PNAS
www.pnas.org

Supplementary Information for

Effects of Self-Transcendence on Neural Responses to Persuasive Messages and Health Behavior Change

Yoona Kang*, Nicole Cooper, Prateekshit Pandey, Christin Scholz, Matthew Brook O'Donnell, Matthew D. Lieberman, Shelley E. Taylor, Victor J. Strecher, Sonya Dal Cin, Sara Konrath, Thad A. Polk, Kenneth Resnicow, Lawrence An, & Emily B. Falk*

*Corresponding authors
Email: yoona.kang@asc.upenn.edu, emily.falk@asc.upenn.edu

This PDF file includes:

SI Methods

Hypothesis Pre-Registration. We hypothesized that the affirmation and compassion tasks, compared to the control task, would lead to increased activity in the VMPFC and PCC as well as greater behavior change. We also had hypotheses regarding neural activity related to counterarguing in the dorsolateral prefrontal cortex (DLPFC) which we report and discuss in *SI Results* below. In addition to the analyses presented in the current paper, we had also pre-registered to compare more self-enhancing values priming to more self-transcendent compassion practice. However, all but four participants in the affirmation condition ranked self-transcendent values as their top values in this study, rendering both the affirmation and compassion interventions as self-transcendent interventions, with the control condition including primarily self-enhancing values.

Baseline demographic characteristics.

	Affirmation (*n*=88)	Compassion (*n*=44)	Control (*n*=88)	Statistic (*p*)
Demographic				
Age (yrs)	32.98 (11.18)	31.91 (11.20)	35.45 (12.16)	F = 1.70 (.18)
Female	57 (64.8%)	27 (61.4%)	60 (68.2%)	X^2= 0.63 (.73)
Black	38 (43.2%)	21 (47.7%)	37 (42.0%)	X^2= 0.40 (.82)
Education (yrs)	16.00 (2.99)	15.80 (2.83)	15.81 (3.05)	F = 0.11 (.90)
Baseline Characteristics				
BMI	32.66 (5.42)	31.93 (6.26)	32.27 (7.02)	F = 0.21 (.81)
Sedentary	235.80 (114.47)	244.90 (95.09)	238.40 (105.21)	F = 0.09 (.92)

Table S1. Baseline demographic characteristics by condition. Note: Mean values and sample sizes are displayed with standard deviations and percentages, respectively, in parentheses where applicable. Baseline sedentary minutes per day were measured by accelerometer and excluding sleep and non-wear time. BMI=Body Mass Index.

Inclusion and exclusion criteria. We recruited participants from August 2014 to February 2017 in Philadelphia PA and surrounding communities. Participants responded to an online advertisement and flyers for a study on “daily activities” to avoid selection bias related to physical activity. Eligibility criteria, based on self-reports collected via online survey, included: 1) engagement in less than 200 minutes of walking, moderate, and vigorous physical activity throughout the seven days prior to the screening (using a short form International Physical Activity Questionnaire [IPAQ]), 2) a body mass index (BMI) over 25, derived from self-reported weight and height, 3) standard fMRI scanning criteria (no metal in body, not claustrophobic, not pregnant/nursing, right-handed), 4) no history of serious psychiatric/medical conditions, and 5) no current use of illicit drugs or psychotropic medications. Research assistants contacted eligible participants via phone to reconfirm their eligibility and scheduled study visits.

Sample sizes for the affirmation and control conditions were determined by power analyses based on effect sizes found in prior work (1), and the sample size for the compassion condition was determined by funding availability from an additional pilot grant. Due to these constraints, the compassion condition had half as many participants (n=44) as the affirmation (n=88) or control (n=88) conditions. Despite the unequal sample sizes, we do not observe

significant heteroscedasticity (Levene's test *p*s>.41) and all results held assuming equal variances or not.

Participants were excluded from the neural data collection, neural outcome analyses, and/or behavioral outcome analyses for the following reasons (Fig. S1): Failure to complete the fMRI study appointment (*n*=20), frontal distortion (*n*=5), excessive motion (*n*=7; 10 or more 1mm spikes and/or 4mm or higher total displacement per run), technical difficulties in scanning (*n*=3), or ineligibilities discovered after the baseline visit, including metal in body (*n*=2), brain abnormalities (*n*=7) or coronary heart diseases (*n*=1). In addition, participants who did not complete the endpoint study appointment (*n*=5) or either declined to wear accelerometers or experienced equipment failure (*n*=16) were excluded from the behavioral outcome analyses (Fig. S1) The rate of total data loss was equivalent across conditions (χ^2= 2.076, *p*=.354).

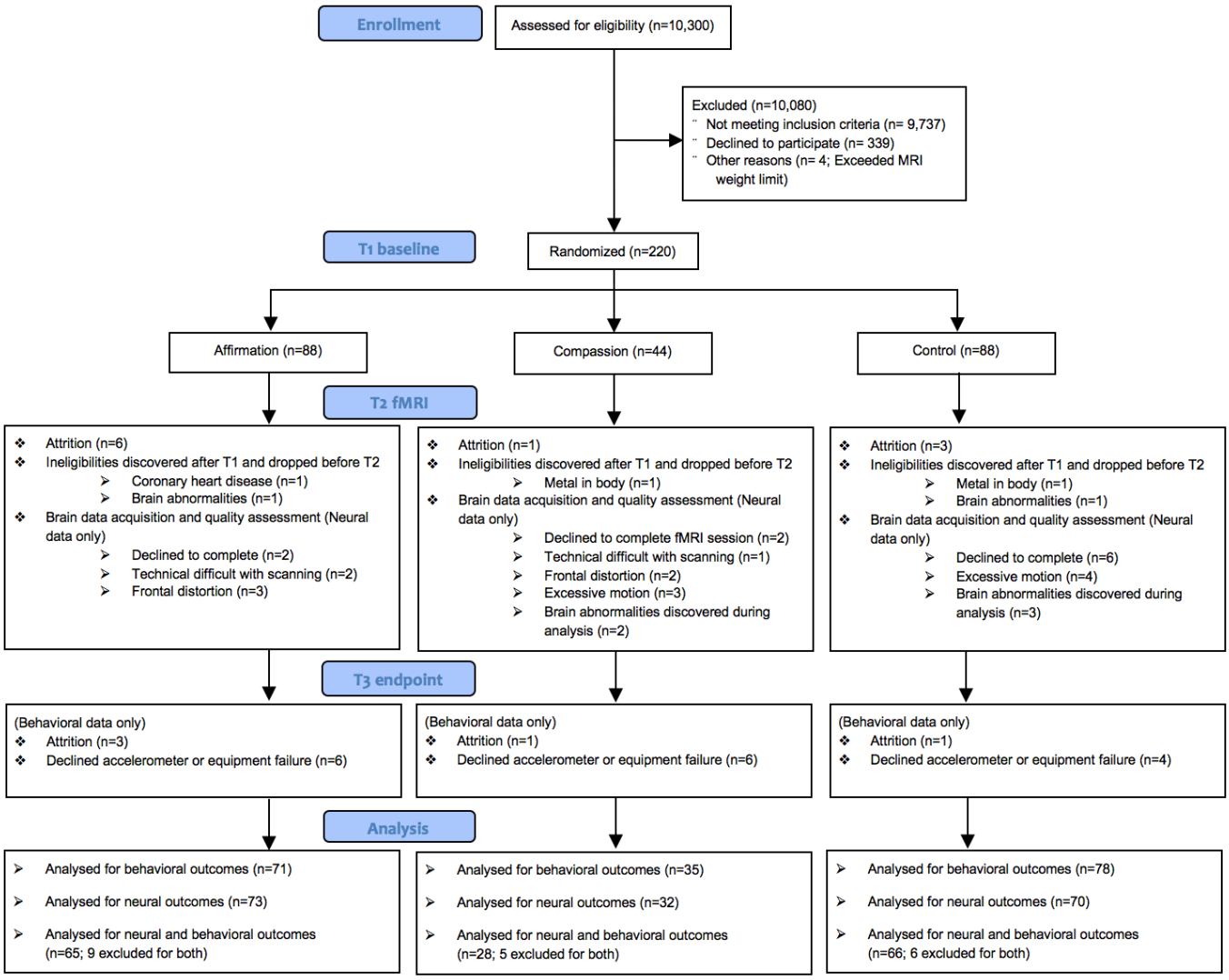

Fig. S1. Flow diagram of the study progress.

Accelerometer data calibration and preprocessing. Participants were directed to wear the waterproof GENEA accelerometer (2) at all times. As part of the baseline accelerometer calibration, participants performed sedentary (i.e., completing surveys while seated at a

computer terminal for at least 30 min) and moderate/vigorous activities (walking/climbing up and down the stairs for 6 min). For each participant, the third quartile (75^{th} percentile) of the activity during the sedentary period was used as a sedentary threshold, such that activity below that threshold was tagged as "sedentary." Activity greater than the second quartile (50^{th} percentile) of the peak level during the moderate/vigorous period was tagged as "moderate/vigorous."

Using the epoch converter function in the GENEActiv software, the raw triaxial data recorded at 20Hz were downsampled to 1 minute epochs, and the Sum Vector Magnitude (an integration of x, y and z acceleration) was used to provide an activity intensity score. Periods in which participants were sleeping or were not wearing the accelerometers were tagged by three research assistants blind to study hypotheses and condition assignments and excluded from the analysis. During the remaining periods in which participants were awake and wearing the device, days with less than 5 hours of wear were excluded (6,242 days of 7,092 tagged, or 88%, met this criterion).

Average daily proportions of activity during the T1-T2 baseline and T2-T3 post-intervention periods were computed by dividing the durations of moderate/vigorous and sedentary times, separately, by the total usable time that excludes sleep and non-wear time for each day (tagged by three blind coders), and averaging these daily scores across the ten-day baseline and one-month post-intervention periods. Baseline to post-intervention change scores were computed by subtracting average baseline intervention proportion scores from post-intervention proportion scores.

fMRI task descriptions.

***Affirmation (Control) task*.** At T1, all participants were presented with six value types including three self-transcendent (compassion and kindness, family and friends, spirituality) and three self-enhancing (wealth, power, fame) values, and ranked them in the order of importance. At T2, participants in the affirmation and control conditions were guided through either an affirmation or control task in the fMRI scanner to reflect on their highest (affirmation condition) or lowest (control condition) ranked values determined at T1. Those in the affirmation condition were provided with an opportunity to think about highest value situations in the future as vividly as they could (e.g., if family and friends were their highest value: "Have fun with family and friends").

As a within-subjects comparison to facilitate fMRI analyses (i.e., to control for low level stimulus features and psychological processes that were not the main focus of the task such as future oriented thinking), they were also asked to imagine value-neutral everyday activities with the same base instruction to think about the situations in the future as vividly as they could (e.g., "Check the weather to see the forecast"). Participants in the between subjects control condition were presented with a series of situations pertaining to their lowest ranked value as well as the within-subjects everyday activities (identical to the ones presented in the affirmation condition). This control employs the same format and value elicitation procedure as the main affirmation task, with the only difference being that participants in the affirmation condition reflect on their highest rated value and participants in the control condition reflect on their lowest rated value.

Forty situations (20 values, 20 everyday activities) were presented across two scanning runs (20 in each run) in a randomized order. Each block consisted of an initial screen showing the block type (value/everyday activity; 2s), followed by the situation description (10s) and an importance rating (4s). Blocks were separated by fixation rest periods (3s); every fifth block contained a longer (10s) period of rest.

Compassion task. To explicitly elicit a self-transcendent and other-directed mindset without necessarily having to use specific, personalized values for different participants, we drew on methodologies and theories from compassion literature. During compassion practice,

individuals make well-wishes for a growing circle of others, ranging from close family members and friends to acquaintances, and eventually everyone in the world.

Most neuroimaging studies of compassion have used resting state as within-subjects contrast trials, which might not be ideal, because regions active during the compassion task can also be active during rest, thereby washing out the compassion effect in the brain. To address this issue, we developed a scanner-adapted guided compassion task that presents other-directed wishing trials in short intervals compatible with a standard block design, and with a structure that parallels our affirmation and control tasks.

We designed a scanner-adapted compassion task that uses directed well-wishing techniques, in which we asked participants to make positive wishes for three target groups, including close others, acquaintances, and everyone in the world, and as vividly as they could, imagine situations in which these wishes come true in the future. Forty wish blocks (20 well-wishes, 20 control wishes) were presented across two runs (20 wishes in each run) in a randomized order. Each wish block consisted of an initial wish phrase (2s; “May you be well”), followed by the target group to direct positive wishes to (10s; “Everyone in the world”) and importance rating (4s). Control wishes focused on everyday activities to allow comparisons of neural activity during transcendent vs. non-transcendent processing, using the same everyday activities that formed the within-subjects control trials for the affirmation and control groups (e.g., “May it be done easily: Check the weather to see the forecast”). Blocks were separated by fixation rest periods (3s); every fifth block contained a longer (10s) period of rest.

Health messages intervention. Immediately following the priming task, all participants received the same 30 health messages targeting sedentary, high-BMI adults (matching the participant demographics). Messages were presented across two runs (15 health messages in each run) in a randomized order, and each message text was accompanied by a simple pictogram and audio of the text to control for reading speed. Each block consisted of a health message (8s) that highlighted a reason why participants should be less sedentary or more active (“You can live longer to enjoy the things you love if you start to sit less.”), how participants might implement the suggestions (“Make a habit of walking up and down the stairs whenever you can.”), or increased risk for chronic disease due to sedentary lifestyle and elevated BMI (“You are more likely to die early if you stay sedentary.”), followed by a relevance rating (4s). All message types were controlled for reading levels using the Simple Measure of Gobbledygook (SMOG) grade and argument strength. Blocks were separated by fixation rest periods (3s); every fifth block contained a longer (12s) block of rest. The task also included blocks with advice regarding other daily behaviors unrelated to physical activity that are not the focus of the current report (n=30).

Self localizer task. A standard self localizer task (43) was used to create group-level functional ROIs (fROIs) associated with self-related processing. Functionally defined ROIs in the VMPFC and PCC were identified. Participants saw 32 personality traits, selected from a list of 120 based on ratings made at the T1 appointment. Participants made binary judgments about a series of personality traits on their self-relevance (me/ not me) and, as a within-subject control trial, valence (good/bad). The task also included trials in which participants judged the case of the lettering (upper/lower) that are not the focus of the current report. Thirty-two personality traits were each presented once for each type of judgment, for a total of 96 trials across one run. Trial types were blocked, such that participants always saw four trials of the same type consecutively. This resulted in eight blocks of each judgment type. Each block consisted of an initial screen showing the block type (2s), followed by four consecutive personality trait words and judgment ratings (3.2s each). Blocks were separated by fixation rest periods (4s, range 2-12s).

Analysis plan. A series of models were computed to test the neural basis of self-transcendence during the affirmation and compassion tasks, the hypothesized effects of affirmation and compassion on neural activity during health message exposure, and subsequent changes in physical activity. We also tested for interaction effects between condition and brain activity within key ROIs on behavior change. All analyses controlled for demographic variables (centered age, sex, ethnicity, centered years of education), and models predicting changes in sedentary and moderate/vigorous activities controlled for baseline activity levels. Planned comparison orthogonal contrasts were used to compare the effects of the three conditions. All reported p values are two-tailed. All analyses were performed in R (v3.0.1, www.r-project.org) using the R-studio interface (v0.98.1103).

fMRI data preprocessing and whole-brain analysis. The imaging data were acquired on 3 Tesla Siemens Trio scanners equipped with a 32 or 64 channel head coil. The head coil type was not associated with any of the neural outcomes (*p*s>.10). High-resolution T1-weighted structural images were collected using an MPRAGE sequence (TI=1,100ms, 160 slices, slice thickness=1mm, voxel size=0.9 × 0.9 × 1). T2*-weighted functional images were recorded (repetition time=1,500ms, echo time=25ms, flip angle=70°, -30° tilt relative to AC-PC line, 54 slices, field of view=200mm, slice thickness=3mm, multiband acceleration factor=2, voxel size=3.0 × 3.0 × 3.0 mm).

Participants were self-guided through two runs of the priming task (affirmation, compassion, or control) (294 volumes each; 588 volumes total) immediately followed by two runs of the health messages task (run1=376, run2=344 volumes, 720 volumes total), and one run of the self localizer task (308 volumes).

The anatomical and functional data were acquired and preprocessed using a standard processing stream using Statistical Parametric Mapping (SPM8; Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK) for all stages apart from the initial despiking, which was carried out using the 3dDespike program as implemented in the AFNI toolbox. Differences in time of acquisition were corrected using a sinc interpolation algorithm with the first slice as reference. Next, data were spatially realigned to the first slice of each volume, and co-registered to functional and structural images using two six-parameter affine stages. The mean image across all blood oxygen level-dependent (BOLD) functional images was registered to high-resolution T1 images (total of 12 parameter affine). Following co-registration, the high-resolution T1 images were segmented into gray matter, white matter and cerebrospinal fluid to create a brain mask used to determine voxels to be included in first and second-level models. Structural and functional images were then normalized to the skull-stripped MNI template ("MNI152_T1_1mm_brain.nii") provided by the FMRIB Software Library (FSL). In the final preprocessing step, the functional images were smoothed using a Gaussian kernel (8-mm FWHM). To allow for the stabilization of the BOLD signal, the first five volumes (7.5s) of each run were discarded before analysis. Movement parameters (a total of six rigid-body parameters, three for translation and three for rotation) derived from spatial realignment were included as nuisance regressors in all first-level models. Data were high pass filtered with a cutoff of 128s.

For ROI analyses, for each person, we compared parameter estimates of activity during the 20 highest (affirmation task) and lowest value (control task) trials, and 20 well-wishes (compassion task) trials with 20 everyday activity trials, within subjects, using MarsBaR (3), and converted to percent signal change.

For the primary whole-brain results, 3dClustSim was used to calculate the cluster threshold (http://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html). The estimated smoothness from each analysis was calculated using SPM (17.1, 16.6, 16.2 for the priming tasks, and 16.5, 16.6, 15.8 for the health messages task). Based on the results from

3dClustSim, *k*=243 and *k*=206 were applied for the whole-brain analysis of the priming tasks, and health messages task, respectively, at *p*<.005, corresponding to *p*<.05, corrected.

Self localizer task analysis. A self localizer task was used to create group-level functional ROIs (fROIs) associated with self-related processing (4, 5). A fixed-effects model of the self localizer task was constructed using a single boxcar function for each block with three block types (self-relevance, valence, case). A contrast between blocks in which people made self-relevance vs. valence judgment was used. Second-level random-effects models were constructed by averaging across participants from which peak voxels in the VMPFC and PCC regions were extracted to create fROIs. Data from 9 people included in the main manuscript analyses were not used in defining self-related fROIs due to technical difficulties (*n*=6), participants' wish to discontinue (*n*=2), and motion (*n*=1) that were specific to the localizer task. The self-relevance blocks were contrasted with the valence judgment blocks for each participant, which produced large group-level clusters along the midline including the VMPFC and PCC. Next, group-level peak voxels were identified in the MPFC and PCC masks drawn from SPM's WFU PickAtlas tool. The MPFC mask included the bilateral frontal superior medial, rectus, anterior cingulum, frontal medial orbital gyri defined by the AAL atlas, and was truncated at Z=32 to approximate the extent of BA10. The PCC mask included the bilateral posterior cingulum and the precuneus defined by the AAL atlas. We extracted the peak coordinates of the group-level self-relevance > valence contrast from the MPFC and PCC masks, and drew 9mm spherical ROIs around these peaks (VMPFC peak coordinates: -3, 50, -8; PCC peak coordinates: -6, -55, 28).

SI Results

Neural regions associated with self-transcendence priming. Contrasts were computed focusing on activation for each condition, and then comparing the conditions. N's below refer to the number of participants with usable data in each condition.

Region	*x*	*y*	*z*	*size*	*t*
Affirmation (*n*=73)					
Highest value > Everyday activity trials					
precuneus	-6	-55	34	457	8.49
L middle temporal gyrus	-63	-10	-17	835	6.65
ventromedial prefrontal cortex	-6	59	19	2217	6.63
R middle temporal gyrus	51	11	-32	452	5.46
L inferior frontal gyrus	-45	2	28	266	-6.07
L middle frontal gyrus	-48	47	10	1268	-6.54
L middle frontal gyrus	-24	8	52	441	-7.29
R middle frontal gyrus	30	11	52	3169	-7.86
R middle temporal gyrus	60	-49	-8	775	-8.58
L inferior temporal gyrus	-57	-58	-8	1360	-8.86
L superior parietal lobule	-33	-79	34	5479	-9.10
L midcingulate gyrus	0	-34	40	-	-8.92
R angular gyrus	39	-64	49	-	-8.84

Compassion (*n*=32)
Well-wishes > Everyday activity trials

ventromedial prefrontal cortex	-6	56	16	2400	6.06
/ventral striatum	0	8	-11	-	3.13
L temporal parietal junction	-45	-70	22	315	4.66
R temporal parietal junction	54	-61	28	297	4.34
R inferior parietal lobule	39	-52	49	305	-4.15
R middle frontal gyrus	51	26	43	789	-5.32
L middle frontal gyrus	-42	5	31	1600	-5.58
L superior parietal lobule	-30	-61	46	1079	-5.91
occipital poles	-18	-100	-5	2374	-11.10

Control (*n*=70)
Lowest value > Everyday activity trials

L temporal pole	-48	14	-32	1439	7.26
R temporal pole	48	20	-32	736	7.21
ventromedial prefrontal cortex	3	41	-23	1757	6.41
/ventral striatum	3	-1	-8	-	2.98
R occipital pole	36	-91	13	699	5.34
precuneus	-6	-52	28	269	5.31
L occipital pole	-33	-97	13	1156	5.15
L insula	-39	-13	-8	363	-3.57
R cerebellum	36	-61	-44	1099	-4.77
L middle frontal gyrus	-27	38	-14	250	-7.10
cingulate gyrus	-6	-34	43	6580	-8.05
L precuneus	-15	-58	13	-	-7.73
L fusiform gyrus	-33	-40	-14	-	-6.36

Affirmation > Control (*n*=143)

R lingual gyrus	15	-70	-8	390	5.11
ventromedial prefrontal cortex/	-3	47	-2	570	4.39
ventral striatum	-3	14	-5	-	2.81
L supramarginal gyrus	-36	-52	37	365	-4.48
R middle frontal gyrus	30	8	49	1640	-5.30
R inferior parietal lobule	48	-58	49	1340	-5.89

Compassion > Control (*n*=102)

middle temporal gyrus	-45	-76	22	291	4.64
ventromedial prefrontal cortex/	-12	53	1	1368	4.46
ventral striatum	-6	20	-14	-	3.77
L inferior frontal gyrus	-57	26	19	944	-4.41
R middle frontal gyrus	54	35	28	487	-5.10
occipital poles	-18	-100	-2	2608	-11.14

Compassion > ***Affirmation*** (*n*=105)

L middle temporal gyrus	-39	-82	34	316	5.87
R superior frontal gyrus	27	32	37	425	5.17
R temporal parietal junction	39	-82	40	649	4.93

Affirmation > Compassion (*n*=105)					
occipital poles	-18	-100	-2	2106	10.29
Self-Transcendence (Affirmation + Compassion) > Control (*n*=175)*					
lingual gyrus	15	-70	-5	160	5.28
ventromedial prefrontal cortex/	-6	50	-2	627	4.69
ventral striatum	-6	20	-11	-	3.57
L middle frontal gyrus	-24	29	40	25	3.94
R midcingulate gyrus	3	-13	37	39	3.72
L insula	-42	-13	22	20	3.69
L midcingulate gyrus	-12	-31	34	24	3.69
L middle temporal gyrus	-42	-70	16	21	3.66
R temporoparietal junction	57	-37	25	85	3.64
R insula	42	5	4	25	3.49
L insula	-39	5	-2	5	3.45
R middle temporal gyrus	60	-67	7	17	3.45
L temporoparietal junction	-66	-37	28	7	3.37
L postcentral gyrus	-24	-43	64	8	3.27

Table S2. Whole-brain results showing regions with increased activity during the priming tasks, $p<.005$, $k=243$, corresponding to $p<.05$, corrected, *$p<.001$, uncorrected. Note: L = left; R = right; Peak voxels and local maxima are reported in clusters that extend across the ventromedial prefrontal cortex and ventral striatum.

Whole-brain analysis results pooling the self-transcendence tasks. Whole-brain analysis results pooling the self-transcendence tasks (affirmation + compassion > control) produced neural activation in the VMPFC, VS, and bilateral TPJ ($p<.001$, uncorrected; Fig. S2).

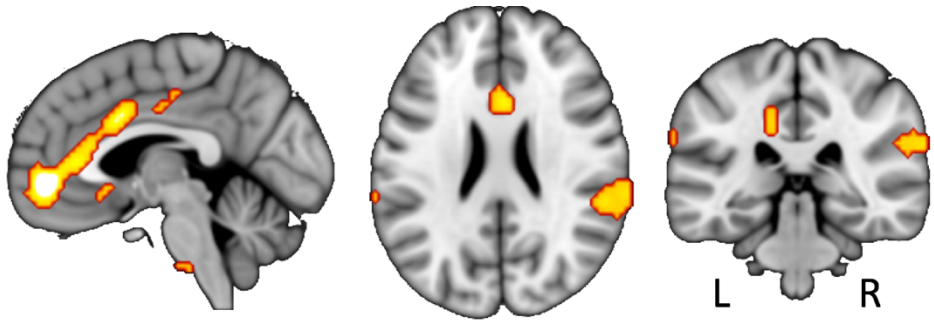

Fig. S2. Within-subjects contrasts of self-transcendence (affirmation + compassion) > control conditions ($p<.001$, uncorrected).

Neural responses predicting later increases in moderate to vigorous physical activity.

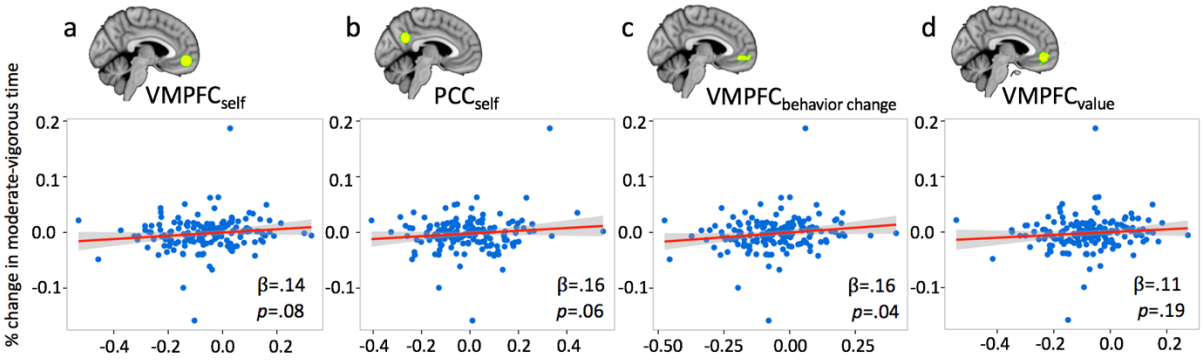

Fig. S3. Changes in moderate to vigorous behavior from baseline to endpoint visit predicted by activity during health message exposure in the primary ROIs.

Positive valuation ROIs during messages task and behavior change. Although not in our pre-registration document, in the years since we wrote the pre-registration, we have gained significant new insights into the processes relevant to behavior change, and have come to focus heavily on activity in brain regions implicated in positive valuation/reward as indicative of receptivity to persuasive information (1, 6–8). As such, we also examined the effect of affirmation and compassion priming on neural activity during health messages task within ROIs implicated in positive valuation/reward processing (VMPFC, VS) (9).

Effects of task on brain activity during health messaging. Those who completed the affirmation task prior to health messages, compared to controls, showed greater activity in $VMPFC_{value}$ (β=.23, $t(137)=2.68$, $p=.008$) and VS_{value} (β=.20, $t(137)=2.37$, $p=.02$) during health message exposure. Participants in the compassion condition also showed greater activity in $VMPFC_{value}$ compared to control participants (β=.21, $t(96)=2.07$, $p=.04$) during health message exposure, but did not differ from controls in their VS_{value} activity (β=.06, $t(96)=0.59$, $p=.55$). Participants in the affirmation and compassion conditions did not differ in their activity in the $VMPFC_{value}$ (β=.05, $t(99)=0.50$, $p=.62$) or VS_{value} ROIs (β=.13, $t(99)=1.30$, $p=.20$) during health message exposure.

Effects of brain activity on behavior change. Next, we tested whether activity in the value ROIs during the messages task was predictive of later behavior change. Activity in the $VMPFC_{value}$ during the health messages task predicted subsequent decreases in sedentary behavior (β=-.18, $t(152)=-2.24$, $p=.03$) but not changes in moderate/vigorous activity (β=.11, $t(152)=1.33$, $p=.19$). Activity in the VS_{value} during health messages task was marginally associated with greater decreases in sedentary behavior (β=-.15, $t(152)=-1.86$, $p=.07$), but not with changes in moderate/vigorous activity (β=.06, $t(152)=0.75$, $p=.46$). These results suggest that positive valuation, in addition to self-related processing during message exposure (i.e., subjective valuation), plays an important role in perceptions of subjective value and subsequent behavior change.

Shared self and valuation neural regions during messages predicting behavior change. To specifically examine the role of subregions of VMPFC previously implicated in both the self and value tasks in predicting behavior change, we extracted an ROI of an overlap between $VMPFC_{self}$ (from self localizer) and $VMPFC_{value}$ (9) ROIs, and subtracted this self/value ROI from regions associated with self and value processing to create five ROIs (Fig. S4): 1) self/value overlap VMPFC, 2) self-only VMPFC ($VMPFC_{self}$ minus $VMPFC_{value}$), 3) value-only VMPFC ($VMPFC_{value}$ minus $VMPFC_{self}$), 4) self-only network ($VMPFC_{self}$ and PCC_{self} minus $VMPFC_{value}$),

and 5) value-only network ($VMPFC_{value}$ and VS_{value} minus $VMPFC_{self}$). We tested whether activity within the self/value overlap ROI during message exposure predicted behavior change controlling for activities in non-overlapping self-only or value-only ROIs. We focused on sedentary minutes, where we found greater effect using the pre-registered $VMPFC_{self}$ ROI.

Activity in the shared self/value region ($VMPFC_{self/value}$) was robustly associated with decreased sedentary behavior over the month following the intervention (β=-.18 $t(152)$=-2.23, p=.03). Separately, decreases in sedentary minutes were associated with greater activity in non-overlapping self-only VMPFC (β=-.15, $t(152)$=-1.94, p=.054), value-only VMPFC (β=-.17, $t(152)$=-2.11, p=.04), and value-only network (β=-.17, $t(152)$=-2.17, p=.03) ROIs, but not in the self-only network ROI (β=-.10 $t(152)$=-1.21, p=.23); though the self-only network was associated with increases in moderate to vigorous activity (β=.17 $t(152)$=2.12 , p=.04). However, these became non-significant when controlling for the activity in the overlapping self/value ROI (ps>.10), emphasizing the difficulty in dissociating these processes.

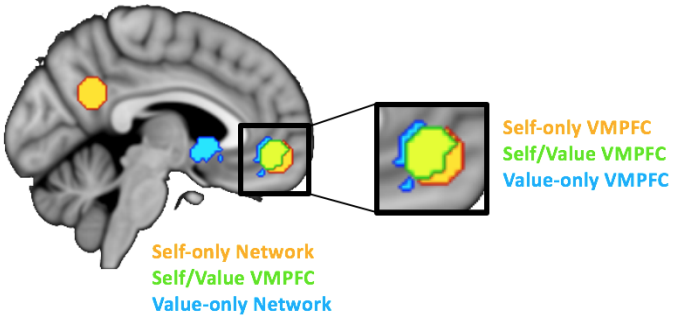

Fig. S4. Shared and distinctive neural regions of interest associated with self and value processing.

Path analyses. We conducted two path analyses that each specified three direct effects, including condition to neural activity during the priming tasks (affirmation, compassion, control tasks) to neural activity during the messages task to behavior change (Fig. S5). The $VMPFC_{value}$ ROI was meta-analytically defined (9). The ROIs for the messages task included $VMPFC_{self}$ (from the self localizer task) and $VMPFC_{self/value}$ (an overlapping cluster between $VMPFC_{self}$ and $VMPFC_{value}$). The hypothesized structural model had acceptable model fit. Self-transcendence priming increased activity within regions associated with positive value or reward processing during priming tasks. This, in turn, led to increased activity within regions associated with self-relevance processing (marginal) and self/value processing during the messages task, which then predicted decreases in sedentary minutes.

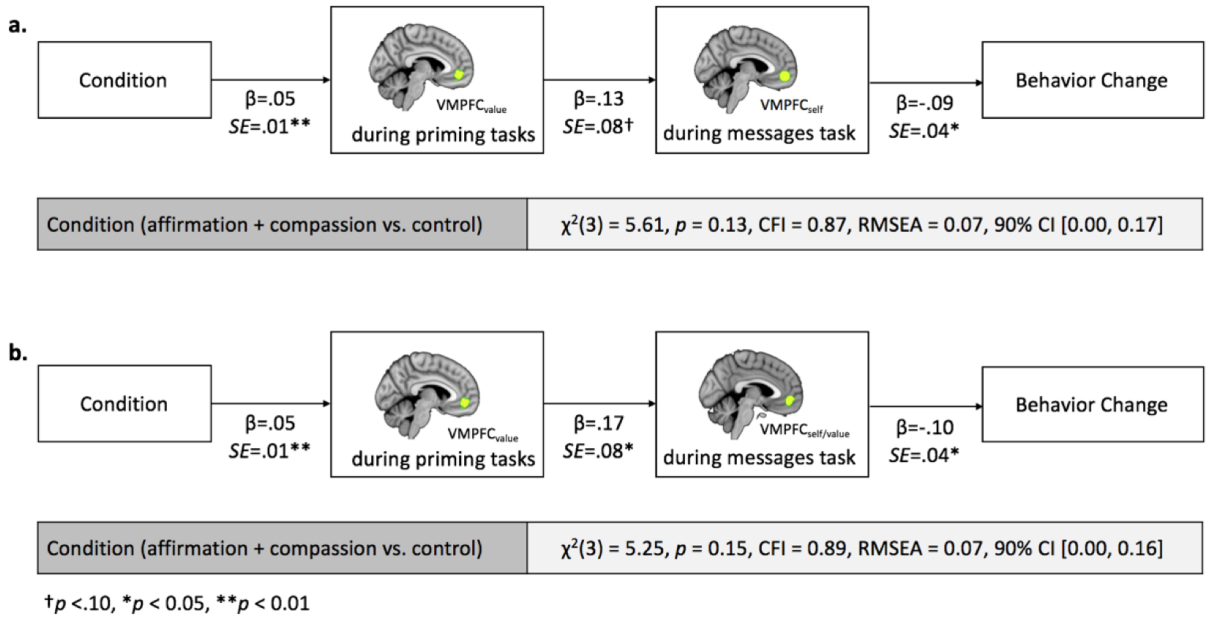

Fig. S5. Self-transcendence effects path models, using activity during the messages task within **a.** $VMPFC_{self}$ and **b.** $VMPFC_{self/value}$. Maximum likelihood estimates (unstandardized coefficients) are reported. Behavior change = changes in average daily proportions of sedentary minutes at post-intervention from baseline; CFI = comparative fit index; RMSEA = residual mean square error of approximation; VMPFC = ventromedial prefrontal cortex.
*$p < 0.05$

Neural regions associated with behavior change. Additional whole-brain analyses of the health messages task identified areas associated with later increases in moderate/vigorous activity and decreases in sedentary behavior (Fig. S6). This analysis revealed a single cluster in the ventral tegmental area (VTA), a central reward circuitry in the brain that projects to VS (10), that extends to the left precentral gyrus, insula, and VS, and was associated with increases in later moderate/vigorous activity (p<.005, k=206, corresponding to p<.05, corrected, Fig. S6a). For the regions associated with decreases in sedentary behavior, no clusters survived the correction (k=206), and Fig. S6b shows uncorrected clusters at p<.005, uncorrected. Increased activity in the VMPFC, VS, and VTA during the messages task was associated with later decreases in sedentary behavior, uncorrected.

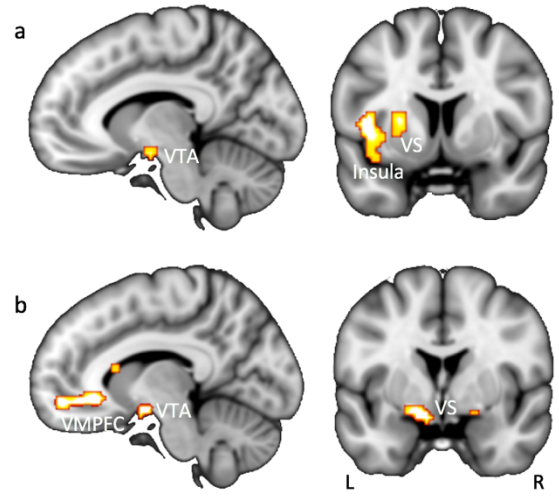

Fig. S6. Whole-brain results showing active regions during the health message exposure that later predicted **a.** increases in moderate/vigorous activity (p<.005, k=206, corresponding to p<.05, corrected), and **b.** decreases in sedentary behavior (p<.005, uncorrected).

Condition-general effects of VMPFC on behavior change. In our pre-registration, we proposed a condition * VMPFC interaction in predicting behavior change because we expected that the intervention might be more effective for some people than others, which in turn might be reflected in VMPFC activity, particularly within the experimental conditions. Contrary to our prediction, we did not observe any interactions between condition and activity within our ROIs in predicting behavior change (ps>.10) (Fig. S7). Further, in control participants, we observed a positive relationship between activity within $VMPFC_{self}$ during the messages task and decreases in sedentary behavior (β=-0.33, t(59)=-2.61, p=.01), suggesting that to the extent that people in the control condition showed heightened activity in MPFC, we also observed downstream behavioral effects. We did not observe this relationship in PCC (p=.76).

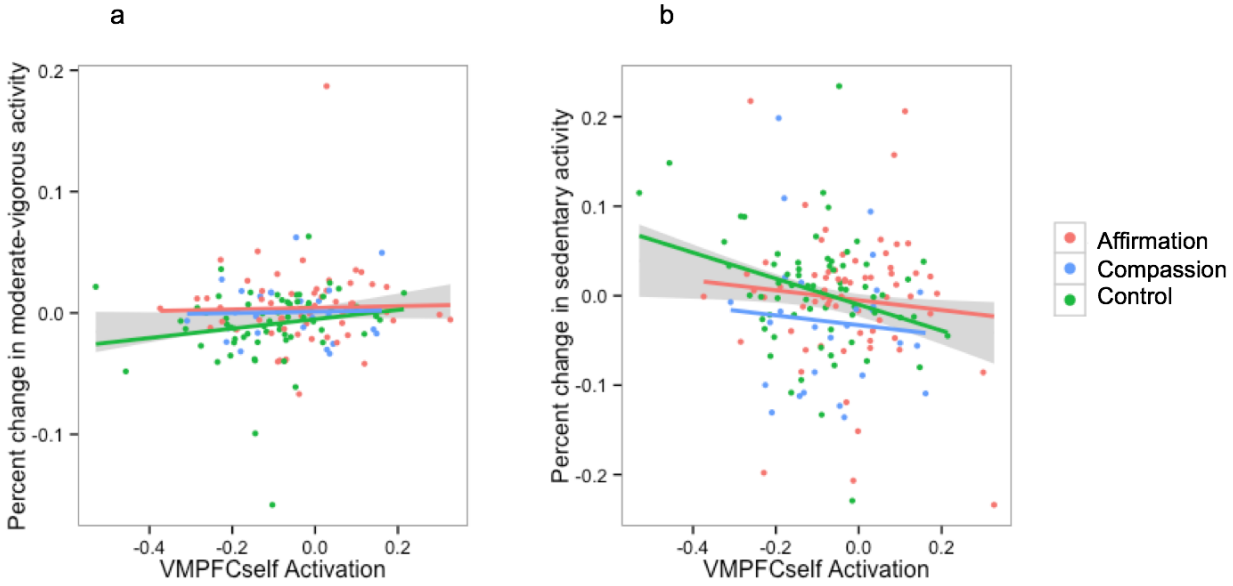

Fig. S7. Activity during health message exposure in $VMPFC_{self}$ predicting changes in **a.** moderate to vigorous and **b.** sedentary behavior from baseline to endpoint visit. There was no condition * brain activity interaction in predicting behavior change (*p*s>.10).

TPJ activity during the priming tasks and subsequent message receptivity. A whole-brain analysis comparing the self-transcendence manipulations vs. the control task revealed activity in bilateral TPJ (*p*<.001, uncorrected), which are key regions implicated in mentalizing/other-focus (11, 12). Given that one possible pathway through which self-transcendence might exert its effects is through other-directed focus in the priming task, we explored whether participants who showed greater activations in TPJ during the priming tasks also showed greater receptivity to the messages. A bilateral TPJ ROI was extracted from the self-transcendence vs. control contrast during the priming tasks (*p*<.001, uncorrected). Greater activations in TPJ during the self-transcendence tasks were associated with greater activations of $VMPFC_{self/value}$ during the messages task (β=.16, *t*(169)=2.17, *p*=.03), as well as greater decreases in shame following message exposure (β=-.16, *t*(169)=-2.17, *p*=.03). Further, a path model indicated that activity in bilateral TPJ during the priming tasks is significantly associated with changes in activity within the $VMPFC_{self/value}$ ROI (an overlapping cluster between $VMPFC_{self}$ and $VMPFC_{value}$) during message receptivity (Fig. S8).

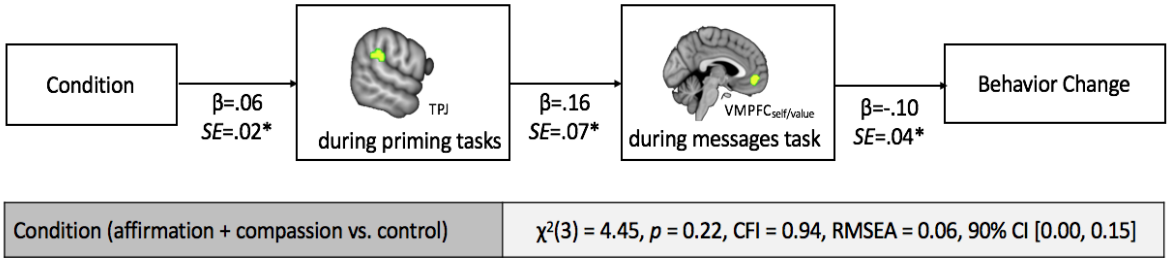

Fig. S8. A self-transcendence effects path model for TPJ activity during priming tasks and $VMPFC_{self/value}$ activity during messages task. Maximum likelihood estimates (unstandardized coefficients) are reported. Behavior change = changes in average daily proportions of sedentary minutes at post-intervention from baseline; CFI = comparative fit index; RMSEA = residual mean square error of approximation; TPJ = bilateral temporoparietal junction; VMPFC = ventromedial prefrontal cortex.
*$p < 0.05$

Self-report measures. Changes in participants' self-reported measures of attitudes, intentions and self-efficacy were not associated with changes in sedentary or moderate/vigorous activities following the intervention (*p*s>.20). Primary analyses results linking neural activities to changes in physical activity remained parallel controlling for self-reported attitudes, intentions, and self-efficacy with an exception of the effect of $VMPFC_{self}$ on decreases in sedentary behavior which becomes marginal (β=-.15, *t*(147)= -1.78, *p*=.08) when controlling for all self reports.

Mood was assessed using the modified Differential Emotion Scale (mDES) (13), and change scores were computed by subtracting the T1 baseline scores from the scores immediately following the T2 fMRI scan. Participants in the self-transcendence conditions (affirmation and compassion), compared to controls, reported less negative mood (β=-.18, *t*(178)=-2.41, *p*=.02). In particular, participants in self-transcendence conditions reported less shame (β=-.16, *t*(178)=-2.10, *p*=.04), previously associated with lower receptivity to health messages (14, 15), following exposure to the health messages. These results suggest that self-

transcendent manipulations lead to decreased negative self-focus following exposure to potentially threatening health messages.

Counterarguing ROI analyses and results. Defensive processing including generation of counterarguments hinders behavior change (16–18). We tested whether affirmation and compassion priming would decrease counterarguing processes during subsequent message exposure, as indexed by a counterarguing localizer task as part of the current study. A counterarguing localizer was used to create a group-level fROI associated with counterarguing processes. Participants were presented with 50 non-threatening statements (e.g. "People should do the crossword") and were asked to generate arguments against them (against blocks) or in favor of them (in favor blocks), or as another control condition, passively judge whether statements were true or false (control blocks). In against and in favor blocks, participants were instructed to press the button box for each reason they generated. An against or in favor block consisted of an initial screen showing the block type (3s), followed by a statement and judgment ratings (12s). A control block consisted of a screen indicating the upcoming condition (3s), followed by three consecutive statements and judgment ratings (4s each). Blocks were separated by fixation rest periods (2s); every fourth block contained a longer (12s) block of rest.

Effects of condition on counterarguing processing during health messaging. To test whether affirmation and compassion priming may decrease counterarguing processing during subsequent message exposure, we drew $DLPFC_{counterarguing}$ fROIs using results from the counterarguing localizer task. Data from 11 people included in the main manuscript analyses were not used in defining the fROI due to technical difficulties (n=2), no variations in button response across block types (n=4), participants' wish to discontinue (n=4), and motion (n=1) that were specific to the localizer task. To identify processes associated with arguing against messages (i.e., counterarguing processes), we compared activity during the against blocks with the in favor blocks for each participant, which produced a cluster in the left DLPFC. Based on our prior theory, we restricted our search to the DLPFC as defined by the WFU PickAtlas tool in the left superior frontal gyrus, middle frontal gyrus, and inferior frontal gyrus (triangular part) defined by the AAL atlas. We extracted the peak coordinates of the group-level against > in favor contrast in the DLPFC, and drew a 9mm spherical ROI around this peak (left DLPFC coordinates: -51, 35, 22). Contrary to our predictions, those in the affirmation condition, compared to controls, showed greater activity in the left $DLPFC_{counterarguing}$ (β=.17, $t(137)$=1.99, p=.05). Activity in the left $DLPFC_{counterarguing}$ did not differ between participants in the compassion and control conditions (β=.13, $t(96)$=1.24, p=.22) or between affirmation and compassion conditions (β=.081, $t(99)$=0.80, p=.42). Furthermore, contrary to our prediction, the left $DLPFC_{counterarguing}$ activity during message exposure was not associated with subsequent changes in moderate/vigorous activity (β=.10, $t(152)$=1.32, p=.19) or sedentary behavior (β=-.04, $t(152)$=-0.49, p=.62). These data suggest that the pathways through which affirmation and compassion priming exert their effects on receptivity to health messaging may not occur through decreases in cognitive counterarguing processing indexed by left DLPFC, but rather emphasize the receptivity pathway described in the main manuscript.

References

1. Falk EB, et al. (2015) Self-affirmation alters the brain's response to health messages and subsequent behavior change. *Proceedings of the National Academy of Sciences* 112(7):1977–1982.

2. Esliger DW, et al. (2011) Validation of the GENEA Accelerometer. *Med Sci Sports Exerc* 43(6):1085–1093.

3. Brett M, Anton J-L, Valabregue R, Poline J-B (2002) Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage* 16(2):S497.

4. Chua HF, et al. (2011) Self-related neural response to tailored smoking-cessation messages predicts quitting. *Nat Neurosci* 14(4):426–427.

5. Schmitz TW, Johnson SC (2006) Self-appraisal decisions evoke dissociated dorsal—ventral aMPFC networks. *Neuroimage* 30(3):1050–1058.

6. Cooper N, Tompson S, Brook O'Donnell M, Emily BF (2015) Brain Activity in Self- and Value-Related Regions in Response to Online Antismoking Messages Predicts Behavior Change. *Journal of Media Psychology* 27(3):93–109.

7. Falk EB, Berkman ET, Mann T, Harrison B, Lieberman MD (2010) Predicting persuasion-induced behavior change from the brain. *J Neurosci* 30(25):8421–8424.

8. Falk EB, Berkman ET, Whalen D, Lieberman MD (2011) Neural activity during health messaging predicts reductions in smoking above and beyond self-report. *Health Psychol* 30(2):177–185.

9. Bartra O, McGuire JT, Kable JW (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76:412–427.

10. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482(7383):85–88.

11. Saxe R, Kanwisher N (2003) People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19(4):1835–1842.

12. Scholz J, Triantafyllou C, Whitfield-Gabrieli S, Brown EN, Saxe R (2009) Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS One* 4(3):e4869.

13. Cohn MA, Fredrickson BL, Brown SL, Mikels JA, Conway AM (2009) Happiness unpacked: positive emotions increase life satisfaction by building resilience. *Emotion* 9(3):361–368.

14. Tangney JP, Dearing RL (2003) *Shame and Guilt* (Guilford Press).

15. Vartanian LR, Shaprow JG (2008) Effects of weight stigma on exercise motivation and behavior: a preliminary investigation among college-aged females. *J Health Psychol* 13(1):131–138.

16. Blumberg SJ (2000) Guarding against threatening HIV prevention messages: an information-processing model. *Health Educ Behav* 27(6):780–795.

17. Zuwerink Jacks J, Cameron KA (2003) Strategies for Resisting Persuasion. *Basic Appl Soc Psych* 25(2):145–161.

18. Liberman A, Chaiken S (1992) Defensive Processing of Personally Relevant Health Messages. *Pers Soc Psychol Bull* 18(6):669–679.