ochsner & Gross, 2005; Urry et al., 2006; Wager, Davidson, Hughes, Lindquist, & Ochsner, 2008), and fronto-parietal control networks (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008). We also predicted that compassion training would be associated with decreased amygdala activation, given the amygdala's role in responding to negative stimuli and distress (Zald, 2003). Further, we hypothesized that greater prosocial behavior after compassion training would be associated with higher levels of activation in anterior insula, which has been implicated in studies of empathy and compassion (Immordino-Yang, McColl, Damasio, & Damasio, 2009; Lamm et al., 2011; Lutz, Brefczynski-Lewis, et al., 2008; Singer et al., 2006) and predicts helping behavior (Hein, Silani, Preuschoff, Batson, & Singer, 2010; Masten, Morelli, & Eisenberger, 2011). This greater prosocial behavior would also be correlated with increased activation in nucleus accumbens (NAcc), which has been linked to charitable giving (Harbaugh, Mayr, & Burghart, 2007; Moll et al., 2006) and positive appraisals of aversive stimuli (Wager et al., 2008).

We specifically tested these hypotheses against the reappraisal group, in which the psychological goal was self-focused (to decrease one's own suffering) rather than other-focused (to decrease other people's suffering through compassion). Although many of the same regions are implicated in reappraisal as in compassion (Wager et al., 2008), we expected that the hypothesized changes (e.g., increases in PFC activity) would not be associated with altruistic behavior because the behavior is not congruent with reappraisal's goals.

Method

Participants

Fifty-six participants completed the entire protocol, and the final sample consisted of 41 participants who believed that they were interacting with real players in the redistribution game (the other 15 participants expressed suspicion about the manipulation and were therefore excluded from the analysis; see Tables S1 and S2 and Supplementary Method and Analyses in the Supplemental Material available online for further information about the sample). Each participant was randomly assigned to receive either compassion training (n = 20; 12 female, 8 male; mean age = 21.9 years) or reappraisal training (n = 21; 13 female, 8 male; mean age = 22.5 years), completed 2 weeks of training (11 out of 14 practice days were required), and attended an fMRI session both before the start of training and after training finished. The groups did not differ in age, gender, baseline trait compassion, or the amount of practice time they received. Participants were healthy adults (18–45 years of age), right-handed, and had no previous experience in meditation or cognitive-behavioral therapy. No participant had issues that would pose a risk for his or her safety in the scanner. The experiment was approved by the University of Wisconsin–Madison Health Sciences Institutional Review Board. All subjects gave informed consent and were paid for participation.

Procedure

Overview. Participants came to the laboratory on three occasions. At Visit 1, each participant was randomly assigned to compassion training or reappraisal training and briefly instructed in the assigned strategy, following which he or she practiced the fMRI task in a mock MRI scanner. Visit 2 occurred approximately 1 week later; during this visit, participants completed the pretraining fMRI scan and began training later that day. Visit 3 occurred immediately after the 2 weeks of training were completed; this visit included the posttraining fMRI scan and the altruistic behavior task (performed outside of the scanner). For more details about the procedure, see Supplementary Method and Analyses in the Supplemental Material.

Trainings. Training consisted of practicing either compassion or reappraisal using guided audio instructions (via the Internet or compact disc) for 30 min per day for 2 weeks. Compassion trainees practiced cultivating feelings of compassion for different targets (a loved one, the self, a stranger, and a difficult person), and reappraisal trainees practiced reinterpreting personally stressful events to decrease negative affect (see Trainings and Fig. S1 in the Supplemental Material).[1]

Altruistic behavior task: redistribution game. We tested whether compassion training could affect altruistic behavior outside of the training context using the redistribution game. This economic decision-making task models both unfair treatment of a victim and costly redistribution of funds to the victim. Using anonymous online interactions, participants first observed a dictator (endowed with $10) transfer an unfair amount of money ($1) to a victim who had no money (Fig. 1a). After witnessing this violation of the fairness norm (Fehr & Fischbacher, 2003), participants could choose to spend any amount of their own endowment ($5) to compel the dictator to give twice that amount to the victim (Fig. 1b). Participants were paid the amount that was left in their endowment after making the decision. (See Supplementary Method and Analyses in the Supplemental Material for full details of the redistribution game.)

Participants were told that they were playing the game with live players over the Internet. Effects of demand characteristics on behavior were minimized by presenting the game as a unique study, describing it in purely economic terms, never instructing participants to use the training they received, removing the physical presence of players and experimenters during game play, and enforcing real monetary consequences for participants' behavior. Because compassionate behavior is specifically evoked by unfairness, all participants observed the same preprogrammed unfair dictator offer. At the end of the entire protocol, participants were debriefed and asked whether they believed they were playing against real people in the game. Data were analyzed only for participants who believed the paradigm (see Table S2 in the Supplemental Material).

fMRI task and stimuli. To determine whether altruistic behavior was predicted by changes in neural responses to human suffering, we scanned participants using fMRI before and after training while they employed their assigned emotion-regulation strategy. Participants in the two groups were presented with images of human suffering and nonsuffering (neutral condition). Compassion trainees were instructed to evoke feelings of compassion while silently repeating compassion-generating phrases. In contrast, reappraisal trainees were instructed to decrease negative emotions by silently reinterpreting the emotional meaning of the images. (See Supplementary Method and Analyses in the Supplemental Material for fMRI data-acquisition parameters).

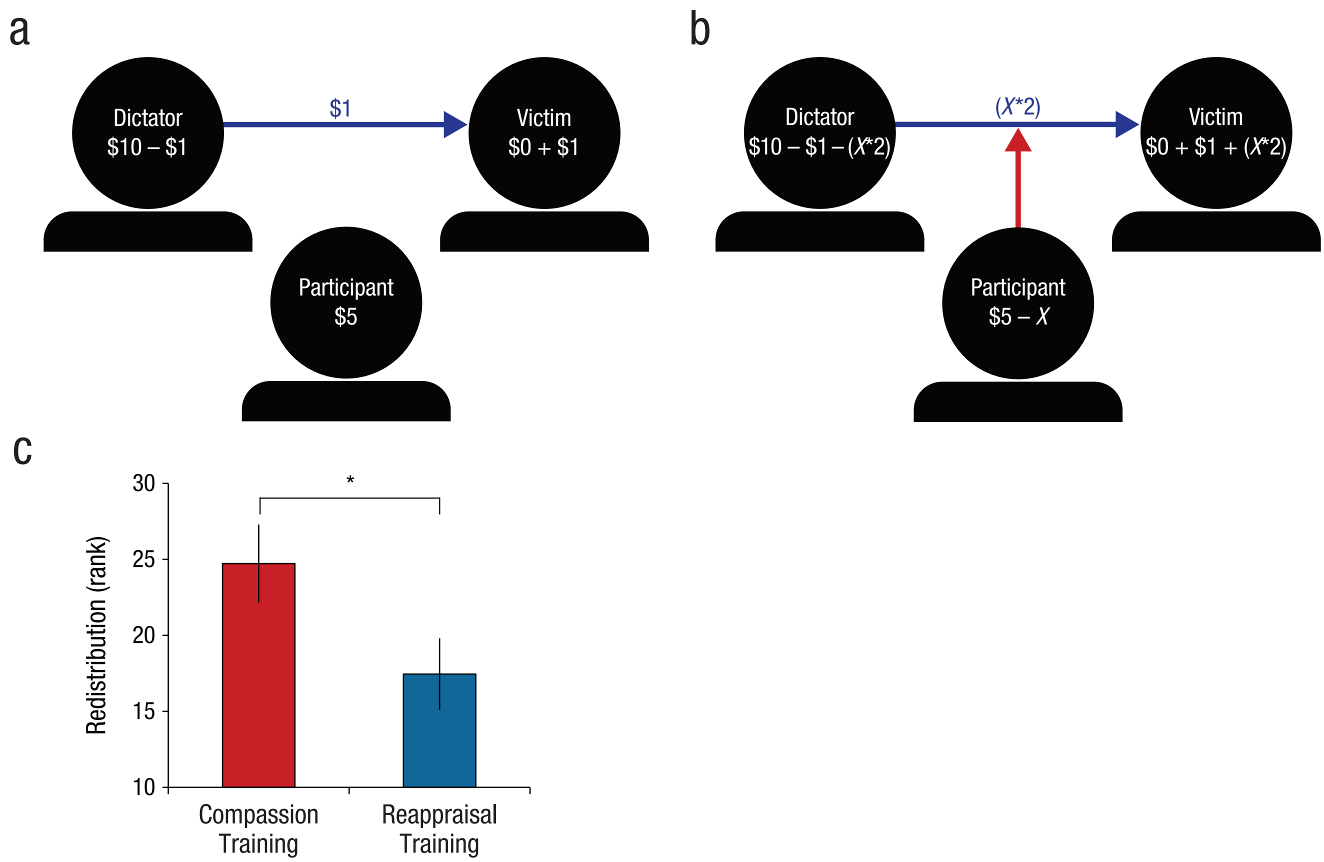

Fig. 1. Paradigm of and results for the redistribution game. In the first stage of the redistribution game (a), the dictator (endowed with $10) transfers an unfair amount of money ($1) to the victim while the participant (endowed with $5) observes. In the second stage (b), the participant can spend any amount (*X*) up to $5 to compel the dictator to give twice that amount to the victim. The graph (c) shows the average rank-transformed redistribution amount as a function of the type of training participants received. Redistribution of $4 (i.e., *X* = $2; rank = 35.5 of 41) results in an equal distribution between the dictator and the victim ($5 each). The asterisk indicates that there was a significant difference between groups ($p < .05$). Error bars denote standard errors of the mean.

Images of suffering depicted emotional distress, physical pain, or acts of violence (e.g., a burn victim, a crying child). Neutral images depicted people in nonemotional situations, such as working or walking down a street. Two parallel sets of images (20 suffering and 16 neutral) were created to ensure that participants viewed different images before and after training. Set order was counterbalanced and randomized. Images were pseudorandomized so that three or more images from either condition were not presented in a row. Image randomization was performed once for each set and then fixed. Images were balanced across sets for published normative ratings of valence and arousal, as well as for properties of hue, luminance, and saturation (all ps > .1).

Participants were instructed to regulate their emotional responses to the images over three blocks. Each block began with a 20-s fixation baseline period. Participants then received both an auditory and visual instruction (3 s) stating that they should invoke either "compassion" or "reappraisal" (depending on group assignment), which was followed by a fixation cross (5–7 s). They then applied the assigned regulation strategy to a series of 12 images. Each image was presented for 12 s and separated by a fixation interval (5–11 s, randomized). Blocks ended with a final fixation baseline (17–38 s). After each block, participants saw each image again for 2 s and rated the arousal of each image (1 = *least arousing*, 7 = *most arousing*).

Behavioral analysis

Across all participants, the redistribution response was positively skewed (skewness = 1.5, *SE* = 0.37), and 2 participants qualified as outliers (> 3 *SD* from the population mean). Because of these violations of normality, we rank-ordered the behavioral response across both groups so that strong assumptions were not made about the scaling or normality of the residuals. Parametric tests were then performed on the ranked data. To test the mean difference between groups, we performed an independent-samples t test on the ranks. For in-depth analyses and discussion of redistribution values and ranks, see Supplementary Method and Analyses in the Supplemental Material.

fMRI analyses

Overview. A series of tests were conducted to identify regions in which changes due to training predicted greater redistribution in compassion training than in reappraisal training. A whole-brain interaction contrast (Group × Redistribution Rank) was tested on neural-change scores. Follow-up tests were conducted using both across- and within-subjects analyses to identify regions that were functionally connected to clusters identified in the interaction analysis and networks involved in emotion regulation. Finally, we investigated whether reported arousal was associated with either redistribution or neural changes. See Supplementary Method and Analyses in the Supplemental Material for full details.

Interaction analysis. Individual functional and structural MRI brain data were preprocessed and normalized to Montreal Neurological Institute (MNI) space. Each participant's neural response to suffering during regulation was estimated with the contrast between activation to images of suffering and activation to neutral images at each fMRI scan time point (before training, after training) using standard first-level analyses (see Supplementary Method and Analyses in the Supplemental Material), and beta coefficients were converted to percentage signal change (PSC). Training-induced changes were calculated by subtracting PSC values before each scan from PSC values after each scan. To identify regions where training-related changes specifically predicted greater redistribution in compassion trainees than in reappraisal trainees, a second-level Group × Redistribution Rank voxel-wise analysis was performed, controlling for main effects of group and redistribution.

First, whole-brain analyses were conducted and corrected for multiple comparisons ($p < .01$ after an initial voxel-wise threshold of $p < .01$) using Monte Carlo simulation. This analysis identified the right inferior parietal cortex (IPC). To decompose the interactions, we extracted mean PSC-change scores from the clusters for each participant and analyzed them to yield parameter estimates and determine the directionality of the relationship for each group. These values were used for descriptive and diagnostic purposes only (Vul, Harris, Winkielman, & Pashler, 2009). In region-of-interest (ROI) analyses, data from the Group × Redistribution Rank interaction were corrected for multiple comparisons ($p < .01$ after an initial voxel-wise threshold of $p < .01$) using Monte Carlo simulation within bilateral a priori ROIs of the amygdala, insula, and NAcc.

IPC conjunction analysis. To identify regions that may be functionally connected to the IPC in order to increase the amount participants redistributed in each training group, we performed a conjunction analysis requiring voxels to be (a) correlated with changes in IPC activation across participants in both groups (voxel-wise $p < .01$) and (b) identified in the original Group × Redistribution Rank interaction (voxel-wise $p < .01$). This analysis identified a cluster in the dorsolateral PFC (DLPFC; whole-brain corrected at $p < .01$ after a conjunction voxel-wise threshold of $p < .001$).

Psychophysiological interaction (PPI) analysis. To determine regions in which altered PFC connectivity predicted higher amounts of redistribution in compassion trainees than in reappraisal trainees, we performed a PPI analysis using the DLPFC seed region identified by the IPC conjunction analysis. The PPI regressor consisted of comparing DLPFC connectivity in response to images of suffering with DLPFC connectivity in response to neutral images. Training-induced PPI changes were calculated by subtracting PPI betas before each scan from PPI betas after each scan. To identify regions where training-related PPI changes specifically predicted greater redistribution in compassion trainees than in reappraisal trainees, we performed a second-level Group × Redistribution Rank voxel-wise analysis, controlling for main effects of group and redistribution. Voxel-wise regression maps were corrected for multiple comparisons ($p < .01$ after an initial voxel-wise threshold of $p < .01$) using Monte Carlo simulation within each bilateral ROI (amygdala, insula, and NAcc). For descriptive purposes, mean PPI-change betas were extracted from the clusters for each participant and analyzed to yield parameter estimates and determine the directionality of the relationship for each group.

Correlational analyses. Compassion training may increase altruistic behavior by decreasing personal distress evoked by suffering (Batson, 1991; Eisenberg et al., 2006). To test this, we computed arousal-change scores (analogous to the neural-change scores) and correlated them with altruistic redistribution in each group. To examine whether changes in arousal were associated with changes in neural responses to suffering, we computed correlations between arousal-change scores and neural-change scores identified in the previous fMRI analyses in each group.

Results

Altruistic redistribution

Findings in an independent validation sample[2] (N = 72) confirmed that altruistic redistribution is a behavioral signature of compassion: Individuals who endorsed greater levels of trait empathic concern (Davis, 1980) redistributed more money, $r(70) = .43$, $p < .001$ (Fig. S2 in the Supplemental Material). In the main study, after 2 weeks of training, compassion trainees spent more money to redistribute funds to the victim compared with reappraisal trainees (Fig. 1c), independent-samples $t(39) = 2.09$, $p < .05$, $d = 0.65$. Compassion trainees also spent more than individuals with no training in the validation sample (Fig. S3 in the Supplemental Material). Compassion trainees spent 1.84 times more money than reappraisal trainees ($1.14 vs. $0.62, respectively) and increased the distribution between the dictator and the victim by 57%. In contrast, reappraisal trainees increased the distribution by only 31%. This demonstrates that purely mental training in compassion can result in observable altruistic changes toward a victim, even when individuals are not explicitly cued to generate compassion.

Neuroimaging

Group differences in neural change and altruistic redistribution. We hypothesized that greater altruism resulting from compassion training would be predicted by training-related changes in the neural responses to images of suffering. The whole-brain Group × Redistribution Rank interaction test revealed that training-induced changes in right IPC activation were differentially associated with altruistic redistribution in the two training groups (Fig. 2a; $p < .01$, corrected; see Tables S3 and S4 in the Supplemental Material). In compassion trainees, greater IPC activation due to training was associated with greater redistribution, but this was not found in reappraisal training (Fig. 2b; also see Table S5 in the Supplemental Material). Within the a priori ROIs, no region survived correction at $p < .01$. See Fig. S4 and Supplementary Method and Analyses in the Supplemental Material for exploratory analyses within the ROIs.

The IPC is implicated in experience sharing as part of the mirror-neuron network (Gallese, Keysers, & Rizzolatti, 2004; Lamm et al., 2011), and we investigated whether the IPC was functionally connected to other regions that also differentially predicted redistribution between groups. The IPC conjunction test identified only the DLPFC (Fig. 2c; also see Tables S3 and S4 in the Supplemental Material; $p < .01$ whole-brain corrected), where greater increases in DLPFC activation predicted greater altruistic redistribution in compassion trainees, and the opposite relationship was found in reappraisal trainees (Fig. 2d; also see Table S5 in the Supplemental Material). The changes in IPC and DLPFC were highly coupled—compassion training: $r(18) = .92$, $p < .001$; reappraisal training: $r(19) = .79$, $p < .001$, and both regions differentially predicted redistribution between groups. These findings suggest that fronto-parietal executive control networks (Dosenbach et al., 2008; Vincent et al., 2008) may be recruited by compassion training in order to regulate emotions and increase altruistic behavior.

DLPFC PPI connectivity changes and altruistic redistribution. Emotion regulation is thought to involve the influence of the PFC over other regions such as the amygdala, insula, and NAcc (Ochsner & Gross, 2005; Urry et al., 2006; Wager et al., 2008). Using PPI, we tested whether changes in task-related functional connectivity between DLPFC and amygdala, DLPFC and NAcc, or DLPFC and insula predicted greater altruistic redistribution in compassion training than in reappraisal training. Using the DLPFC cluster defined by the IPC conjunction test as a seed (Fig. 3a), we found a significant interaction in the NAcc, demonstrating that DLPFC-NAcc connectivity was differentially associated with redistribution in compassion training compared with reappraisal training (Fig. 3b; $p < .01$, corrected within the ROI; see Tables S3 and S4 in the Supplemental Material). Compassion trainees who showed greater DLPFC-NAcc connectivity redistributed more funds after training, whereas reappraisal trainees who showed greater DLPFC-NAcc connectivity redistributed less money after training (Fig. 3c; see also Table S5 in the Supplemental Material; see Supplementary Method and Analyses in the Supplemental Material for discussion of the directionality of the connectivity). No relationship was found in the insula or the amygdala.

Arousal correlations with altruistic redistribution and neural change. Compassion training may increase altruistic behavior by decreasing personal distress evoked by suffering (Batson, 1991; Eisenberg et al., 2006). We found that decreases in reported arousal to images of suffering were correlated with increased redistribution in compassion trainees, $r(18) = −.45$, $p < .05$, but not in reappraisal trainees, $r(19) = .09$, $p = .70$. We further investigated whether decreases in arousal were associated with neural changes and found that greater DLPFC-NAcc connectivity was correlated with decreases in arousal in compassion trainees, $r(18) = −.64$, $p < .01$, but not in reappraisal trainees, $r(19) = −.13$, $p = .59$. Decreases in arousal were not associated with IPC or DLPFC changes in either group (all ps $\geq .21$).

Discussion

Individuals who trained in compassion for 2 weeks were more altruistic toward a victim after witnessing an unfair social interaction compared with individuals who trained

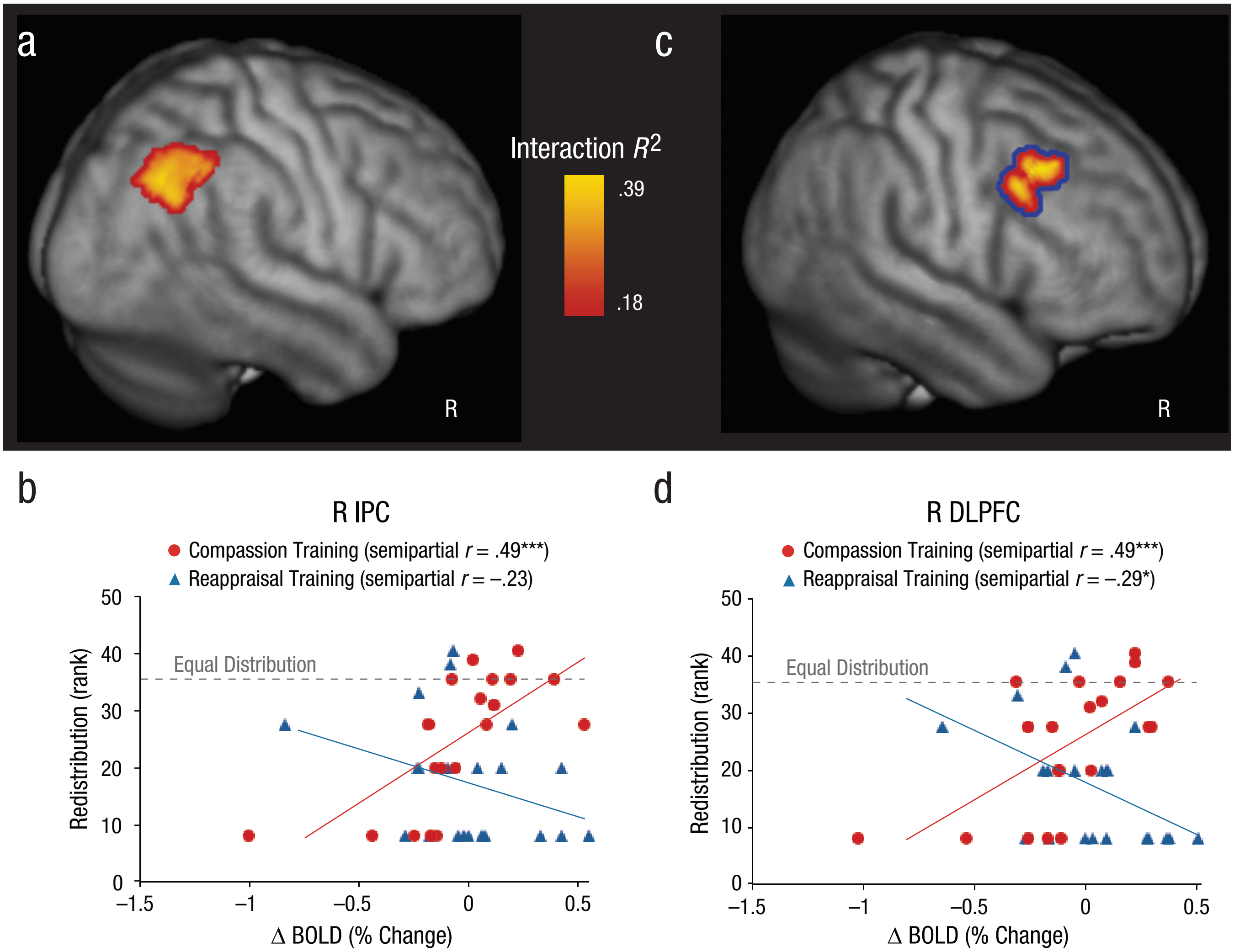

Fig. 2. Activation in right inferior parietal cortex (IPC) and right dorsolateral prefrontal cortex (DLPFC). The brain images in (a) and (c) show blood-oxygen-level-dependent (BOLD) changes in right IPC and right DLPFC, respectively, from before to after training in compassion trainees; in both sessions, participants regulated their emotional responses while viewing images of human suffering. Images are in Montreal Neurological Institute space. The color coding indicates the amount of variance accounted for by the Group × Redistribution Rank interaction. The blue outline in (c) indicates that the cluster was identified by a conjunction test. The scatter plots (with best-fitting regression lines) in (b) and (d) show rank-transformed redistribution amounts for the two training groups as a function of the percentage signal change in BOLD responses from before training to after training. The dashed line indicates the rank of a redistribution of $4 (rank = 35.5 of 41), which results in an equal $5 distribution between the dictator and the victim. Asterisks indicate significant results (*p < .05, ***p < .001). R = right.

in reappraisal and individuals in a validation control group. This demonstrates that a purely mental training can generalize to behavioral domains by affecting social behavior outside of the training context. Furthermore, increases in altruistic responses were correlated with training-related changes in the neural response to suffering, which provides evidence for functional neuroplasticity in the circuitry underlying compassion and altruism.

The pattern of neural changes in compassion training suggests that increased altruistic behavior is achieved by enhancing neural mechanisms that support the understanding of others' states, greater fronto-parietal executive control, and up-regulation of positive emotion systems. Greater IPC activation specifically predicted greater redistribution in compassion trainees and not in reappraisal trainees, which suggests that IPC recruitment is a unique neural marker for altruism induced by compassion training. This region has been implicated in the human mirror-neuron system (Gallese et al., 2004) and may reflect increased simulation of the suffering of other people. If

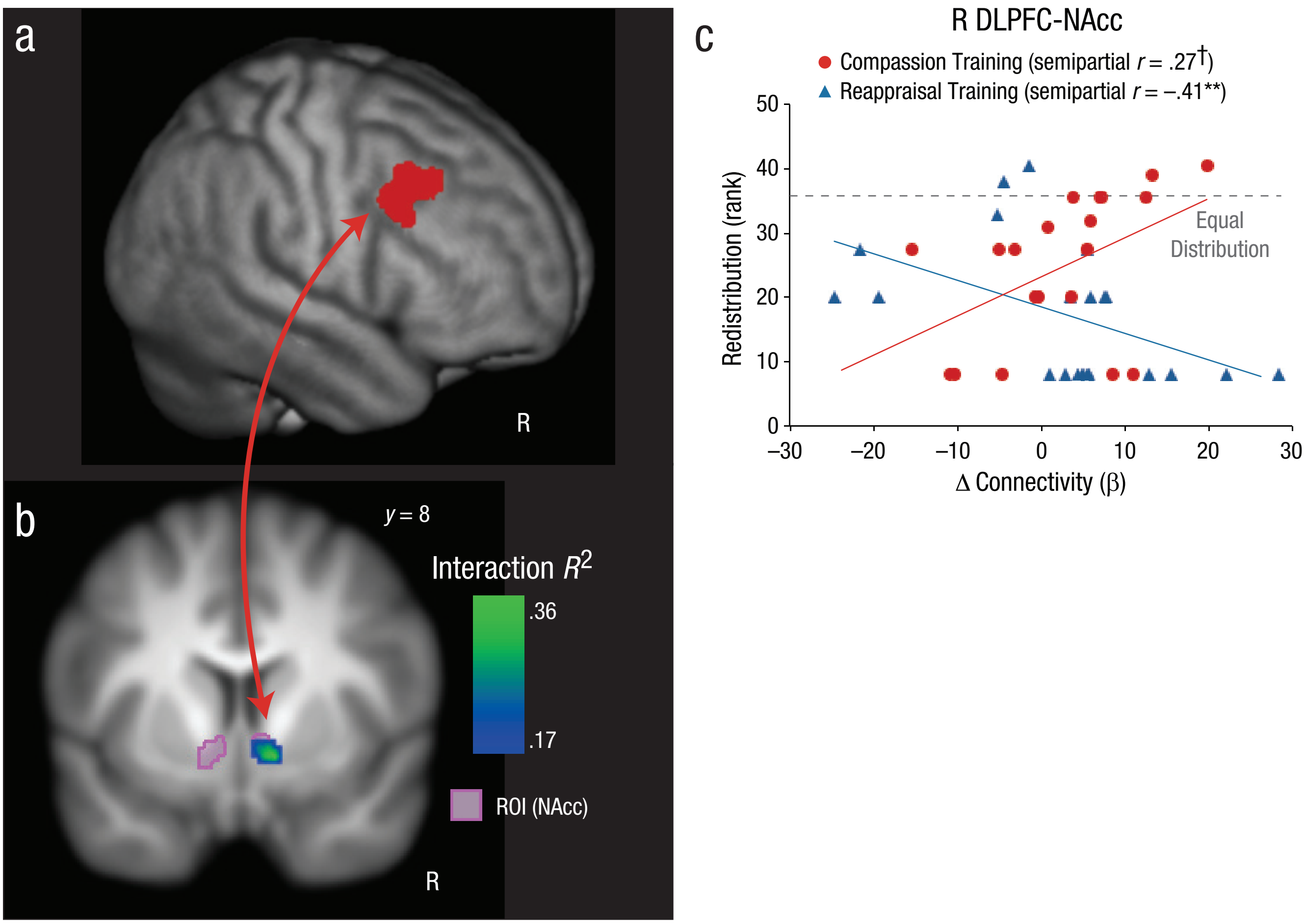

Fig. 3. Results for connectivity between right dorsolateral prefrontal cortex (DLPFC) and nucleus accumbens (NAcc). The image in (a) shows the DLPFC cluster identified by the conjunction test, which was used as the seed region in the psychophysiological interaction analysis. The image in (b) shows the regions of NAcc in which there was differential activation in response to images of human suffering before and after training in compassion trainees. Images and coordinates are in Montreal Neurological Institute space. The color coding indicates the amount of variance accounted for by the Group × Redistribution Rank interaction. The scatter plot (with best-fitting regression lines) in (c) shows rank-transformed redistribution amounts for the two training groups as a function of changes in right DLPFC-NAcc connectivity. The dashed line indicates the rank of a redistribution of \$4 (rank = 35.5 of 41), which results in an equal \$5 distribution between the dictator and the victim. Asterisks indicate significant results (** $p < .01$, † $p = .07$). R = right; ROI = region of interest.

the signal of other people's suffering is indeed increased by compassion training, this leads to an emotion-regulatory challenge that requires trainees to approach rather than avoid suffering in order to engage in prosocial behavior. This transformation of emotional response may have been instantiated by a fronto-parietal executive control network (Dosenbach et al., 2008; Vincent et al., 2008) in order to increase altruistic behavior in compassion trainees. The coordinated activation of the IPC and DLPFC in compassion trainees may reflect greater sustained attention and goal maintenance (Miller & Cohen, 2001) to help other individuals, as well as integration of information from systems that process both external information (of other people's suffering) and internal information (the goal to help; Vincent et al., 2008).

Regulation of internal information may include increasing positive emotions toward other people's suffering, as reflected by the increased DLPFC-NAcc connectivity that predicted redistribution in compassion trainees. This may represent increasing positive appraisals of aversive stimuli (Wager et al., 2008) by enhancing the reward value of the victim's well-being (e.g., caring) and increasing the anticipated reward (Knutson & Cooper, 2005) of helping the victim. Furthermore, decreased reported arousal after compassion training may be due to enhancement of reward-related neural systems. These findings also support research suggesting that compassion training enhances positive emotions and neural substrates of affiliation (Klimecki et al., 2012).

The relationship between training-induced neural changes (DLPFC activation and DLPFC-NAcc connectivity) and altruistic behavior was not unique to compassion trainees. In fact, greater changes in these regions predicted less redistribution in reappraisal trainees. This finding may be due to the differing regulatory goals between compassion training and reappraisal training. In compassion training, the goal was to increase caring for people who are suffering and to help, whereas the goal in reappraisal training was to decrease personal negative emotions. In a social context, the goals of compassion training and reappraisal training are opposing (other-focused vs. self-focused), and this may explain the crossover interaction effects. In reappraisal training, neural changes may have resulted in decreased helping of other individuals in order to serve the primary goal of decreasing personal negative affect.

A clear limitation of this study is that altruistic behavior was not measured at pretraining, although a separate validation sample was used to estimate pretraining levels. Future research may build on this study's findings by measuring altruism at baseline, which may strengthen claims that compassion training increases altruism (Leiberg et al., 2011). Furthermore, emotional valence and arousal may be measured using methodology that is less susceptible to demand characteristics, such as facial electromyography and skin conductance response. In future research, longitudinal designs should be employed to determine the length of compassion training needed to have sustained behavioral effects.

In sum, these results build on existing evidence that the adult human brain may demonstrate functional and structural changes after mental training (Davidson & McEwen, 2012; Klingberg, 2010; Lutz, Slagter, Dunne, & Davidson, 2008) and extend these previous findings to include socioemotional domains such as compassion and altruism. Our findings support the possibility that compassion and altruism can be viewed as trainable skills rather than as stable traits. This lays the groundwork for future research to explore whether compassion-related trainings can benefit fields that depend on altruism and cooperation (e.g., medicine) as well as clinical subgroups (Hofmann, Grossman, & Hinton, 2011) characterized by deficits in compassion, such as psychopaths (Blair, 2007).

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by the National Institutes of Health (P01-AT004952 and T32-MH018931—R. J. Davidson, principal investigator; P30-HD003352—M. Seltzer, principal investigator); University of Wisconsin–Madison Department of Psychology (Hertz Award to A. S. Fox and H. Y. Weng); the Fetzer Institute, John Templeton Foundation, Impact Foundation, J. W. Kluge Foundation, and Mental Insight Foundation (R. J. Davidson); the Mind and Life Institute (H. Y. Weng); and gifts from Bryant Wangard, Ralph Robinson, Keith and Arlene Bronstein, and Ann Down (R. J. Davidson).

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

Notes

1. Training audio files and written scripts can be downloaded at www.investigatinghealthyminds.org/compassion.html.
2. See Supplementary Method and Results in the Supplemental Material for detailed method and analyses regarding the independent validation study.

References

Batson, C. D. (1991). *The altruism question*. Hillsdale, NJ: Erlbaum.

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*, 387–392.

Condon, P., Desbordes, G., Miller, W., & DeSteno, D. (in press). Meditation increases compassionate responses to suffering. *Psychological Science*.

Darwin, C. (2004). *The descent of man and selection in relation to sex*. London, England: Penguin. (Original work published 1871)

Davidson, R. J., & McEwen, B. S. (2012). Social influences on neuroplasticity: Stress and interventions to promote well-being. *Nature Neuroscience*, *15*, 689–695.

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

Decety, J., & Jackson, P. L. (2006). A social-neuroscience perspective on empathy. *Current Directions in Psychological Science*, *15*, 54–58.

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*, 99–105.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 3: Social, emotional, and personality development* (pp. 646–718). New York, NY: Wiley.

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

Fredrickson, B. L., Cohn, M. A., Coffey, K. A., Pek, J., & Finkel, S. M. (2008). Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. *Journal of Personality and Social Psychology*, *95*, 1045–1062.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*, 396–403.

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin*, *136*, 351–374.

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

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, *68*, 149–160.

Hofmann, S. G., Grossman, P., & Hinton, D. E. (2011). Loving-kindness and compassion meditation: Potential for psychological interventions. *Clinical Psychology Review*, *31*, 1126–1132.

Hutcherson, C. A., Seppala, E. M., & Gross, J. J. (2008). Loving-kindness meditation increases social connectedness. *Emotion*, *8*, 720–724.

Immordino-Yang, M. H., McColl, A., Damasio, H., & Damasio, A. (2009). Neural correlates of admiration and compassion. *Proceedings of the National Academy of Sciences, USA*, *106*, 8021–8026.

Klimecki, O. M., Leiberg, S., Lamm, C., & Singer, T. (2012). Functional neural plasticity and associated changes in positive affect after compassion training. *Cerebral Cortex*. Advance online publication. doi:10.1093/cercor/bhs142

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*, 317–324.

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*, 411–417.

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

Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS ONE*, *6*(3), e17798. Retrieved from http://www.plosone.org/article/info:doi/10.1371/journal.pone.0017798

Lutz, A., Brefczynski-Lewis, J., Johnstone, T., & Davidson, R. J. (2008). Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise. *PloS ONE*, *3*(3), e1897. Retrieved from http://www.plosone.org/article/info:doi/10.1371/journal.pone.0001897

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, *12*, 163–169.

Mascaro, J. S., Rilling, J. K., Tenzin Negi, L., & Raison, C. L. (2013). Compassion meditation enhances empathic accuracy and related neural activity. *Social Cognitive and Affective Neuroscience*, *8*, 48–55.

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for "social pain" and subsequent prosocial behavior. *NeuroImage*, *55*, 381–388.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences, USA*, *103*, 15623–15628.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*, 242–249.

Pace, T. W. W., Negi, L. T., Adame, D. D., Cole, S. P., Sivilli, T. I., Brown, T. D., . . . Raison, C. L. (2009). Effect of compassion meditation on neuroendocrine, innate immune and behavioral responses to psychosocial stress. *Psychoneuroendocrinology*, *34*, 87–98.

Salzberg, S. (1997). *Lovingkindness: The revolutionary art of happiness*. Boston, MA: Shambhala.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*, 466–469.

Smith, A. (2010). *The theory of moral sentiments*. London, England: Penguin Classics. (Original work published 1759)

Sober, P. E., Wilson, P. D. S., & Wilson, D. S. (1999). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Urry, H. L., van Reekum, C. M., Johnstone, T., Kalin, N. H., Thurow, M. E., Schaefer, H. S., . . . Davidson, R. J. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *Journal of Neuroscience*, *26*, 4415–4425.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, *100*, 3328–3342.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290.

Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, *59*, 1037–1050.

Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, *15*, 675–680.

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews*, *41*, 88–123.

Compassion training alters altruism and neural responses to suffering

Helen Y. Weng, Andrew S. Fox, Alexander J. Shackman, Diane E. Stodola, Jessica Z. K. Caldwell, Matthew C. Olson, Gregory M. Rogers, Richard J. Davidson

Supplemental Material - Table of Contents

Acknowledgements

Authors thank Michael Anderle, Lisa Angelos, Donna Cole, John Curtin, Tenzin Dhanze, Isa Dolski, Ron Fisher, Michael Koenigs, Hyejeen Lee, John Ollinger, Michael Place, Tim Salomons, Brianna Schuyler, Dana Tudorascu, Nathan Vack, Linda Wuestenberg, Carolyn Zahn-Waxler, and the staffs of the Center for Investigating Healthy Minds, Laboratory for Affective Neuroscience, and Waisman Laboratory for Brain Imaging and Behavior.

Compassion and Reappraisal Training Study

Participants. Participants were recruited from the Madison, WI community and consented to a protocol approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board. To ensure that the population was not biased towards learning any particular training, participants were recruited based on their interest in learning how to reduce negative emotions. Potential participants were screened over the phone, and enrolled if they were 18-45 years in age, right-handed, reported no history of psychiatric diagnosis or psychotropic medication, had no previous experience in meditation or cognitive-behavioral therapy, and had no contra-indications for receiving an MRI scan (e.g., implanted ferromagnetic devices). Participants were randomized to the compassion (COM) or reappraisal (REP) training group before the first study visit. 56 participants completed the entire protocol (COM n = 28, REP n = 28). 7 subjects did not complete the study (COM n = 5, REP n = 2) due to lack of time (n = 2), lack of evidence that trainings were completed (n = 2; practice could not be verified by either online completion or paper logging of practice time and ratings), excessive motion during the pre-training fMRI scan (n = 1), and discomfort in the scanner environment (n = 2). Participants were paid $8/hour or $165 for completing the entire study, plus whatever was earned in the Redistribution Game. Of the 56 completed participants, 41 had valid redistribution data where they believed they were interacting with live players (COM n = 20, REP n = 21). Data analyses were restricted to these 41 participants (4 minorities in each group). Training groups did not differ in age, trait empathic concern, or training time (**Table S1**). Believers of the redistribution paradigm did not significantly differ from Non-Believers of the paradigm in the N of each training group, age, trait empathic concern, training time or practice ratings, or average redistribution (**Table S2**).

Overview. After screening, participants came to the laboratory for 3 study visits. Participants were randomized to COM or REP before the first visit. During Visit 1, they were introduced to the assigned training strategy (compassion or reappraisal), and practiced the appropriate strategy in the fMRI task using a mock MRI scanner. Visit 2 occurred approximately one week later, where they completed the pre-training fMRI scan, and training began later that day. Training consisted of practicing COM or REP using guided audio instructions for 30 minutes/day for 2 weeks. Visit 3 occurred immediately after the 2 weeks of training was completed, and consisted of the post-training fMRI scan and the altruistic Redistribution Game.

Trainings.

Training Structure. Trainings consisted of pre-recorded guided audio instructions that could be accessed at participants' convenience via the Internet or compact disc. Each training script was written by an expert, and was professionally recorded at a university recording studio to be approximately 30 minutes long. The same speaker read the scripts for each of the trainings. Training audio files and written scripts can be downloaded at http://www.investigatinghealthyminds.org/compassion.html.

Both groups were required to practice at least 11 out of 14 days of training (~79%). After each training session was completed, participants recorded the number of minutes practiced, level of focus, and improvement (1 = *least*, 9 = *most*).

Compassion Meditation Training (COM). The purpose of COM was to increase feelings of compassion for different targets, starting with easier targets (a Loved One and the Self), and moving to increasingly difficult targets (a Stranger and a Difficult Person). From this perspective, practicing compassion is treated like exercising a muscle, strengthening over time as increasing difficulty or weight is applied. This structure was adopted from lovingkindness meditation (Salzberg, 1997). The meditation was specifically designed for beginners and was written in secular language. The COM script was written by Linda Wuestenberg, M.Ed., a Licensed Clinical Social worker (LCSW) and a Certified Substance Abuse Counselor (CSAC). Her clinical practice began in 1982, and she has been a practitioner of compassion meditation since 1993. She practices compassion meditation in the Drikung Kagyu tradition of Tibetan Buddhist meditation, under the tutelage of Khenchen Konchog Gyaltsen and Garchen Rinpoche. She has used compassion directly or indirectly in her clinical practice for approximately 15 years.

Participants practiced compassion for targets by 1) contemplating and envisioning their suffering and then 2) wishing them freedom from that suffering. They first practiced compassion for a Loved One, such as a friend or family member. They imagined a time their loved one had suffered (e.g., illness, injury, relationship problem), and were instructed to pay attention to the emotions and sensations this evoked. They practiced wishing that the suffering were relieved and repeated the phrases, "May you be free from this suffering. May you have joy and happiness." They also envisioned a golden light that extended from their heart to the loved one, which helped to ease his/her suffering. They were also instructed to pay attention to bodily sensations, particularly around the heart. They repeated this procedure for the Self, a Stranger, and a Difficult Person. The Stranger was someone encountered in daily life but not well known (e.g., a bus driver or someone on the street), and the Difficult Person was someone with whom there was conflict (e.g., coworker, significant other). Participants envisioned hypothetical situations of suffering for the stranger and difficult person (if needed) such as having an illness or experiencing a failure. At the end of the meditation, compassion was extended towards all beings. For each new meditation session, participants could choose to use either the same or different people for each target category (e.g., for the loved one category, use sister one day and use father the next day). After the meditation, participants rated how much compassion they felt on a likert scale (1 = *least*, 7 = *most*) for each target after the 30 minutes of practice.

Reappraisal Training (REP). The purpose of REP was to decrease the intensity of negative emotions evoked by personally stressful events by reinterpreting the nature of the stressful event. The REP script was modeled after homework exercises used in cognitive-behavioral therapy (Beck, Rush, Shaw & Emery, 1979). The guided instructions were accompanied by written assignments that were completed online or on a worksheet. The training was written by Gregory Rogers, Ph.D., a clinical psychologist who was licensed in 1999. Dr. Rogers has been certified as a qualified cognitive therapist by the Academy of Cognitive Therapy, and is a member of the Association for Behavioral and Cognitive Therapies.

REP participants were asked to recall a stressful experience from the past 2 years that remained upsetting to them, such as arguing with a significant other or receiving a lower-than-expected grade. They were instructed to vividly recall details of the experience (location, images, sounds). They wrote a brief description of the event, and chose one word to best describe the feeling experienced during the event (e.g., sad, angry, anxious). They rated the intensity of the feeling *during* the event, and the intensity of the *current* feeling on a scale (0 = *No feeling at all,* 100 = *Most intense feeling in your life*). They wrote down the thoughts they had during the event in detail. Then they were asked to reappraise the event (to think about it in a different, less upsetting way) using 3 different strategies, and to write down the new thoughts. The strategies included 1) thinking about the situation from another person's perspective (e.g., friend, parent), 2) viewing it in a way where they would respond with very little emotion, and 3) imagining how they would view the situation if a year had passed, and they were doing very well. After practicing each strategy, they rated how reasonable each interpretation was (0 = *Not at all reasonable, 100* = *Completely reasonable*), and how badly they felt after considering this view (0 = *Not bad at all*, 100 = *Most intense ever*). Day to day, participants were allowed to practice reappraisal with the same stressful event, or choose a different event. Participants logged the amount of minutes practiced after the session.

Training Adherence. Most participants used the Internet to access trainings (63% used the internet for at least 75% of training days), which provided a rigorous method for determining training adherence. Participant ID and password were required to access the training site, and thus provided a time stamp for experimenters to monitor practice. The online questionnaires were also monitored for completion. Practice was tracked daily, and if participants were in danger of not completing 11/14 days of practice, they were contacted by phone or e-mail. Trainings for participants who listened via CD were counted by the date and number of paper daily training questionnaires completed.

Training efficacy.

COM: Effects of time and target. To test whether feelings of compassion increased over the training period, we first averaged compassion ratings (collected after each practice session) from the early (first 3 days) and late (last 3 days) periods of training. We also predicted that compassion ratings would differ by target. The Time (Early, Late) × Target (Loved One, Self, Stranger, Difficult Person) interaction on compassion ratings was significant ($F_{3,17} = 3.31$, $p < 0.05$). Main effects of Time ($F_{3,17} = 317.89$, $p < .001$) and Target ($F_{3,17} = 41.09$, $p < .001$) were also significant.

To decompose the effect of time within each target, we performed contrasts of Early vs. Late compassion ratings for each target. Feelings of compassion increased from early to late periods of training for every target (all p's < 0.05) except the Loved One ($p = 0.72$; **Fig. S1A**). We predicted that compassion ratings for targets would exhibit the following pattern: Loved One > Self, Self > Stranger, Stranger > Difficult Person. The direction of our predictions were confirmed during the first 3 days of practice, although the difference between Self > Stranger did not reach significance ($p = .14$, **Fig. S1B**). These differences were less apparent during the last 3 days of practice, where only Loved One > Self was significant ($p < .001$; **Fig. 1SB**). Although Self > Stranger and Stranger > Difficult Person were not significant (p's > .14; **Fig. S1B**), Self > Difficult Person was significantly different ($p < .05$). The interaction is therefore driven by greater increases in compassion ratings due to practice (Early vs. Late) as the difficulty of the target increases (greatest increases for Difficult Person).

REP: Average effect of reappraisal. Averaging across all strategies over 11 days of training, REP was successful in reducing intensity of negative feelings compared to baseline (Baseline intensity = 42.32, Average reappraisal intensity = 30.46, $t_{20} = 3.9$, $p < .001$). Each individual strategy was successful in reducing negative feelings (all p's < 0.05), and each successive strategy reduced more negative emotion than the previous strategy (Strategy 1 > Strategy 2 > Strategy 3, all p's < 0.01).

REP: Effect of time. To test whether REP decreased negative emotions over the training period, we averaged ratings from the early period (first 3 days) and late period (last 3 days) of training. Although trending in the expected direction, no significant differences were found in Early vs. Late period emotion ratings in either the average of all strategies or each individual strategy (all p's > .10). This may be because REP trainees rarely repeated the same stressful event over 11 days of training. Within each training day, negative emotions were reduced, but they were not reduced over time as new events were reappraised. However, COM may show improvement over time because trainees used the same example within a target category more frequently (e.g., repeatedly using "coworker" as your difficult person). The number of repeats within each target category in COM was significantly greater than the number of repeats of stressful events in REP (all p's < 0.001). Both groups reported greater focus and improvement in regulatory strategies comparing Early vs. Late training periods (all p's < 0.05).

Procedure

Visit 1: mock fMRI scan. Participants signed consent forms to participate in the training study as well as the economic game study. The economic game portion of the study was presented as a separate study from the training study to reduce demand characteristics on game behavior.

Participants received instructions for the in-scanner emotion regulation task that were specific to their group. Specifically, the COM group was taught to use compassion, and the REP group was taught to use reappraisal in response to the social pictures in the scanner: COM trainees were instructed to feel compassion for the people in the images and to pay attention to bodily sensations, especially around the heart. They were told to repeat the phrase "May you be free from your suffering. May you have joy and happiness" to help them evoke compassionate feelings. REP trainees were instructed to reinterpret the meaning of the pictures to decrease negative emotions. These strategies were similar to those used in other studies of cognitive reappraisal (Ochsner & Gross, 2005; Urry et al., 2006), and included reinterpreting the situation in the picture (e.g., an injured person will receive help and make a full recovery) and reinterpreting their relationship to the individual in the picture (e.g., imagining they are a neutral bystander).

After learning their assigned strategy, participants completed a mock fMRI scan in which they were introduced to the scanner environment and practiced a short version of the task. They regulated their response to 8 picture trials (4 negative and 4 neutral), and verbalized the COM or REP regulation strategies so the experimenter could verify that they understood the task. Social images were chosen from the International Affective Picture System (Lang, Bradley, & Cuthbert, 1999) that represented SUFFERING (negative images) and NEUTRAL (neutral images) scenes.

Visit 2: Pre-Training fMRI Scan. Participants completed the pre-training fMRI scan (see Methods in main manuscript) approximately one week after their mock fMRI session. They employed the assigned training strategy (COM or REP) in the fMRI experiment.

Visit 3: Post-Training fMRI Scan. After completing two weeks of training, participants performed the identical fMRI task with the parallel image set (see Methods in main manuscript).

Visit 3: Altruistic Redistribution Task. See Page 15 for the full description of the Redistribution Game, which was validated in an independent sample and shown to be associated with individual differences in trait compassion.

Participants played a variant of the Redistribution Game in which only one unfair offer was observed. The task was presented as an on-going study in the laboratory, independent of the training study. They were brought to a room with a computer after the post-training scan and instructed that they would be playing with live players who were located in a different building over the Internet. They read the instructions on the game website, which described the task in purely economic terms and did not use judgmental language such as "victim", "compassion," or "reappraisal". They were not asked to use their training in any way, and they were instructed that they were entirely free to make whatever decision they wished. Experimenters asked participants to describe the game in their own words to verify that they understood the rules. They understood that the decision they made would directly affect their payment in the study.

Because compassionate behavior is specifically evoked by unfairness, we ensured that all participants observed the same unfair dictator offer. The game website was programmed to reveal an unfair dictator transfer of 10/100 points to the victim (this is the paradigm depicted in Fig. 1a and b, with points converted into dollars). Because individuals behave differently if they are not playing with live players, deception was used to convince participants that they were playing with live players in real time. Before game play, the experimenter made a mock call to another "scientist" who was running the other players in a different location. The experimenter pretended to converse with the scientist, who requested some more time for the other players to read the instructions. The experimenter waited a few minutes, then confirmed that all players were ready. The experimenter left the room, giving the participant complete privacy to make his/her decision. Only one trial of the game was played. At the end of the session, participants were debriefed and asked, "When you were playing the game, did you believe you were playing against real people?" Only participants who believed the paradigm (responded "yes" to the question) were analyzed for this study (N = 41/56; see **Table S2** for statistics comparing Believers vs. Non-Believers of the redistribution paradigm).

Behavioral data analysis. Across all participants, the redistribution response was positively skewed (skewness = 1.5, SE = .37), and two participants qualified as outliers (1 from each group spent 40/45 points which was > 3 SDs from the population mean). Because of these violations of normality, we rank-ordered the behavioral response across both groups so that strong assumptions were not made about the scaling or normality of the residuals. Parametric statistics were then performed on the ranked data (Conover & Iman, 1981). To test the mean difference between groups, an independent samples t-test was performed on the redistribution ranks. When analyzed with the fMRI data, the ranks were mean-centered to facilitate interpretation of the parameter estimates.

Interpretation of redistribution behavior and ranks. Rank-ordering the redistribution data may reduce interpretability, so we further describe how to interpret the redistribution metric and the ranks. One issue is the question of the "optimal" redistribution response. Is compassion linearly related to greater altruistic redistribution, or associated with a certain outcome of redistribution such as an equal distribution between the dictator and the victim? To address this, we use the data from the Validation study (see page 14 for more methodological details). We computed two redistribution metrics to associate with trait compassion: 1) redistribution *percentage* (ranges from 0-100%, see page 15) which assumes greater compassion will be associated with linear increases in redistribution, and 2) redistribution *equality* which assumes that compassion will be associated with the equality of the distribution between the dictator and the victim (100 = perfect equality where both players have $5, 0 = perfect inequality where one player has $10 and the other has $0). For example, we can contrast the two "optimal" responses for each metric using the case where the Dictator has $9 and gives $1 to the Victim (this is the offer seen by participants in the Training study). For the redistribution percentage metric, the optimal response would be the maximum amount the participant could spend constrained by how much the Dictator had spent ($4.50/$4.50 = 100%). For the redistribution equality metric, the optimal response would be to spend the amount of money that achieves an equal distribution between the Dictator and the Victim (the participant spends $2 which results in both the Dictator earning $5 and the Victim earning $5, equality = 100).

Correlating these metrics with trait compassion (empathic concern subscale of the Interpersonal Reactivity Index; Davis, 1980) in the Validation study demonstrates that redistribution percentage is more highly associated with trait compassion than redistribution equality (percentage $r_{70} = 0.43$, $p < 0.001$, **Fig. S2a**; equality $r_{70} = 0.29$, $p < 0.05$, Fig. **S2b**). This suggests that compassionately-motivated redistribution behavior is associated with the equality of the distribution, but is best represented by participants spending as much of their endowment as possible. This results in greater funds for the Victim even at the expense of the dictator. To emphasize this point, participants who redistribute over and above equality (Victim $ > Dictator $, $n = 9$) report greater trait empathic concern than those who redistribute at or below equality (Victim $ = Dictator $ or Victim $ < Dictator $, $n = 63$) ($t_{70} = 2.56$, $p = 0.01$). Participants who redistribute over and above equality are shaded in black in **Fig. S2a** and **S2b**. Redistribution equality was also computed as a percentage metric (this takes into account how much the dictator contributed to equality) and this yielded a similar correlation as the raw redistribution equality metric ($r_{70} = 0.28$, $p < 0.05$).

Although we demonstrate with the Validation study that redistribution percentage is more associated with compassion than redistribution equality, we aid interpretation of the ranked data in the Training study by indicating where redistribution equality is represented in the data (participants spend $2, rank = 35.5/41). This is indicated by a dashed line in **Fig. 2b, 2d**, and **3c** in the main manuscript. In addition, we describe the average levels in COM and REP in terms of distribution equality. After two weeks of training, COM participants behaved more altruistically in the Redistribution Game compared to REP participants (**Fig. 1c;** independent sample $t_{39} = 2.09$, $p < 0.05$, $d = 0.65$). The mean rank out of 41 participants for COM was 24.725 or $1.14, where the REP mean rank was 17.45 or $0.62. Participants needed to spend $2 to achieve an equal distribution between the Dictator and Victim. COM trainees spent 1.84 times more money than REP, where the $1.14 spent by COM made the distribution between the Dictator and the Victim 57% more equitable (inequality of $3.44 instead of $8), while the $0.62 spent by REP only increased the equality by 31% (inequality of $5.52 instead of $8).

Behavioral comparison with Validation Study. In the Training study, participants who trained in two weeks of COM chose to redistribute more funds compared to REP ($t_{39} = 2.09$, $p < 0.05$, $d = 0.65$). In this design, redistribution was only measured post-training, so it remains unclear how redistribution levels changed compared to pre-training levels in each group. To address this, we compared redistribution after COM and REP to average redistribution in the Validation Study. As a population with no training experience, redistribution levels in the

Validation sample can serve as an estimation of pre-training redistribution. To compare redistribution responses across groups, redistribution percentages (see Page 16 for the rationale of using a percentage metric in the Validation study) were ranked across the full sample (N = 113). Statistics were performed on these ranks to compare (1) COM vs. Validation to determine whether COM increased redistribution compared to a sample with no training, (2) REP vs. Validation to determine whether REP impacted redistribution compared to a sample with no training, and 3) COM vs. REP using the new ranks to confirm the original behavioral finding.

To make these comparisons, methodological differences between the Training and Validation studies were taken into account (see Page 14 for full methodological details of the Validation study). In the Validation sample, participants played with live players that were visible in the same experimental room, whereas in the Training sample, participants played alone with no live players visible. Therefore, participants in the Validation study were likely more influenced by social desirability than those in the Training study. This was confirmed by the finding that greater social desirability (as measured by the Marlowe-Crowne Social Desirability scale; Marlowe & Crowne, 1960) significantly predicted more redistribution in the Validation sample ($r_{70} = 0.32$, $p < 0.01$), but not in COM ($r_{18} = -0.31$, $p = 0.18$) or REP ($r_{19} = -0.04$, $p = 0.88$) groups. Average social desirability also differed by group, where COM participants reported more social desirability than Validation participants ($t_{90} = 2.20$, $p < .05$). No social desirability differences were found between REP and Validation or COM and REP (p's > 0.12).

To address the influence of social desirability, we controlled for both the main effect of social desirability and the interaction of group × social desirability when comparing redistribution responses between Training and Validation groups (Tests 1 and 2). We did not control for social desirability when comparing COM vs. REP (Test 3) because social desirability was not associated with their redistribution behavior. We first performed two planned comparisons of COM vs. Validation and REP vs. Validation using hierarchical linear regressions. To test whether COM shows greater redistribution than Validation, we entered the main effect of Social Desirability and the interaction of Group × Social Desirability in Step 1. The Group variable (COM, Validation) was entered in Step 2 and was found to predict an additional 5.9% variance ($F_{3,89} = 5.80$, $p < 0.05$) over and above the other factors. The Group × Social Desirability interaction also predicted significant variance in Redistribution ($t = -2.46$, $p < 0.05$), but the main effect of Social Desirability was not significant. Computing the covariate adjusted means showed that COM gave significantly more in the Redistribution Game than the Validation sample, suggesting that COM increases redistribution compared to a sample with no training (COM adjusted mean rank = 99.43, Validation adjusted mean rank = 28.05, see **Fig. S3a**).

In the analogous test comparing REP vs. Validation, with Social Desirability and the Group × Social Desirability interaction entered in Step 1, the Group variable (REP, Validation) entered in Step 2 did not predict any additional variance in predicting redistribution ($F_{3,90} = 0.66$, $p = 0.42$, **Fig. S3a**). This demonstrates that REP does not impact redistribution compared to average responses in participants with no training. Neither Group nor Group × Social Desirability variables were significant (p's > 0.14). When comparing COM vs. REP with the newly-computed ranks for the full sample using an independent sample t-test, we again found that participants who practiced two weeks of COM redistributed more funds than those trained in REP (COM mean rank = 63.53, REP mean rank = 44.64, $t_{39} = 2.05$, $p < 0.05$; **Fig. S3b**). Social Desirability was not in this model because it does not predict redistribution in the Training sample. Overall, these findings show that COM increases redistribution compared to a sample with no training (which putatively represents redistribution at pre-training) as well as REP, whereas REP showed no difference with the sample with no training. This suggests that COM specifically increases altruistic behavior compared to pre-training levels, whereas REP is having no effect.

Functional magnetic resonance imaging (fMRI) data acquisition. Whole-brain functional and anatomical images were acquired using a General Electric 3 Tesla MRI scanner (GE Medical Systems, Waukesha ,WI) with LX software (version ESE12M4), a transmit-receive quadrature birdcage coil, and Nvi (40mT/m; 150 mT • m^{-1} • ms^{-1} slew rate) gradients. Functional images were acquired using a T2*-weighted gradient-echo, echo planar imaging (EPI) pulse sequence [30 sagittal slices, 4mm thickness, 1 mm interslice gap; 64 × 64 matrix; 240 mm field of view (FOV); repetition time (TR)/echo time (TE)/Flip, 2000 ms/30 ms/90º; 146 whole-brain volumes per block]. A high-resolution T1-weighted anatomical image was also acquired (T1-weighted inversion recovery fast gradient echo; 256 × 256 in-plane resolution; 240 mm FOV; 124 × 1.2 mm axial slices).

fMRI data analysis.

Pre-Processing. Image analysis was performed with AFNI (Cox, 1996) unless otherwise noted. Data were slice-time corrected and motion corrected with realignment to the first volume. They were then field map corrected (Jezzard & Clare, 1999) using prelude (Smith et al., 2004) from FSL and in-house software (http://brainimaging.waisman.wisc.edu/~jjo/fieldmap_correction/fieldmap_correction.html), and spatially smoothed using a Gaussian kernel with a full width at half maximum of 6 mm. Anatomical images from both time points were first normalized to the Montreal Neurological Institute (MNI152) template using an affine transformation with FLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), and then re-normalized to the MNI152 template using a nonlinear algorithm implemented by FNIRT in FSL. The resulting warp matrices were applied to the functional data (re-sampled to 2 mm^3). Anatomical images were averaged across study participants for display purposes.

Addressing issues of non-independence. It is worth emphasizing that our analytic strategy is in accord with consensual recommendations about non-independent or 'circular' analyses (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Poldrack & Mumford, 2009; Vul, Harris, Winkielman, & Pashler, 2009). Data analyses were constrained to *a priori* anatomically-defined regions of interest (ROIs), and corrected for multiple comparisons within ROIs in each hemisphere or the entire brain. For hypothesis testing purposes, second-level analyses of mean and individual differences (correlations with redistribution behavior) were conducted simultaneously using orthogonal contrasts. Follow-up analyses aimed at decomposing omnibus effects relied on cluster means (percentage signal change averaged over all suprathreshold voxels)—not peaks. Nevertheless, it is important to emphasize that the mechanism for selecting voxels and that for estimating effect sizes were not independent, leading to some degree of bias in the estimates. Consequently, scatter plots and the accompanying effect size estimates (Pearson's *r*)—estimated using cluster averages—are provided for descriptive and diagnostic purposes only (i.e., ensuring that the data adequately satisfy inferential test assumptions).

Interaction Analysis.

GLM 1: First Level (Subject) Analysis. Functional data were modeled with a general linear model (GLM). Canonical hemodynamic response functions (3dDeconvolve's 'GAM' function) were convolved with delta functions representing stimulus onsets for each condition

(SUFFERING, NEUTRAL) and the initial instruction. A second-order polynomial was used to model the baseline and slow signal drift, and the first two whole-brain volumes were excluded from the GLM to ensure steady-state magnetization.

To isolate activity specific to negative trials, beta coefficients were computed from a general linear test of SUFFERING-NEUTRAL, and were converted to percent signal change (PSC: 100 × β/baseline). PSC maps were computed for both pre- and post-training fMRI scans, and training-induced neural changes were computed by subtracting the pre- from the post-training map.

Second (Group) Level Analysis. In accord with the study's aims, we identified regions where training-induced changes in activity were differentially related to altruistic redistribution across groups, by estimating the Group (COM, REP) × Redistribution (mean-centered rank) interaction using the PSC differences described above. The group-level GLM also modeled the main effects of Group and Redistribution using 3dRegAna in AFNI. All group-level statistical tests were first thresholded at a voxelwise level of $p < 0.01$, and correction for multiple comparisons were performed after this initial thresholding. Although mean-centered redistribution ranks were used in the statistical tests, figures display raw ranks for ease of interpretation for the reader.

Whole brain analysis. To search for significant regions within the entire brain, the interaction map was corrected for multiple comparisons at $p < 0.01$ within a whole-brain ROI using cluster-extent thresholding based on Monte Carlo simulation. This was accomplished using AlphaSim in AFNI and an estimate of the spatial autocorrelation derived from single-subject residuals. The whole-brain test identified the right inferior parietal cortex (IPC; **Fig. 2a**, see **Table S3**). The $p < 0.01$ correction threshold was chosen by setting the corrected p value for each test at $p < 0.05$ and dividing by the number of whole-brain tests and ROIs (5 total: whole-brain tests of Group × Redistribution interaction test and IPC conjunction test, ROIs of amygdala/hippocampus, insula, and nucleus accumbens; see below for more details). The $p < 0.01$ corrected threshold is used for subsequent tests.

Region of Interest (ROI)-based analyses. After whole-brain interrogation, we investigated *a priori* regions of interest (bilateral ROIs of amygdala, insula, nucleus accumbens) with the Group × Redistribution interaction test. ROIs for the insula and nucleus accumbens (NAcc) were taken from the Harvard-Oxford probabilistic atlas (Desikan et al., 2006) provided in FSL (atlas thresholded at 50%). The amygdala ROI was taken from the Juelich probabilistic atlas (conjunction of basolateral, centromedial, and superior amygdala regions; Amunts et al., 2005) provided in FSL (atlas thresholded at 50%).

All statistical maps were corrected for multiple comparisons at $p < 0.01$ using cluster-extent thresholding based on Monte Carlo simulation within each ROI (after voxelwise thresholding at $p < 0.01$). This was accomplished using AlphaSim in AFNI and an estimate of the spatial autocorrelation derived from single-subject residuals. None of the ROI analyses survived at $p < 0.01$. See *Exploratory ROI Analyses* below for additional information.

IPC network conjunction analysis. The IPC has been implicated in shared representations of others' pain (Lamm, Decety, & Singer, 2011) as well as the human mirror neuron system (Gallese, Keysers, & Rizzolatti, 2004). To follow up on this finding, we investigated what other regions the IPC may be connected with in order to increase redistribution in COM vs. REP. We performed a conjunction test that required neural change voxels to be (1) correlated with changes in IPC activation, and (2) differentially correlated with altruistic redistribution between COM and REP groups. These voxels were identified by subjecting the POST-PRE change scores to a conjunction test of (1) a regression with IPC change scores (using the extracted average PSC values from the IPC cluster; thresholded voxelwise at $p < 0.01$), and (2) a Group × Redistribution interaction (thresholded voxelwise at $p < 0.01$). Because this conjunction analysis consists of two independent tests, the initial voxelwise thresholding on the conjunction map was set at $p < 0.001$ ($p < 0.01$ x $p < 0.01$) and then was corrected for multiple comparisons at $p < 0.01$. A region in the dorsolateral PFC (DLPFC) survived this test, whole-brain corrected at $p < 0.01$ (**Fig. 2c-d**). The fact that the IPC and DLPFC are functionally correlated and both differentially predict greater redistribution in COM vs. REP suggests that a fronto-parietal executive control network (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008) may contribute to altruistic changes in COM. Because of the rich literature that implicates connectivity between the PFC and subcortical regions as a neural substrate of emotion regulation (Johnstone, van Reekum, Urry, Kalin, & Davidson, 2007; Ochsner & Gross, 2005; Urry et al., 2006; Wager, Davidson, Hughes, Lindquist, & Ochsner, 2008), we performed subsequent connectivity analyses using the DLPFC as the seed region (see below). For descriptive purposes of the relationship, average PSC values were extracted from the IPC and DLPFC clusters and correlated in SPSS 18.0 (SPSS, Chicago, IL). Changes in the IPC and DLPFC were highly correlated across all subjects ($r_{39} = 0.86$, $p < 0.001$) as well as in each group (COM $r_{18} = 0.92$, $p < 0.001$; REP $r_{19} = 0.79$, $p < 0.001$).

Decomposing the interaction. In order to decompose significant interactions, mean PSC change scores (i.e., averaged across suprathreshold voxels) were extracted from the clusters for each participant. These values were analyzed in SPSS using multiple regression to yield parameter estimates and determine the directionality of the relationship for each group (Tables S4 and S5). These values were also used for descriptive and diagnostic purposes (see ‘Addressing issues of non-independence’ above).

DLPFC functional connectivity. *Intra-individual analysis: Psychophysiological interactions (PPI).* We performed analyses to see if changes in intra-individual DLPFC effective connectivity with the three ROIs (amygdala, insula and NAcc) would also predict greater redistribution in COM vs. REP. This would demonstrate that DLPFC connectivity within individual subjects during voluntary generation of compassion, specifically in the context where suffering is encountered, impacts subsequent altruistic behavior. We used the psychophysiological interactions (PPI) technique (Friston et al., 1997) with the DLPFC cluster identified in the IPC conjunction test as the seed region (**Fig. 3a**).

A new first-level GLM was computed for each participant. This contained the same regressors as GLM 1, as well as additional regressors that modeled the entire DLPFC time series, and the interaction of DLPFC activity in SUFFERING vs. NEUTRAL trials. The PPI regressor thus modeled task-related connectivity with the DLPFC, independent of general activation to the task and intrinsic connectivity with the DLPFC across conditions. PPI beta coefficients were computed for each subject at each time point and normalized to the MNI152 template. Training-induced effects were isolated by subtracting Pre-training from Post-training coefficients. At the second level, we determined whether changes in DLPFC-insula or DLPFC-NAcc connectivity predicted redistribution differentially between COM and REP groups.

Group-level Interaction Analysis. A group-level Group × Redistribution interaction was computed with ranked, mean-centered redistribution data as a predictor of the PPI change scores, controlling for the main effects of Group and Redistribution (using the AFNI regression package 3dRegAna). Similarly to previous analyses, the interaction map was initially thresholded voxelwise at $p < 0.01$ within each ROI, then was corrected for multiple comparisons using

Monte Carlo simulations at a threshold $p < 0.01$. The NAcc survived correction within the ROI **Fig. 3b**; **Table S3**). The insula and amygdala ROIs were not significant. We extracted mean cluster PPI change scores for each participant, and for descriptive and diagnostic purposes, these were used to generate parameter estimates in SPSS (**Table S4 and S5**) and scatterplots (e.g., to ensure that results were not driven by outliers).

Arousal and Valence correlations with redistribution. To test the hypothesis that decreased distress in COM predicts altruistic behavior, we tested whether decreases in arousal or valence correlated with greater redistribution in each group. Ratings change scores were computed as POST(SUFFERING-NEUTRAL) - PRE(SUFFERING-NEUTRAL) to match the neural change metrics. We found that changes in arousal predicted redistribution in COM ($r_{18} = -0.45$, $p < .05$) but not REP ($r_{19} = 0.09$, $p = 0.70$). Changes in valence only marginally predicted redistribution in COM ($r_{18} = -0.42$, $p = .07$) and was not correlated in REP ($r_{19} = 0.13$, $p = 0.58$), so follow-up analyses were only conducted with arousal. No group mean differences in arousal change ($t_{39} = 1.18$, $p = 0.25$) or valence change ($t_{39} = 0.06$, $p = 0.95$) were present.

Arousal rating correlations with brain changes. We tested whether training-induced changes in neural activation were associated with changes in self-reported arousal. Arousal change scores were computed as POST(SUFFERING-NEUTRAL) - PRE(SUFFERING-NEUTRAL) to match the PSC and PPI change metrics. Arousal change scores were correlated with change metrics in each region identified (cluster extracted average) in the previous analyses for each group in SPSS (see p. 13 for results).

Image visualization. Brain images were re-sampled to 1 mm^3 for visualization purposes only. All images are displayed using MRIcron (http://www.cabiatl.com/mricro/).

Supplementary Analyses and Results

Exploratory ROI analyses. Because of the strong *a priori* hypotheses about the amygdala and insula, the ROIs were further inspected for exploratory purposes. After voxelwise thresholding the Group × Redistribution interaction test at $p < 0.01$, we examined whether activated voxels in the amygdala and insula also extended to neighboring regions. Activated voxels within the left amygdala (basolateral region) extended into the left hippocampal entorhinal cortex as defined by the Juelich probabilistic atlas thresholded at 50% (Amunts et al., 2005). If combined into a single cluster, activated voxels within the left amygdala and hippocampal entorhinal cortex survived correction at $p < 0.05$ within the combined bilateral ROI of both regions. The cluster did not survive correction at $p < 0.01$. Tentatively, this suggests that training-induced changes in the left amygdala/hippocampal entorhinal cortex were differentially associated with altruistic redistribution in COM vs. REP (**Fig. S4a**; **Table S3**).

Although these exploratory results should be interpreted with caution, we further investigated the directionality of the association in each group. Cluster averages were extracted from each group and analyzed in SPSS for descriptive purposes only. In COM, trainees who were more altruistic after training showed larger decreases in amygdala/hippocampal entorhinal cortex activation in response to images of suffering. In REP, the opposite relationship was found in the region (decreased activation was associated with less redistribution; **Fig. S4b**; **Table S4**). The amygdala has been associated with processing negative emotions (Zald, 2003), and together with the hippocampus has been implicated in processing emotional memories (Phelps, 2004). Further, the hippocampus may be involved in compassion because the process likely uses reflection on one's own experiences (Immordino-Yang & Singh, 2011).

Because emotion regulation is thought to involve connectivity between the PFC and amygdala (Urry et al., 2006; Wager et al., 2008), exploratory analyses also tested whether changes in DLPFC activation were correlated with changes in amygdala/hippocampal entorhinal cortex activation across participants. In COM, increased DLPFC activation was correlated with decreased amygdala/hippocampal entorhinal cortex activation ($r_{18} = -0.47$, $p < 0.05$), where there was no significant relationship in REP ($r_{19} = -0.14$, $p = 0.55$). The anti-correlation in COM may reflect emotion regulation that dampens the aversive conditioning evoked by another's suffering in order to promote altruistic behavior. However, these results should be interpreted with caution, and future studies may confirm whether these regions are consistently altered through compassion training.

Inspection of the activated voxels in the insula from the Group × Redistribution interaction test yielded no further activation in neighboring regions. PPI analyses (voxelwise thresholded at $p < 0.01$) also yielded no further activation in neighboring regions of the amygdala and insula.

Effect of training on arousal ratings. We found that changes in neural responses to suffering due to training were associated with redistribution, so we first analyzed the arousal ratings using analogous change scores (POST[SUFFERING-NEUTRAL]-PRE[SUFFERING-NEUTRAL]). We found that decreases in reported arousal to the images were correlated with greater redistribution in COM ($r_{18} = -0.45$, $p < .05$) but not REP ($r_{19} = 0.09$, $p = 0.70$). When comparing average arousal change scores between groups, the groups did not differ ($t_{39} = 1.18$, $p = 0.25$). These data suggest that although the groups are equivalent in arousal change on average, when looking on an individual differences level, participants who decrease arousal are more altruistic in COM but not REP. This suggests that decreasing one's personal distress to others' suffering in COM (where the goal is to help another), but not REP (where the goal is to help oneself) leads to greater altruistic behavior.

To further understand how both trainings impact arousal, we decomposed the change scores by looking at the means in each condition (PRE, POST, SUFFERING, NEUTRAL). As expected, there was a main effect of Condition where SUFFERING images were rated as more arousing than NEUTRAL images ($F_{1,39} = 174.34$, $p < 0.001$). There was a main effect of Group ($F_{1,39} = 11.62$, $p < 0.01$), where COM rates images as more arousing than REP (collapsed across Time and Condition). A Group × Time interaction was also significant ($F_{1,39} = 4.57$, $p < 0.05$), where REP significantly decreases arousal to all images after training (PRE mean = 4.64, POST mean = 4.23, $p < 0.05$), but COM did not change (PRE mean = 5.35, POST mean = 5.46). The Group × Time × Condition was not significant, suggesting that the previous pattern of results was similar in both SUFFERING and NEUTRAL conditions.

These data suggest that overall, a compassion regulation strategy increases arousal to both SUFFERING and NEUTRAL images compared to reappraisal. Increased arousal is likely due to increasing the importance of and focus on others' well-being, particularly if they are suffering. However, too much arousal is sub-optimal because it may result in personal distress (Batson, 1991; Eisenberg, Fabes, & Spinrad, 2006) and take away cognitive resources to help (Goetz, Keltner, & Simon-Thomas, 2010), so participants who are able to down-regulate their arousal after COM are the ones who are most altruistic. In REP, arousal may be too low after decreasing one's own negative emotions, and altruistic behavior is decreased compared to COM (and no different from individuals with no training). This pattern of results may suggest that an optimal level of arousal in response to other people's suffering is needed to support altruistic behavior.

Arousal rating correlations with brain changes. We tested whether training-induced changes in neural activation were associated with changes in self-reported distress. Changes in arousal were not associated with changes in DLPFC activation (COM r_{18} = -0.29, p = 0.22; REP r_{19} = -0.27, p = 0.24) or IPC activation (COM r_{18} = 0.36, p = 0.12; REP r_{19} = -0.30, p = 0.21), but did significantly correlate with DLPFC-NAcc connectivity in COM (r_{18} = -0.64, p < 0.01; REP r_{19} = -0.13, p = 0.59). As an exploratory analysis, changes in amygdala-hippocampal entorhinal cortex activition were not correlated with changes in arousal in COM (COM r_{18} = 0.36, p = 0.12) or REP (r_{19} = 0.29, p = 0.20).

Tests of Mediation. Because we found associations between changes in multiple brain regions and redistribution (i.e., IPC and DLPFC, IPC and redistribution), we tested several models using the Sobel test of mediation (Preacher & Hayes, 2004) in each group. This test provides a means of calculating and testing the significance of the indirect effect of a predictor variable (IPC change) on a criterion variable (redistribution) by way of a hypothesized mediator (DLPFC change). This mediation test was not significant in either group (COM Z = 0.31, p = 0.76; REP Z = -1.44, p = 0.15). We also tested whether changes in arousal mediated the relationship between changes in DLPFC-NAcc connectivity and redistribution. The mediation test was not significant in either group (COM Z = 0.89, p = 0.38; REP Z = -0.07, p = 0.95).

Directionality of DLPFC-NACC PPI connectivity. To investigate the directionality of PPI between the DLPFC and NAcc, PPI betas were extracted from the NAcc cluster from both groups and PRE and POST time points. Right DLPFC-NAcc connectivity was on average negative before training (COM beta = -0.58, REP beta = -1.17), indicating greater connectivity in response to neutral images compared to negative images, and positive after training (COM beta = 1.85, REP beta = 1.57), indicating greater connectivity in response to negative images compared to neutral images. No group differences in connectivity were found at either time point. When looking at both groups together, DLPFC-NAcc connectivity was marginally increased pre to post-training (t_{40} = 1.94, p = 0.06). Overall, DLPFC-NAcc connectivity showed subtle increases from negative to positive connectivity after both trainings, but these increases were only associated with greater altruistic redistribution in COM. Increases in connectivity in REP were associated with decreased altruistic redistribution (**Fig. 3**).

Redistribution Game: Independent Validation Study

Rationale. In order to test whether compassion training impacts altruistic behavior outside of the training context, we developed an economic decision-making task that would be sensitive to individual differences in trait compassion. According to psychological theories of compassion, compassion is evoked in response to suffering or unfair treatment, and results in a desire to help (Goetz et al., 2010). These components were modeled using design features of the third-party punishment game, where a participant views “suffering” when a dictator violates the social fairness norm and distributes an unfair amount of funds to a powerless recipient (Fehr & Fischbacher, 2004). Both helping and punishment may be associated with compassion because helping a victim is an important behavioral outcome of compassion, and punishment decreases future norm violations and promotes cooperation (Fehr & Fischbacher, 2003). Therefore, both helping and punishment behaviors were incorporated into the Redistribution Game. To determine that the Redistribution Game was a valid behavioral measure of compassionately-motivated behavior, we investigated whether game behavior in an independent sample would be positively associated with trait compassion. If this association was found, we reasoned that the

Redistribution Game would be a valid means of assessing whether compassion training alters altruistic behavior after witnessing unfair treatment.

Redistribution Game. The experimenter endows a dictator with 100 points, a recipient with 0 points, and a participant with 50 points. In the first interaction of the game, the dictator may choose to transfer any number of the 100 points to the recipient, while the participant observes. In the second interaction, the participant may choose to spend points in order to redistribute funds from the dictator to the recipient. Each point spent by the participant results in two points taken from the dictator and given to the recipient. Participants can redistribute any amount without exceeding the value of the dictator's remaining points after transferring to the recipient. When the game is over, points are converted to dollars (10 points = $1), and each player is paid based on the number of points acquired. Therefore, decisions directly affect monetary outcomes.

Questionnaires. To assess whether redistribution behavior is predicted by trait compassion, participants completed the Empathic Concern subscale of the Interpersonal Reactivity Index, which measures the tendency to feel warmth, compassion and concern for others undergoing negative experiences (Davis, 1980). An example item is, "I often have tender, concerned feelings for people less fortunate than me." They also completed the Compassion subscale of the Dispositional Positive Emotions Scale (Shiota, Keltner, & John, 2006) and the Compassionate Love Scale (Sprecher & Fehr, 2005). To control for possible confounds of social desirability and current mood, the Marlowe-Crowne Social Desirability scale (Crowne & Marlowe, 1960) and Positive and Negative Affective Scales (Watson, Clark, & Tellegen, 1988) were administered. Family income was also measured.

Participants and Procedure. 147 adults from the Madison, WI community consented to participate in the study, which was approved by the University of Wisconsin-Madison Social and Behavioral Sciences Institutional Review Board. 141 participants produced useable data (77 female), and 72 participants were included in subsequent analyses of redistribution responses to unfair dictator offers. Participants were brought to the computer laboratory in groups ($n \geq 9$), and read the instructions on the game website. Experimenters confirmed that they understood the rules of the game, and then three rounds of the game were played. Participants used a web interface to ensure that each game interaction was played 1) with live players 2) anonymously and 3) with unique participants. This design allowed for real-time interactions with live players while minimizing reputation effects. To maximize data points, each participant played in each role (dictator, recipient, third party) with the order randomized. Participants were free to choose any decision in each position, and no deception was used. Payment was determined by game outcome (average earnings ≈ $14). Trait questionnaires were completed either before or after game playing.

The Validation Study design differed from the Training Study design in that participants played 3 trials (once in each role) with live anonymous players, instead of only 1 trial as the third-party redistributer. This difference in study design is addressed in data analyses below. In addition, participants in the Training Study saw the same pre-programmed unfair offer of $1/$10 to ensure experimental control. See *Visit 3: Altruistic Redistribution Task* (Pages 5-6) for more Training Study paradigm details.

Data Analysis and Results. Redistribution was calculated as a percentage of the total possible redistribution amount, where the raw number of points was divided by the maximum points that could be spent (constrained by the remaining dictator endowment after transferring to the recipient). A redistribution score of 50%, for example, could represent spending 50/100 points as well as 40/80 points. Because compassion is evoked by unfair treatment, we constrained analyses to participants who witnessed an unfair dictator transfer (≤ 25%, n = 72, 36 female). Unfair offers have been similarly defined in other economic decision-making studies (Güth, Schmittberger, & Schwarze, 1982; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). To determine the relationship between game behavior and trait compassion, redistribution percentage was correlated with questionnaire response. All analyses were performed with SPSS.

After witnessing an unfair transfer, participants who rated themselves as more compassionate redistributed more funds (Empathic Concern $r_{70} = 0.43$, $P < 0.001$, **Fig. S2a**; Compassion $r_{70} = 0.4$, $P < 0.001$; Compassionate Love $r_{70} = 0.35$, $P < 0.005$), demonstrating that redistribution is a valid behavioral measure of compassion. Compassion scales were highly intercorrelated (all *r*'s > .74), so Empathic Concern was used for all subsequent analyses because it showed the highest correlation with redistribution. Analyses conducted on raw redistribution scores ($r_{70} = 0.41$, $P < 0.001$) or using nonparametric correlations ($rho_{70} = 0.47$, $P < 0.001$) yielded similar results.

Controlling for Confounding Variables. Redistribution did not differ between participants who completed trait questionnaires before (n = 32) vs. after the game (n = 40), $t_{70} = -0.54$, $P > .05$. Both groups showed the association between empathic concern and redistribution (*r*'s > 0.41, *P*'s < .01).

A hierarchical regression model also demonstrated that relations between redistribution and empathic concern were not primarily determined by other potentially confounding variables. On the first step, we entered measures of social desirability, behavior when playing as the dictator, family income, player order, previous experience in the game (earnings and punishment before playing as the third party), and current positive and negative affect. On the second step, we entered empathic concern. This revealed that the potentially confounding variables accounted for 27% of the variance in redistribution ($\Delta R^2 = 0.27$, $P = .01$), and empathic concern accounted for an additional 16% of variance ($\Delta R^2 = 0.16$, $P < .001$).

Supplementary References

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., Habel, U., et al. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anatomy and Embryology*, *210*(5-6), 343–352.

Batson, C. D. (1991). *The altruism question*. Hillsdale, NJ: Erlbaum.

Beck, A. T. B., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive Therapy of Depression*. New York, NY: Guilford Press.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, *35*(3), 124–129.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354.

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

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980.

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg (Vol. ed.), W. Damon & R. M. Lerner (Series eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 646–718). New York: Wiley.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*(3), 218–229.

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin*, *136*(3), 351.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388.

Immordino-Yang, M. H., & Singh, V. (2011). Hippocampal contributions to the processing of social emotions. *Human Brain Mapping*. doi:10.1002/hbm.21485

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841.

Jezzard, P., & Clare, S. (1999). Sources of distortion in functional MRI data. *Human Brain Mapping*, *8*(2-3), 80–85.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Instruction manual and affective ratings. *The Center for Research in Psychophysiology, University of Florida*.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*(5), 242–249.

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*(2), 198–202.

Poldrack, R. A., & Mumford, J. A. (2009). Independence in ROI analysis: where is the voodoo? *Social Cognitive and Affective Neuroscience*, *4*(2), 208–213.

Preacher, K., & Hayes, A. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods*, *36*(4), 717–731.

Salzberg, S. (1997). *Lovingkindness: The Revolutionary Art of Happiness*. Boston, MA: Shambhala.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science (New York, N.Y.)*, *300*(5626), 1755–1758.

Shiota, M. N., Keltner, D., & John, O. P. (2006). Positive emotion dispositions differentially associated with Big Five personality and attachment style. *The Journal of Positive Psychology*, *1*(2), 61–71.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23 Suppl 1*, S208–219.

Sprecher, S., & Fehr, B. (2005). Compassionate love for close others and humanity. *Journal of Social and Personal Relationships*, *22*(5), 629 –651.

Urry, H. L., van Reekum, C. M., Johnstone, T., Kalin, N. H., Thurow, M. E., Schaefer, H. S., Jackson, C. A., et al. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *Journal of Neuroscience*, *26*(16), 4415–4425.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology*, *100*(6), 3328–3342.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Perspectives on Psychological Science*, *4*(3), 274 –290.

Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, *59*(6), 1037–1050.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063.

Group	*N*	*Age (Years)*	*Gender (M, F)*	*Baseline Compassion*	*Training Days*	*Training Minutes*
COM	20	21.9 (6.72)	8, 12	20.8 (4.69)	11.8 (0.89)	351.7 (27.47)
REP	21	22.5 (3.2)	8, 13	18.4 (4.67)	12.1 (0.8)	352.9 (27.07)

Table S1. Training population statistics. Group differences in demographic and training variables were not significant (all *P*'s > 0.11). Standard deviations are in parentheses. Baseline compassion was assessed using the Empathic Concern subscale of the Interpersonal Reactivity Index (Davis, 1980).

	Group	
	Believers	*Non-Believers*
Total N	41	15
COM N	20	8
REP N	21	7
Gender (M, F)	16, 25	8, 7
Age	22.15 (7.00)	24.07 (5.16)
Baseline Compassion	19.29 (4.80)	19.93 (4.78)
Training Days	11.95 (0.92)	12.00 (1.07)
Training Minutes	352.30 (26.92)	356.53 (35.95)
COM Compassion Rating – First 3 Days	4.11 (0.60)	4.19 (9.84)
COM Compassion Rating – Last 3 Days	4.46 (0.92)	4.63 (0.86)
REP Emotion Rating – First 3 Days	32.70 (16.16)	43.46 (17.90)
REP Emotion Rating – Last 3 Days	26.56 (14.94)	37.05 (23.22)
Redistribution (Rank)	26.40 (15.29)	34.23 (16.68)

Table S2. Population statistics comparing participants who believed they were playing with live players in the Redistribution Game (Believers) to participants who did not (Non-Believers). Believers did not differ from Non-Believers in training group and gender N, age, baseline trait compassion, training practice time or ratings, or mean redistribution (ranked across 56 participants) (all *p*'s > 0.14). Standard deviations are in parentheses. Baseline compassion was assessed using the Empathic Concern subscale of the Interpersonal Reactivity Index (Davis, 1980). Compassion ratings are averaged across all targets (loved one, self, stranger, difficult person), and REP Emotion Ratings are emotion ratings after using reappraisal strategies (averaged across the 3 strategies).

Region	*Hemisphere*	*Anatomical Description*	*MNI Coordinates*	*Volume (mm^3)*
Interaction clusters				
IPC[a]	R	Spans the inferior parietal lobule, angular gyrus[HO], superior division of lateral occipital cortex[HO],	46, -62, 36	5280
DLPFC[b†]	R	Middle frontal gyrus[HO]	38, 22, 46	1504
Amygdala/ Hippocampal Entorhinal Cortex[c]	L	Basolateral group[J], Hippocampal Entorhinal Cortex[J]	-18, -2, -30	560
PPI cluster				
DLPFC-NAcc[d]	R		12, 8, -10	152

Table S3. Interaction clusters indicate regions where the Group × Redistribution interaction predicted the training-induced BOLD changes (POST[SUFFERING-NEUTRAL] – PRE[SUFFERING-NEUTRAL] percent signal change). **PPI cluster** indicates a region where the Group × Redistribution interaction predicted the training-induced changes (POST-PRE beta coefficients) in DLPFC functional connectivity (SUFFERING vs. NEUTRAL condition). IPC = inferior parietal cortex, DLPFC = dorsolateral prefrontal cortex, NAcc = nucleus accumbens, R = right, L = left. MNI coordinates indicate the peak voxel within the cluster.

Region corrected for multiple comparisons within:
All statistical tests are set at an initial minimum voxelwise thresholded of $p < 0.01$.
[a] A whole-brain ROI with a voxelwise threshold of $F_{3,38} = 7.36$ ($p < 0.01$) and an extent of 362 voxels (2896 mm^3; $p < 0.001$, corrected)
[b] Cluster was corrected within a whole-brain ROI after a conjunction test of (1) the across-subject correlation with IPC change scores using the full sample (voxelwise $F_{1,40} = 7.32$, $p < 0.01$) and (2) the Group × Redistribution interaction (voxelwise $F_{3,38} = 7.36$, $p < 0.01$) (conjunction voxelwise threshold at $p < 0.001$). This cluster was then corrected for multiple comparisons at an extent of 114 voxels (912 mm^3; $p < 0.01$, corrected).
[c] An exploratory bilateral amygdala-hippocampal entorhinal cortex ROI (50% Juelich atlas) with a threshold of $F_{3,38} = 7.36$ ($p < 0.01$) and an extent of 49 voxels (392 mm^3; $p < 0.05$, corrected). The cluster did not survive correction at $p < 0.01$ (73 voxels, 584 mm^3).
[d] A bilateral NAcc ROI (50% Harvard-Oxford atlas) with a threshold of $F_{3,38} = 7.36$ ($p < 0.01$) and an extent of 18 voxels (144 mm^3, $p < 0.01$, corrected).

[†] Cluster used as the seed region for the PPI analysis
[HO] Anatomical regions identified by the Harvard-Oxford atlas (Desikan et al., 2006).
[J] Anatomical regions identified by the Juelich atlas (Amunts et al., 2005).

Cluster Statistics												
	Group				*Redistribution*				*Group × Redistribution*			
Region	*B*	*SE*	*t*	R^2	*B*	*SE*	*t*	R^2	*B*	*SE*	*t*	R^2
Interaction clusters												
IPC	-0.04	0.04	-0.99	0.01	0.005	0.004	1.25	0.02	0.014	0.004	3.66	0.26***
DLPFC	-0.06	0.04	-1.47	0.03	0.003	0.004	0.97	0.01	0.015	0.004	4.14	0.31***
Amyg/Hipp^	0.007	0.13	0.05	0.002	-0.013	0.012	-1.13	0.03	-0.05	0.012	-4.31	0.33***
PPI cluster												
DLPFC-NAcc	0.23	1.65	0.14	0.000	-0.11	0.14	-0.79	0.009	0.48	0.14	3.35	0.23**

Table S4. Parameter estimates from the interaction clusters displayed in Table S3. Statistics are reported from the main effect of Group, main effect of Redistribution, and the interaction of Group × Redistribution. Parameter estimates were computed for cluster averages and are reported for descriptive and diagnostic purposes only.
^ Amyg/Hipp indicates the exploratory amygdala/hippocampal entorhinal cortex cluster.
** $P < 0.01$, *** $P < 0.001$

Unique group effects								
Region	*COM*				*REP*			
	B	*SE*	*t*	*sr*	*B*	*SE*	*t*	*sr*
Interaction clusters								
IPC	0.019	0.005	3.54	0.49***	-0.009	0.006	-1.68	-0.23
DLPFC	0.018	0.005	3.68	0.49***	-0.011	0.005	-2.20	-0.29*
Amyg/Hipp^	-0.064	0.016	-3.91	-0.52***	0.037	0.017	2.21	0.29*
PPI cluster								
DLPFC-NAcc	0.370	0.200	1.85	0.27†	-0.596	0.208	-2.87	-0.41**

Table S5. Individual differences in redistribution behavior predict training-induced changes in PSC or PPI betas: Parameter estimates reported for each group reported for descriptive and diagnostic purposes only. Parameter estimates are unbiased towards any one group because clusters were identified by the overall interaction test.
^ Amyg/Hipp indicates the exploratory amygdala/hippocampal entorhinal cortex cluster.
† $P = 0.07$, * $P < .05$, ** $P < .01$, *** $P \leq .001$

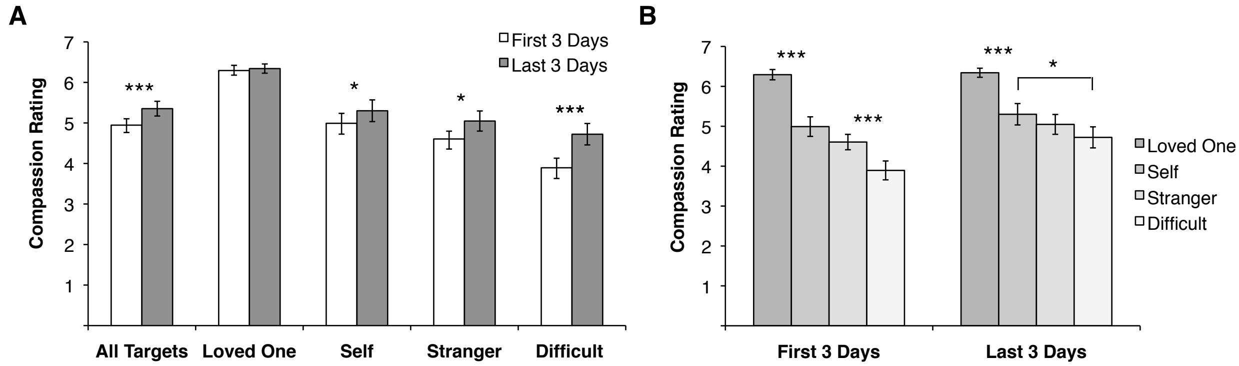

Fig. S1. COM training efficacy. (*A*) COM significantly increases feelings of compassion for each target when comparing the first 3 days to the last 3 days of practice, except for the Loved One. Compassion increased the most for the Difficult Person. (*B*) During the first 3 days of training, compassion ratings decreased as the target became more difficult (Loved One > Self > Stranger > Difficult Person). This difference became less pronounced during the last 3 days of training, where the Stranger was not rated significantly different from the Difficult Person.
* $P < .05$, *** $P < .001$

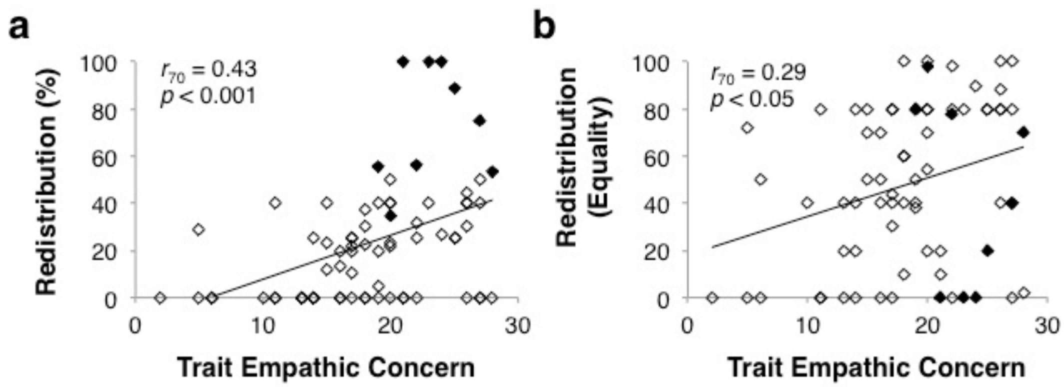

Fig. S2. Comparison of the relationship between trait empathic concern and two redistribution metrics (% and equality). *(a)* In an independent validation study, individual differences in empathic concern (Davis, 1980) predict third-party redistribution after witnessing an unfair dictator transfer (≤ 25%) in the Redistribution Game. Redistribution % represents the percentage of the maximum possible redistribution. Shaded participants represent those who redistributed over and above equality (Victim $ > Dictator $). (*b*) Individual differences in empathic concern are correlated with redistribution equality (how equitable the distribution is between the dictator and victim after redistribution), but less so than the redistribution % metric. Shaded participants represent those who redistributed over and above equality (Victim $ > Dictator $). This suggests that redistribution % is more highly associated with trait compassion than redistribution equality.

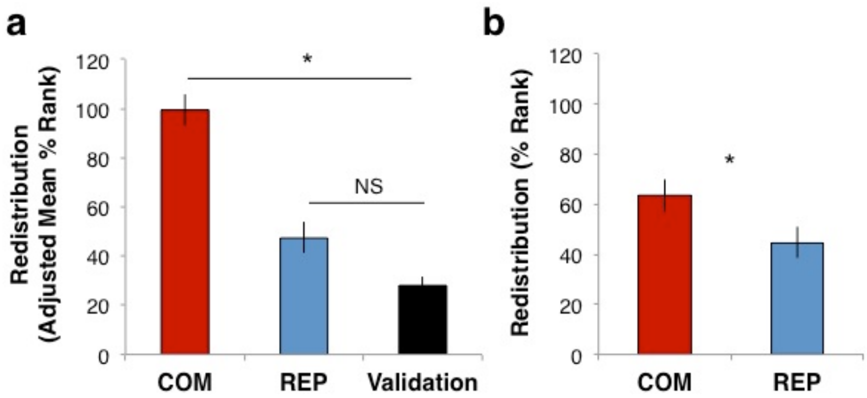

Fig. S3. (*a*) Participants trained in compassion redistribute more wealth compared to an independent validation sample with no training, controlling for the main effect of social desirability and the interaction of group × social desirability (* R^2 change = 0.059, $p < 0.05$ in a hierarchical regression, see page 7 of Supplementary Material). Participants trained in REP did not differ in redistribution compared to the validation sample (NS = not significant at $p = 0.42$). Participant redistribution % responses were ranked across the full sample from both studies (n = 113), and graphs display covariate-adjusted means of the ranks. (*b*) COM participants redistribute more funds compared to REP using the full sample ranks (COM mean rank = 63.53, REP mean rank = 44.64, independent sample $t_{39} = 2.05$,* $p < 0.05$).

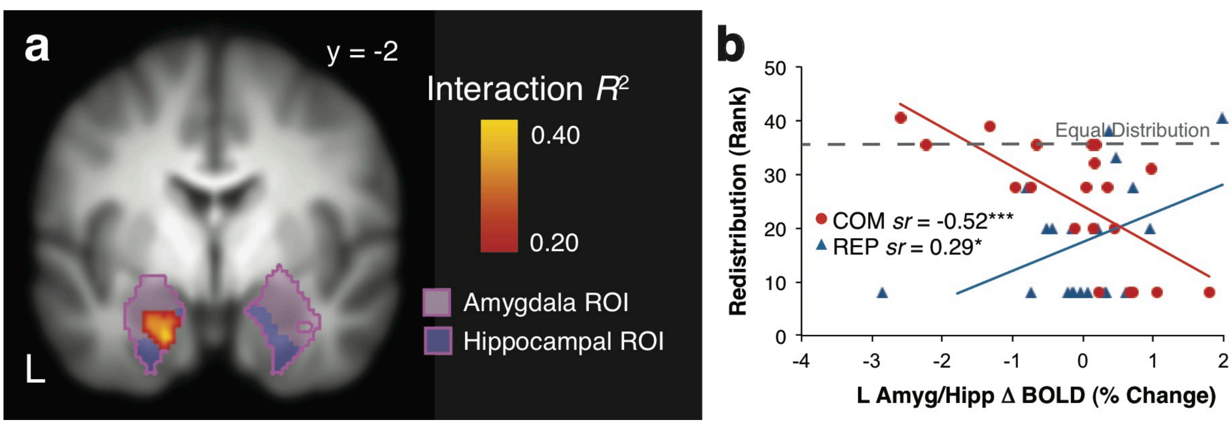

Fig. S4. Exploratory ROI analyses (all results should be interpreted with caution). (*a*) Training-induced (POST-PRE) BOLD changes in an exploratory combined ROI of left amygdala/hippocampal entorhinal cortex while regulating emotional responses evoked by images of human suffering were differentially associated with post-training altruistic redistribution in COM vs. REP ($p < 0.05$ corrected using cluster-extent thresholding based on Monte Carlo simulation within a bilateral combined amygdala/hippocampal entorhinal cortex ROI; amygdala shaded in light purple, hippocampal entorhinal cortex shaded in dark purple; Tables S3 and S4). Images and coordinates are in MNI space. Interaction R^2 indicates the proportion of variance in BOLD change accounted for by the Group (COM, REP) × Redistribution interaction. (*b*) Training-related decreases in left amygdala/hippocampal entorhinal cortex activation were associated with greater redistribution in COM ($n = 20$; *** $p <$

0.001) and less redistribution in REP (n = 21; * $p < 0.05$; Table S5). Δ BOLD (% change) in *b* indicates POST-PRE changes in brain response to human suffering (SUFFERING-NEUTRAL), and *sr* indicates the semipartial correlation of redistribution and neural change in each group. The dashed line indicates redistribution of $2 (rank = 35.5/41) which results in an equal $5 distribution between the dictator and victim.