Health Neuroscience Special Issue

Neural mechanisms of self-affirmation's stress buffering effects

Janine M. Dutcher, [1] Naomi I. Eisenberger,[2] Hayoung Woo,[3] William M. P. Klein,[4] Peter R. Harris, [5] John M. Levine,[6] and John David Creswell[1]

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA, [2]Department of Psychology, University of California, Los Angeles, CA 90095, USA, [3]Department of Psychology, New York University, New York, NY 10003, USA, [4]Behavioral Research Program, National Cancer Institute, Rockville, MD 20852, USA, [5]School of Psychology, University of Sussex, Brighton BN1 9RH, UK and [6]Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260, USA

Correspondence should be addressed to Janine M. Dutcher, 5000 Forbes Ave, Pittsburgh, PA 15213, USA. E-mail: jdutcher@andrew.cmu.edu.

Abstract

Self-affirmation can buffer stress responses across different contexts, yet the neural mechanisms for these effects are unknown. Self-affirmation has been shown to increase activity in reward-related neural regions, including the ventral striatum and ventromedial prefrontal cortex (VMPFC). Given that reward-related prefrontal cortical regions such as the VMPFC are involved in reducing neurobiological and behavioral responses to stress, we hypothesized that self-affirmation would activate VMPFC and also reduce neural responses to stress in key neural threat system regions such as the dorsal anterior cingulate cortex (dACC) and anterior insula (AI). We explored this hypothesis using self-affirmation and evaluative stress tasks following a within-subjects design in the fMRI scanner. Consistent with prior work, self-affirmation blocks led to lower self-reported stress and improved performance. With respect to neural activity, compared to control blocks, self-affirmation blocks led to greater VMPFC activity, and subsequently less left AI (but not dACC) activity during stress task blocks. Functional connectivity analyses revealed greater connectivity between the VMPFC and left and right AI during self-affirmation compared to control. These findings begin to articulate the neural circuits involved in self-affirmation's effects during exposure to stressors, and more broadly specify neural reward-based responses to stressful situations.

Key words: health neuroscience; stress; interventions; self-affirmation

Introduction

Self-affirmation—the process of reflecting on important personal values or attributes—has a host of benefits from reducing stress, to improving performance, to enhancing well-being (Cohen and Sherman, 2014). Across a number of experimental studies, self-affirmation has been shown to lead to reduced physiological responses to stress (Creswell *et al.*, 2005), lower levels of a marker of damage to blood vessels (endothelial cell-derived microparticles) following a stress induction (Spicer *et al.*, 2016) and improved performance, especially for those with

higher levels of chronic stress (Creswell *et al.*, 2013, p. 20). Outside of the lab, there is wide-ranging evidence that a self-affirmation intervention can lead to reduced physiological responses to real-world stressors (Sherman *et al.*, 2009) and foster enduring improvements in academic performance outcomes among vulnerable students (Cohen *et al.*, 2006, 2009; Sherman *et al.*, 2013; Brady *et al.*, 2016; Goyer *et al.*, 2017). Furthermore, self-affirmation has been shown to reduce defensive responses to threatening health information (for a review: Cohen and Sherman, 2014), and to facilitate health behavior change (for reviews: Epton *et al.*, 2015; Ferrer and Cohen, 2019). These findings converge to suggest that self-affirmation reduces stress and mitigates the consequences of stress. While there has been some research into the psychological mechanisms for self-affirmation effects [focus on others (Crocker *et al.*, 2008), attentional processes (Klein and Harris, 2009)], the neural mechanisms underlying the effects of self-affirmation on stress responses are not well understood.

It is not known whether self-affirmation can affect neural threat responses to a stressful experience, and how that effect occurs. One plausible mechanism for these stress resilience effects is via neural reward pathways. Previous studies have found that engaging in a self-affirmation task, compared to a control task, elicits greater neural activity in reward-related regions including the ventromedial prefrontal cortex (VMPFC) and ventral striatum (VS) (Cascio *et al.*, 2016; Dutcher *et al.*, 2016; Kang *et al.*, 2018). Moreover, across human and animal studies, reward system activation can lead to reductions in stress responding, including changes in behavior and physiology (for a review: Dutcher and Creswell, 2018b). Thus, it is plausible that self-affirmation's stress buffering benefits occur via increased reward-related neural activity, which, in turn, attenuates neural responses to stressful experiences.

There are a set of neural regions that play a central role in detecting a threat, appraising the threat and resources available for managing it, and deploying the physiological stress response. These regions include the dorsal anterior cingulate cortex (dACC), amygdala and anterior insula (AI) (Eisenberger and Cole, 2012; Muscatell and Eisenberger, 2012; Gianaros and Wager, 2015). Research has found that activity in these regions is associated with stress physiology, including autonomic, cardiovascular, neuroendocrine and immune responses, as well as psychological stress (Gianaros and Wager, 2015; Cohen *et al.*, 2016). However, it is also critical for an individual to be able to regulate and modulate this stress response. There are important structural connections between reward regions and these threat-related regions (Saper, 1982; Chiba *et al.*, 2001; Cloutman *et al.*, 2012; Dutcher and Creswell, 2018b). For example, neuroanatomical work suggests that connections between medial prefrontal regions and the insula might lead to integration of autonomic systems with behavior and affect (Saper, 1982). Human neuroimaging has also found that VMPFC and VS activity to rewarding stimuli leads to corresponding decreases in dACC and AI activity to pain stimulation (Younger *et al.*, 2010; Eisenberger *et al.*, 2011). Taken together, the extant literature suggests that the VMPFC and VS are key regions modulating dACC and AI responses to threatening or stressful stimuli. Psychophysiological interaction (PPI) analysis is a method for understanding the functional connectivity between regions during a task, and helps to determine if there is transmission of information between circuits in the brain. Thus, we additionally explored whether VMPFC, and VS, activity during self-affirmation leads to greater functional connectivity with dACC or AI, as a mechanism for self-affirmation's stress buffering effects.

The present study had three basic goals: (i) to replicate previous work demonstrating that self-affirmation leads to greater reward-related neural activity (Cascio *et al.*, 2016; Dutcher *et al.*, 2016; Kang *et al.*, 2018); (ii) to determine if self-affirmation leads to reduced neural responses to a stress manipulation and (iii) to explore VMPFC functional connectivity differences during self-affirmation compared to control.

Importantly, to understand how self-affirmation affects neural and behavioral responses to stress in an fMRI paradigm, both self-affirmation and stress needed to be manipulated within-subjects. While self-affirmation research typically manipulates self-affirmation between subjects, the fMRI environment necessitates within-subject comparisons, so a secondary goal of this study was to explore whether self-affirmation has stress buffering effects in a within-subjects design. To explore self-affirmation's effects on neural and behavioral responses to stress, we manipulated self-affirmation using a scanner-adapted version of a self-affirmation task, which includes both self-affirmation and non-affirmation control conditions (Dutcher *et al.*, 2016). To manipulate stress levels, participants completed a stressful math task with a socially evaluative component (Dedovic *et al.*, 2005; Inagaki *et al.*, 2016), which allowed us to assess, within subjects, neural and self-reported stress, as well as performance under stress. This task has been shown to lead to increased activity in the dACC and bilateral AI (Wang *et al.*, 2005; Dedovic *et al.*, 2009; Inagaki *et al.*, 2016).

Based on previous behavioral work (Creswell *et al.*, 2005, 2013), we hypothesized that self-affirmation would lead to lower self-reported stress and better performance compared to non-affirmation control. Based on previous neuroimaging work (Cascio *et al.*, 2016; Dutcher *et al.*, 2016; Kang *et al.*, 2018), we predicted that self-affirmation would lead to greater reward-related neural activity compared to non-affirmation control. Furthermore, we predicted there would be less threat-related neural activity during the stressful math problems that followed self-affirmation compared to those that followed the non-affirmation control. Finally, we predicted that, during self-affirmation compared to non-affirmation control, there would be greater connectivity between reward-related regions (VMPFC and VS) and threat system regions, suggesting a linkage between the neural effects occurring during self-affirmation and the effects occurring in the brain after the affirmation process.

Methods

Participants

Twenty-seven university students (18 female; mean age = 19.3 years, s.d. = 1.35) completed study procedures. All participants met eligibility criteria for fMRI studies (right-handed, not claustrophobic, no metal). To ensure that students believed the performance feedback on the math stress task (that they were underperforming relative to their peers), only students in the arts, social sciences and humanities were recruited (Carnegie Mellon has high performing engineering and computer science programs). Twenty-five participants had usable neuroimaging data and were included in neuroimaging analyses (one was excluded for excessive motion, one participant's data were lost by the scanner). The Carnegie Mellon University Institutional Review Board approved all procedures and all participants gave informed consent.

Procedure

A week before the fMRI session, participants completed a survey in which they rated a list of personal values in order of

Self-Affirmation Trial

Non-Affirmation Trial
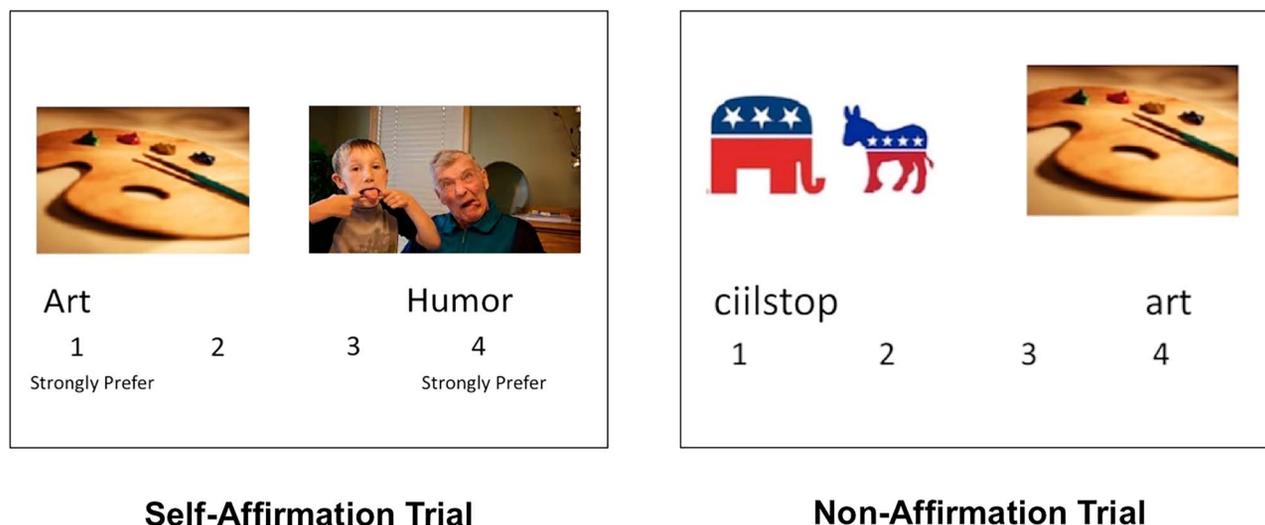
Fig. 1. Example trial displays for self-affirmation and non-affirmation control blocks.

importance to them. We then used these rankings to build a self-affirmation task for each individual participant (details below). When participants arrived for the fMRI scan session, they provided informed consent, were screened for scanner eligibility, and received instructions about the tasks they would be completing. Then participants underwent an fMRI scan in which they completed a high-resolution structural scan (MPRAGE), two runs of the self-affirmation and stressful math combined task, one run of a self-affirmation and threat reactivity task, and one run of a values statements task (always presented in this order). The latter two tasks are not reported here as they were designed to test different hypotheses from the task reported on in the present analyses. During the scan, participants viewed trials on a projector screen and were asked to make responses using a button box. After the scan, participants completed questionnaire measures, were debriefed and then dismissed.

Self-affirmation task. To manipulate self-affirmation, participants completed a task that had participants affirm their important values (self-affirmation) or a control condition (adapted from: Dutcher *et al.*, 2016). This task uses a standard self-affirmation decision making paradigm (Steele and Liu, 1983; Steele, 1988), in which participants were given a series of paired personal-value statements and were asked to indicate their relative preference (adapted from Vernon and Allport, 1931). The self-affirmation blocks were based on the online survey participants completed before their scheduled session. In these self-affirmation blocks, they saw a screen that said, 'Think about these values and indicate which value is more important to you.' The next screen displayed two values (see Figure 1), and participants indicated which value they preferred, on a scale from 1 to 4 [1 = strongly prefer (the value on the left), 2 = slightly prefer (the value on the left), 3 = slightly prefer (the value on the right), 4 = strongly prefer (the value on the right)]. Participants would see their top value paired against each other value, and which side the values appeared on screen was counterbalanced. As a non-affirmation control condition, participants completed alphabetizing trials that controlled for visual content (similar images were displayed). They were shown a screen that said, 'Which label is alphabetized?' The next screen displayed images with scrambled letter sequences below them (see Figure 1), and

participants pressed 1 if the letters in the sequence on the left were shown in alphabetical order and 4 if the letters in the sequence on the right were shown in alphabetical order. Both trial types were shown on screen for 6 s, in blocks of four trials (24 s per block). There were eight self-affirmation blocks and eight non-affirmation blocks across two runs.

Stress task. To manipulate stress levels, we had participants complete a math task with both stressful condition and non-stressful control conditions. In the stressful 'test' condition, participants completed math problems of varying difficulty and were told their performance would be compared to their peers at the university (based on: Inagaki *et al.*, 2016). In the control condition, participants completed 'practice' blocks that included trials with easy math problems (e.g. $4^2 + 1 =$) which they were instructed to calculate in their heads. Once they arrived at an answer, they pressed a button, and then saw a brief fixation followed by the next trial. No feedback was provided in the practice blocks. During the test blocks, participants were shown a more difficult math problem (e.g. $6^2/9 =$) and, under time pressure, were asked to select an answer from a set of possible answers shown on screen (see Figure 2). Participants then saw accuracy feedback: red text stating 'Incorrect. Please try harder.' if they were incorrect, blue text stating 'Correct!' if they were correct, and black text stating 'No response provided' if they did not respond in time. They then saw evaluative feedback regarding how their performance compared to the average of their peers at that point. Participants saw two horizontal bars on screen, one indicating their performance and one indicating their peers' performance. They were told their performance rating was based both on how quickly they responded, as well as how accurate they were. The feedback was designed so their performance bar started to fall behind the peer bar, such that they were performing worse over time relative to the average peer, which made these blocks both socially evaluative and stressful. At the end of each math block (both practice and test), participants were asked to rate how stressful that block was on a 1 (not at all) to 4 (a lot) scale. Blocks were 56-s long: eight math trials presented for 5 s each. In the practice blocks, trials were followed by 2 s of rest, while in the test blocks, trials were followed by 1 s accuracy feedback, and 1 s of evaluative feedback (see Figure 2).

A

B
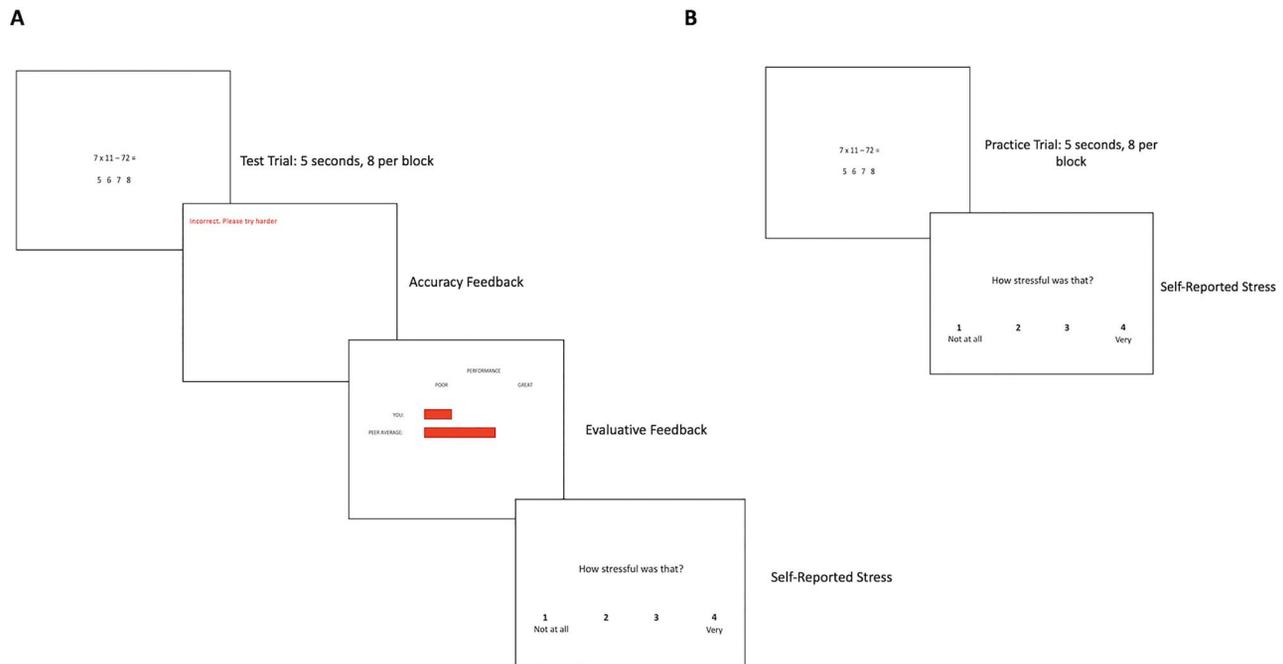
Fig. 2. (A) The images shown during the math test blocks and (B) math practice blocks.

Participants completed eight blocks of each condition across two runs.

Run design. In order to explore the effect of self-affirmation on stress responding, we crossed affirmation condition and stress level in a 2 × 2 design. Hence, the tasks were interspersed such that either a self-affirmation or a non-affirmation block preceded each type of math block. To obtain four blocks of every combination of stress and values blocks, this task was conducted across two runs of the scanner. This 2 (self-affirmation) × 2 (stress) design resulted in four conditions of interest: self-affirmation practice, non-affirmation practice, self-affirmation test and non-affirmation test. While the order of presentation of math blocks was always a practice block followed by a test block, the exact math problems and the presentation order of self-affirmation and non-affirmation blocks was counterbalanced based on four pseudorandom scripted presentation orders.

Image acquisition. Data were acquired on a Siemens Verio 3-T MRI scanner with foam padding surrounding each participant's head to reduce head movement. For each participant, we acquired a structural T2-weighted echo-planar imaging volume (MPRAGE)—spin-echo, repetition time (TR) = 2300 ms, echo time (TE) 1.97 ms, matrix size = 256 × 256, 1.0-mm isovoxel, field of view (FOV) = 256 mm, 176 slices, 1-mm thick, flip angle 9°, bandwidth = 240 Hz/Px (coplanar with the functional scans). Both runs of the values and math stress task lasted 706 s (11 min, 46 s)—gradient-echo, TR = 2000 ms, TE = 25 ms, multiband factor 3, matrix size 70 × 70, 3.0-mm isovoxel, FOV = 210 mm, 51 axial slices, 3-mm thick, flip angle 79°, bandwidth = 1930 Hz/Px, collected at a 38° slice angle.

Data analysis. Imaging data were analyzed using Statistical Parametric Mapping (SPM) software (SPM8; Wellcome Department of Cognitive Neurology, Institute of Neurology, London, England). Prior to preprocessing, images were manually reoriented to maximize preprocessing alignment quality. For preprocessing, functional and anatomical images were realigned, co-registered to the structural scan and normalized using the DARTEL procedure in SPM8. For the self-affirmation task, the 24-s values blocks were modeled as the self-affirmation condition and the 24-s alphabetize blocks were modeled as the non-affirmation condition. For the stress task, the practice math trials that came after a self-affirmation block were modeled as the self-affirmation practice condition, and the practice math trials that came after a non-affirmation block were modeled as the non-affirmation practice condition. Similarly, the test math trials that came after a self-affirmation block were modeled as the self-affirmation test condition, and the test math trials that came after a non-affirmation block were modeled as the non-affirmation test condition. The self-reported stress rating block was modeled separately from the math blocks. Implicit baseline consisted of the rest periods (viewing a fixation cross). Conditions were replicated across both runs of the task. Activation during each block was convolved with a canonical hemodynamic response function. Six motion parameters were included as nuisance predictors plus a predictor for each timepoint that the global signal change (GSC) exceeded 2.5 s.d. of the mean GSC or where estimated motion exceeds 1.5 mm of translation or 1.5° of rotation. We used a 128 Hz high-pass filter, and modeled serial autocorrelation as an AR (1) process.

We computed linear contrasts for each participant comparing BOLD signal for one main contrast of interest from the self-affirmation task: self-affirmation *vs* non-affirmation control. Moreover, we computed linear contrasts for each participant comparing BOLD signal for two main contrasts of interest from the stress task: non-affirmation test *vs* non-affirmation practice, and self-affirmation test *vs* non-affirmation test. These individual contrast images were then used in group-level analyses. We conducted whole brain analyses to gain a comprehensive picture of neural activity during these tasks. Due to this being the first study exploring the neural stress buffering effects of self-affirmation, whole brain analyses were calculated at an a

priori voxel extent threshold of 20 voxels, $P < 0.001$ to balance Type I and Type II errors (Lieberman and Cunningham, 2009), and the tables note which clusters are significant at $P < 0.05$ FDR-corrected.

To examine differences in functional connectivity between self-affirmation and non-affirmation control, we also conducted two PPI analyses. PPI analyses measure the functional connectivity between the time series of a chosen seed and the time series of the remaining voxels in the brain. Due to our a priori interest in the effect of prefrontal regions' influence on threat system activity, we seeded the first analysis with a VMPFC ROI. We built the ROI by taking the prefrontal cluster seen in the whole brain analysis of self-affirmation compared to the non-affirmation control and restricting it to the ventromedial portion (from $-14 < x < 14$, anterior to $y = 34$, and ventral to the corpus callosum, $-20 < z < -4$, see Figure 4B). We ran a second PPI analysis using a VS ROI. We built the ROI by taking the prefrontal cluster seen in the whole brain analysis of self-affirmation compared to the non-affirmation control and restricting it to the portion that overlapped with the VS (from $-24 < x < 24$, $4 < y < 18$, and $-12 < z < 0$). Using the generalized psychophysiological interaction toolbox in SPM (McLaren *et al.*, 2012), we first calculated PPI measures at the subject level for the self-affirmation compared to non-affirmation contrast. These first-level contrast estimates were then compared as a *t*-test at the group level of analysis to examine which regions in the brain demonstrated greater functional connectivity with VMPFC and VS during self-affirmation compared to non-affirmation control, using our a priori threshold.

Results

Preliminary analyses

There were no significant differences in stress ratings or number correct for any of the conditions based on the order trials were presented (all P's > 0.13), so all analyses presented will be collapsed across counterbalanced order.

As a manipulation check, we ran a 2 (self-affirmation condition) × 2 (stress level) ANOVA to ensure the stress task manipulated stress, as well as that the self-affirmation task effectively buffered stress responding. As predicted, there was a main effect of stress level such that participants rated the test math problems as more stressful ($M = 2.597$, s.d. $= 0.468$) than the practice math problems ($M = 1.281$, s.d. $= 0.426$), $F(1,26) = 98.34$, $P < 0.001$. Additionally, there was a significant main effect of self-affirmation condition, such that participants rated math test blocks as less stressful following the self-affirmation blocks ($M = 1.853$, s.d. $= 0.577$), than following the non-affirmation blocks ($M = 2.062$, s.d. $= 0.546$), $F(1,26) = 9.46$, $P = 0.005$. There was also a significant interaction between self-affirmation and stress on stress ratings (see Figure 3A), $F(1, 26) = 5.12$, $P = 0.032$. Consistent with previous research (Creswell *et al.*, 2005; Sherman *et al.*, 2009), paired samples *t*-tests revealed a significant effect such that individuals reported less stress to the math test blocks following self-affirmation ($M = 2.426$, s.d. $= 0.897$), compared to the math test blocks following non-affirmation ($M = 2.769$, s.d. $= 0.778$), $t(26) = -3.225$, $P = 0.003$. This effect was not observed for the practice math blocks following self-affirmation compared to non-affirmation, $t(26) = -1.055$, $P = 0.301$. Finally, we ran a paired samples *t*-test comparing the number of problems correct in each test condition, as only the test math problems required an answer response. Consistent with previous research (Creswell *et al.*, 2013), self-affirmation led to significantly more correct answers on the test math problems ($M = 14.15$,

s.d. $= 3.371$) compared to non-affirmation ($M = 12.33$, s.d. $= 3.679$), $t(26) = 3.399$, $P = 0.002$ (see Figure 3B). These results suggest that the stress task manipulated stress and the self-affirmation task effectively buffered stress responding and enhanced performance.

Neuroimaging analyses

First, we examined the difference in neural activity for the self-affirmation condition compared to the non-affirmation control condition. Whole brain analyses focusing on clusters with greater activity during self-affirmation compared to non-affirmation revealed a number of regions (see Table 1 for full list). Consistent with prior work (Dutcher *et al.*, 2016), self-affirmation led to more activity in a large cluster in medial and VMPFC that extended into the caudate head of the VS (see Figure 4A) [MNI coordinates: $-9, 51, 39$, $t = 8.504$, k (number of voxels) $= 2105$], as well as clusters in the posterior cingulate cortex ($-6, -51, 30$; $t = 10.126$, $k = 980$) and the angular gyrus ($-42, -63, 27$; $t = 10.224$, $k = 627$).

Next, we ran a whole brain analysis focusing on the stressful test trials that followed the non-affirmation condition compared to the non-stressful practice trials that followed the non-affirmation condition to explore the regions more active during the stressful condition compared to control. The non-affirmation test condition led to greater activity (compared to non-affirmation practice) in a number of regions including left and right AI and dACC (cluster: $-12, -87, -9$; $t = 15.303$, $k = 13\,966$) (for list, see Table 2).

To test our hypothesis that self-affirmation would lead to less threat-related neural activity compared to non-affirmation, we conducted a whole brain analysis on the stressful trials that followed the self-affirmation conditions compared to the stressful trials that followed the non-affirmation condition. This analysis produced one cluster of activity in the negative contrast, a cluster in left AI ($-18, 33, -3$, $t = -4.617$, $k = 29$) (see Table 3), such that the self-affirmation test condition led to less activation of the left AI compared to the non-affirmation test condition (see Figure 5). However, this cluster was not significant at the more stringent FDR corrected $P < 0.05$ value.

To examine whether there were differences in functional connectivity during self-affirmation compared to non-affirmation control, we conducted a PPI analysis using a VMPFC seed (see Figure 4B). We then examined the whole brain to determine which regions showed greater functional connectivity with the VMPFC during self-affirmation compared to non-affirmation. Relative to non-affirmation control, there was greater functional connectivity between the VMPFC and left AI ($-36, 30, -6$, $t = 4.536$, $k = 44$), right AI ($27, 30, -6$, $t = 4.940$, $k = 23$), left dorsolateral prefrontal cortex ($-39, 18, 30$, $t = 6.756$, $k = 495$) and supplementary motor area ($-6, 15, 51$, $t = 6.627$, $k = 227$) (for full list, see Table 4).

We also conducted a PPI analysis using a VS seed. We then examined the whole brain to determine which regions showed greater functional connectivity with the VS during self-affirmation compared to non-affirmation. There were no significant clusters in this analysis.

Discussion

The present study explored the effect of self-affirmation on neural and behavioral stress responding. Compared to non-affirmation control, self-affirmation led to lower self-reported feelings of stress and enhanced performance in response to the stressful math problems. These findings are consistent with

Table 1. Whole brain analysis of self-affirmation *vs* non-affirmation contrast ($k > 20$, $P < 0.001$)

Anatomical region	Brodmann's area	Hemisphere	MNI coordinates of peak voxel				
			x	y	z	t (24)	k
Self-affirmation > non-affirmation							
Cerebellum[a]		Right	30	−87	−36	7.055	253
Cerebellum[a]		Left	−27	−87	−39	6.450	87
Temporal pole[a]		Right	54	9	−18	6.922	456
Temporal pole[a]		Left	−60	0	−18	9.179	925
MPFC/VMPFC[a]	9/10/11	Left	−9	51	39	8.504	2105
Hippocampus		Right	27	−9	−18	4.421	42
Hippocampus		Left	−27	−30	−15	4.618	63
Occipital lobe[a]		Left	−9	−96	24	6.628	552
Posterior cingulate cortex[a]		Left	−6	−51	30	10.126	980
Temporoparietal junction[a]		Right	54	−60	24	7.665	546
Angular gyrus[a]	39	Left	−42	−63	27	10.224	627
Non-affirmation > self-affirmation							
Cerebellum		Left	−30	−66	−48	−5.008	26
Cerebellum			0	−54	−30	−5.111	22
Cerebellum		Left	−6	−75	−24	−5.375	27
Inferior occipital[a]		Right	27	−90	−3	−8.372	361
Inferior occipital[a]		Left	−27	−93	−6	−8.869	1057
Inferior frontal gyrus[a]		Right	42	9	27	−6.542	543
Inferior frontal gyrus[a]		Left	−33	−3	30	−6.818	324
Intraparietal sulcus[a]		Right	27	−51	42	−7.384	479

[a]Regions significant with FDR $P < 0.05$ correction.
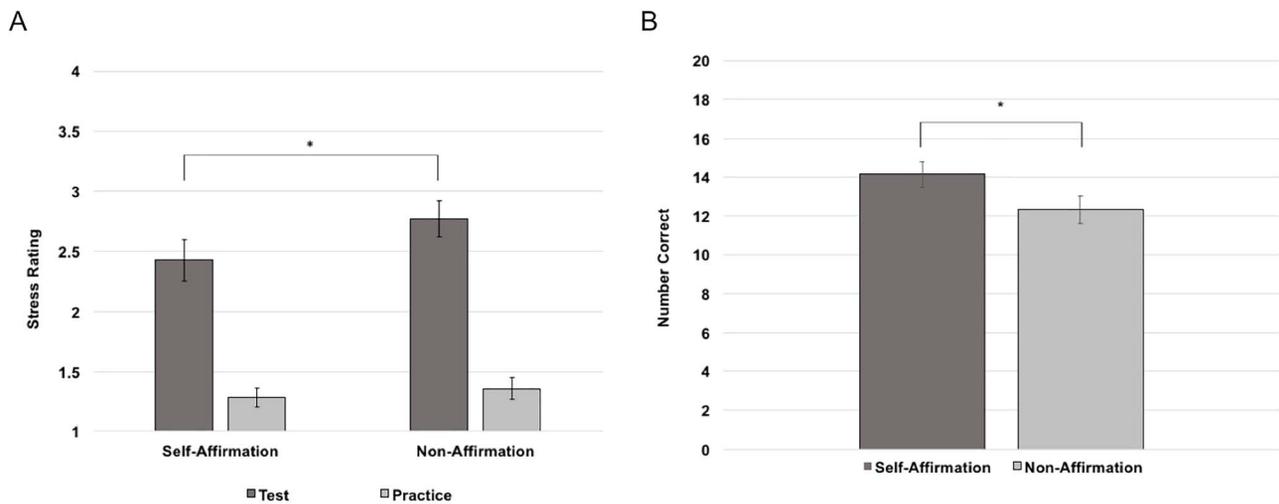
Fig. 3. (A) Interaction between condition and math difficulty on self-reported stress ratings. (B) The effect of condition on the number of math problems correctly answered during test trials. Error bars depict standard errors of the means. *$P < 0.05$.

prior behavioral work demonstrating self-affirmation to be an effective stress-reduction technique (Creswell *et al.*, 2005, 2013; Cohen *et al.*, 2009; Sherman *et al.*, 2009, 2013; Brady *et al.*, 2016; Goyer *et al.*, 2017) and show that this effect can also be observed in a within-subjects design. This has important implications for future laboratory-based studies, and, in particular, facilitates further examination of brain responses to self-affirmation and its subsequent effects.

The primary goal of this study was to characterize the neural mechanisms for self-affirmation's stress buffering effects. First, we replicated findings that self-affirmation (compared to control) elicited greater reward-related neural activity (VMPFC, in particular). We also showed that self-affirmation (*vs* control) led to less AI activity in response to stressful, evaluative math blocks.

We posit that this finding is consistent with neuroscience and behavioral research finding that rewarding stimuli or activities can mitigate stress responding (Dutcher and Creswell, 2018b). Critically, anatomical and structural investigations in the brain find pathways by which key hubs in the reward and threat systems in the brain are linked (Ulrich-Lai and Herman, 2009; Cloutman *et al.*, 2012; Dutcher and Creswell, 2018b), supporting the plausibility of this mechanism. In addition, this finding suggests that self-affirmation may be changing the way the brain responds to exposure to threatening or stressful information, consistent with previous behavioral work demonstrating that self-affirmation leads to changes in attentional bias towards threatening information (Klein and Harris, 2009) and ERP indicators of error responsiveness (Legault *et al.*, 2012).

Table 2. Whole brain analysis of non-affirmation test *vs* non-affirmation practice contrast ($k > 20$, $P < 0.001$)

Anatomical region	Brodmann's area	Hemisphere	MNI coordinates of peak voxel				
			x	y	z	t (24)	k
Non-affirmation test > non-affirmation practice							
Bilateral insula/dACC/precuneus/thalamus[a]		Left	−12	−87	−9	15.303	13 966
Non-affirmation practice > non-affirmation test							
Temporal pole		Left	−45	6	−30	−4.175	31
VMPFC		Right	42	39	−9	−4.367	32
DMPFC		Left	−18	39	45	−5.199	244
Dorsal posterior insula[a]		Right	39	−12	15	−6.740	175
Dorsal posterior insula		Left	−39	−15	21	−4.719	25
Angular gyrus[a]		Left	−45	−72	42	−6.927	295
Posterior cingulate cortex		Right	51	−57	30	−4.938	88
MPFC		Right	15	57	24	−4.386	20
Posterior cingulate cortex		Left	−12	−45	36	−4.169	30
Supplementary motor area		Right	9	−21	54	−4.172	62
DMPFC		Right	18	30	54	−4.716	65

[a]Regions significant with FDR $P < 0.05$ correction.

Table 3. Whole brain analysis of self-affirmation test *vs* non-affirmation test contrast ($k > 20$, $P < 0.001$)

Anatomical region	Brodmann's area	Hemisphere	MNI coordinates of peak voxel				
			x	y	z	t(24)	k
Self-affirmation test > non-affirmation test							
No clusters							
Non-affirmation test > self-affirmation test							
AI		Left	−18	33	−3	−4.617	29

No regions significant with FDR $P < 0.05$ correction.
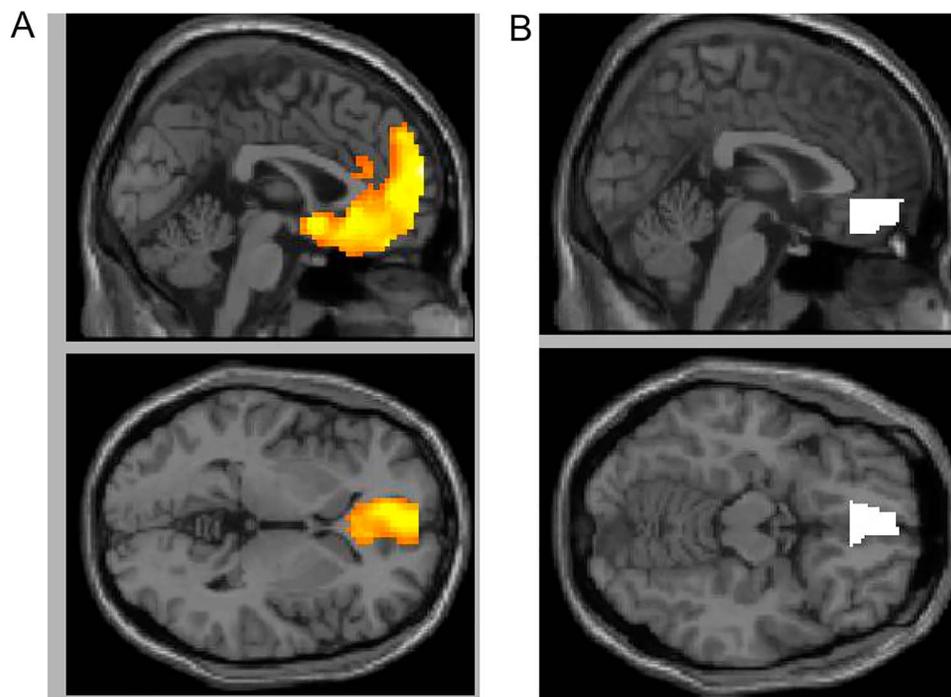
Fig. 4. (A) The prefrontal cluster from whole brain analysis of the self-affirmation > non-affirmation control contrast. (B) This cluster restricted to just the VMPFC. This seed was used in the PPI analysis.

Table 4. PPI analysis comparing functional connectivity in the VMPFC between self-affirmation *vs* non-affirmation control (k > 20, P < 0.001)

Anatomical region	Brodmann's area	Hemisphere	MNI coordinates of peak voxel				
			x	y	z	t (24)	k
Self-affirmation > non-affirmation							
Cerebellum		Left	−9	−75	−30	5.420	172
Occipital lobe[a]		Right	30	−93	−6	8.774	630
Occipital lobe[a]		Left	−30	−93	−6	9.516	678
Midbrain		Left	−15	−18	−9	4.432	32
AI		Right	27	30	−6	4.940	23
AI		Left	−36	30	−6	4.536	44
Thalamus		Right	3	−6	3	5.123	49
Thalamus		Right	15	−21	15	4.852	28
DLPFC[a]		Left	−39	18	30	6.756	495
Posterior parietal cortex[a]		Left	−27	−72	42	8.761	311
DLPFC		Right	48	33	27	5.069	38
Inferior frontal gyrus		Right	33	15	27	4.249	39
Posterior parietal cortex[a]		Right	30	−66	45	6.037	290
Supplementary motor area[a]		Left	−6	15	51	6.627	227
Non-affirmation > self-affirmation							
Occipital lobe		Left	−6	−96	21	−4.810	20
Posterior cingulate cortex		Right	3	−18	39	−4.470	39

[a]Regions significant with FDR P < 0.05 correction.
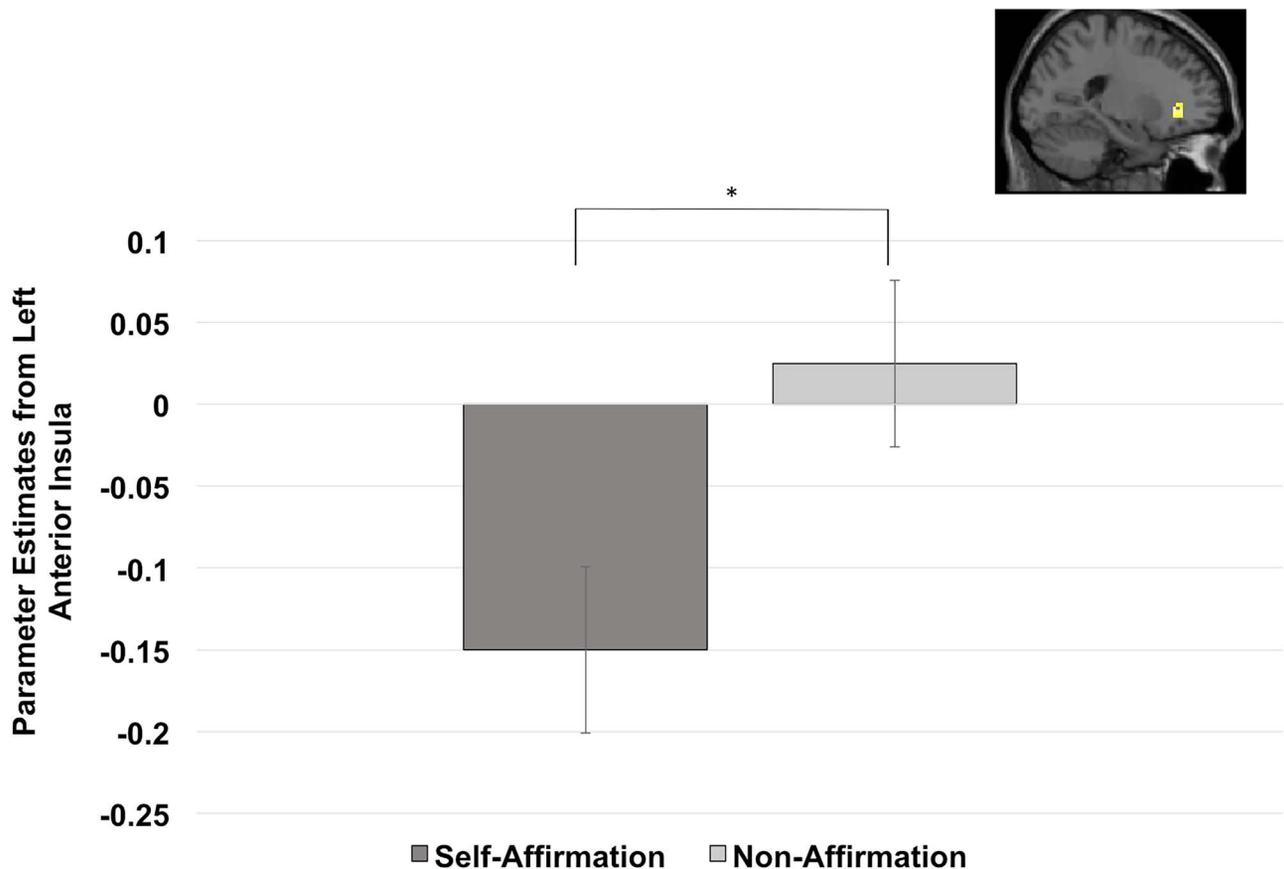
Fig. 5. Parameter estimates from the left AI cluster found in whole brain analysis comparing self-affirmation test to non-affirmation test. *P < 0.001.

Moreover, we explored the functional connectivity of reward-related regions during self-affirmation compared to non-affirmation control to test a pathway for stress buffering effects. Results found greater connectivity between the VMPFC and both left and right AI, suggesting that during self-affirmation, there is greater communication between these regions, mirroring structural AI connectivity (Saper, 1982; Chiba *et al.*, 2001; Cloutman *et al.*, 2012). There were no significant differences in functional connectivity for the VS during self-affirmation compared to non-affirmation control. While PPI analysis cannot

characterize the nature or direction of communication, animal tracing studies have identified projections from VMPFC to insular cortex, supporting the plausibility of VMPFC inputs on insula activity. Furthermore, in one study, activity in the VMPFC was associated with less AI activity to a subsequent threat (Eisenberger et al., 2011). In the present study, increased connectivity occurred during the self-affirmation task, and the subsequent response to stress was characterized by a relative decrease in left AI activity following self-affirmation. However, little work has explored functional connectivity between reward and threat system regions and whether greater connectivity indicates an enhanced ability to down-regulate neural stress responses. Future work can examine whether this increased functional connectivity leads to an inhibitory influence on subsequent AI responses to threat and if it is related to changes in physiological responses to stress.

While prior research has found that self-affirmation can buffer the physiological stress response (Creswell et al., 2005; Sherman et al., 2009; Spicer et al., 2016), this is the first study to examine its effects on neural responses during stress exposure. Activity in the AI has been linked to inflammatory and physiological responses to stress (Gianaros et al., 2005; Slavich et al., 2010; Muscatell and Eisenberger, 2012; Gianaros and Wager, 2015), suggesting these effects present a possible neural mechanism by which self-affirmation affects the physiological responses to stress. Future research can continue to explore this possibility by conducting studies that integrate fMRI with stress physiology measures to determine if reductions in AI activity to stress after a self-affirmation intervention are related to alterations in stress physiology (e.g. blood pressure, heart rate or cortisol).

We used a socially evaluative stress task to be consistent with the extant literature on self-affirmation effects on laboratory stress responding (Creswell et al., 2005, 2013; Sherman et al., 2009). Studies outside the laboratory often examine self-affirmation's effects on stressors that have social implications. It is possible, then, that self-affirmation's stress buffering effects are limited to social stressors. However, self-affirmation has been shown to mitigate the effect of ego depletion on later pain tolerance (Schmeichel and Vohs, 2009) and self-enhancement has been shown to lead to longer persistence in the ice bath during a cold-pressor task (O'Mara and Gaertner, 2017), suggesting that self-affirmation's stress buffering effects are not limited to social stressors. This work has not tested whether self-affirmation has differential effects on different types of stressors, thus future work should explore self-affirmation's stress buffering effects in the context of a variety of different stressors.

We hypothesized that self-affirmation would lead to less AI and dACC activity during stress compared to non-affirmation control. It was therefore surprising that results only support reduced AI activity during stress following self-affirmation, not dACC activity. While we are cautious not to over-interpret a null finding, the roles that dACC and AI play in threat detection and processing are known to differ (Muscatell and Eisenberger, 2012). The dACC is believed to signal distress (Spunt et al., 2012), whereas the AI is believed to be critically involved in awareness of physiology and bodily states (Craig and Craig, 2009). It is possible that self-affirmation's effects are stronger on AI responses to stress than dACC responses to stress, if self-affirmation has stronger effects on physiological stress responses than distress responses. While there was a decrease in self-reported stress to the stressful math problems that followed self-affirmation compared to non-affirmation control, these self-report findings did not correlate with VMPFC responses

to self-affirmation or AI responses during the stressful math trials, and not significantly to VMPFC-left AI connectivity (see Supplementary data for details on these analyses). Thus, self-affirmation might affect other processes in the brain that contribute to distress or stress appraisals that were not captured here, such as affecting the individual's perception of the resources they have to manage a stressor. This could influence retrospective self-reported stress at the end of the block while not necessarily influencing immediate dACC responses to the stress exposure. It will be important for future research to clarify if this dACC null effect is reliable and its interpretation.

There is significant scientific interest in empirically supported stress-reduction interventions. Yet there is still work to be done articulating the neural mechanisms for stress reduction effects and their health consequences (for a review: Dutcher and Creswell, 2018a). Indeed, few studies have explored the effects of any behavioral intervention on neural responses to stressful experience. One pilot study found that mindfulness training led to less stress reactivity in the amygdala and anterior/middle insula compared to control (Kober et al., 2017). Because this is the first study to explore the effect of self-affirmation on neural responses to stress, it also offers potential insight on the effect of stress reduction interventions more generally on neural responses to stress. It is possible that increased reward-related activity and subsequent decreased threat-related activity is a neural mechanism for other stress reduction interventions as well.

The dearth of findings in this area highlights the need for more health neuroscience research on stress-reduction pathways. There are a few limitations to the present work. First, the study design necessitated an acute laboratory stressor and a within-subjects design, which limits the generalizability to chronic stress conditions. However, this study serves as a 'proof of concept' for future studies aiming to explore the effect of self-affirmation on chronic stress and the consequences therein. Additionally, participants were between the ages of 18–23, and much of the work on self-affirmation's stress buffering effects has been in young adults and adolescents. However, studies of the neural correlates of self-affirmation have been conducted across a range of ages, suggesting that the initial reward-related activity associated with self-affirmation is not limited to this age group (Cascio et al., 2016; Dutcher et al., 2016; Kang et al., 2018). Some results were not significant at more stringent FDR-corrected thresholds, including the reduced left AI activity during self-affirmation test compared to non-affirmation test. These effects may or may not be stronger in a larger sample. Thus, the age group and the smaller sample size in this study suggest that it will be important to replicate these findings in a larger study, with broader range of participant demographics. Finally, in order to ensure the believability of the performance feedback, participants were given showing they were underperforming compared to their peers, we recruited humanities, social sciences and art students. While this may have weakened the self-relevance of the task, participants still found the task stressful, providing the opportunity to test self-affirmation's stress buffering effects.

The present results found that, consistent with previous work, self-affirmation led to lower self-reported stress and enhanced performance to a socially evaluative experience compared to a non-affirmation control. Importantly, self-affirmation also led to less AI activity to the stressful task compared to control, suggesting that self-affirmation can reduce both affective and neural responses to stress exposure, and mitigate some of the performance consequences of

stress. Furthermore, the AI is a key region for facilitating the physiological stress response (Gianaros and Wager, 2015), perhaps serving as a mechanism for self-affirmation's effects on stress physiology (Creswell *et al.*, 2005; Sherman *et al.*, 2009; Spicer *et al.*, 2016). While this and previous work have primarily been conducted in laboratory studies, they underscore the effects a self-affirmation intervention could have on stress-related health outcomes. Considering the rising rates of chronic stress (American Psychological Association, 2018), the negative effects stress has on health (Cohen *et al.*, 2007), and the cost-effective and easy to implement nature of a self-affirmation intervention, the present findings suggest self-affirmation interventions for stress and stress-related health are worth exploring.

Supplementary data

Supplementary data are available at *SCAN* online.

Acknowledgements

Conflict of interest

The authors have no conflicts of interest to report with this work.

References

American Psychological Association (2018). *Stress in America: generation Z. Stress in America Survey*, p. 11.

Brady, S.T., Reeves, S.L., Garcia, J., *et al.* (2016). The psychology of the affirmed learner: spontaneous self-affirmation in the face of stress. *Journal of Educational Psychology*, **108**(3), 353.

Cascio, C.N., O'Donnell, M.B., Tinney, F.J., *et al.* (2016). Self-affirmation activates brain systems associated with self-related processing and reward and is reinforced by future orientation. *Social Cognitive and Affective Neuroscience*, **11**(4), 621–9. doi: 10.1093/scan/nsv136.

Chiba, T., Kayahara, T., Nakano, K. (2001). Efferent projections of infralimbic and prelimbic areas of the medial prefrontal cortex in the Japanese monkey, *Macaca fuscata. Brain Research*, **888**(1), 83–101.

Cloutman, L.L., Binney, R.J., Drakesmith, M., Parker, G.J., Ralph, M.A.L. (2012). The variation of function across the human insula mirrors its patterns of structural connectivity: evidence from in vivo probabilistic tractography. *NeuroImage*, **59**(4), 3514–21.

Cohen, G.L., Sherman, D.K. (2014). The psychology of change: self-affirmation and social psychological intervention. *Annual Review of Psychology*, **65**, 333–71.

Cohen, G.L., Garcia, J., Apfel, N., Master, A. (2006). Reducing the racial achievement gap: a social-psychological intervention. *Science*, **313**(5791), 1307–10.

Cohen, S., Janicki-Deverts, D., Miller, G.E. (2007). Psychological stress and disease. *JAMA*, **298**(14), 1685–7.

Cohen, G.L., Garcia, J., Purdie-Vaughns, V., Apfel, N., Brzustoski, P. (2009). Recursive processes in self-affirmation: intervening to close the minority achievement gap. *Science*, **324**(5925), 400–3.

Cohen, S., Gianaros, P.J., Manuck, S.B. (2016). A stage model of stress and disease. *Perspectives on Psychological Science*, **11**(4), 456–63.

Craig, A.D., Craig, A.D. (2009). How do you feel–now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, **10**(1), 59–70.

Creswell, J.D., Welch, W.T., Taylor, S.E., Sherman, D.K., Gruenewald, T.L., Mann, T. (2005). Affirmation of personal values buffers neuroendocrine and psychological stress responses. *Psychological Science*, **16**(11), 846–51.

Creswell, J.D., Dutcher, J.M., Klein, W.M.P., Harris, P.R., Levine, J.M. (2013). Self-affirmation improves problem-solving under stress. *PLoS One*, **8**(5), e62593. doi: 10.1371/journal.pone.0062593.

Crocker, J., Niiya, Y., Mischkowski, D. (2008). Why does writing about important values reduce defensiveness? *Psychological Science*, **19**(7), 740–7.

Dedovic, K., Renwick, R., Mahani, N.K., Engert, V., Lupien, S.J., Pruessner, J.C. (2005). The Montreal imaging stress task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience*, **30**(5), 319.

Dedovic, K., Duchesne, A., Andrews, J., Engert, V., Pruessner, J.C. (2009). The brain and the stress axis: the neural correlates of cortisol regulation in response to stress. *NeuroImage*, **47**(3), 864–71.

Dutcher, J.M., Creswell, J.D. (2018a). Behavioral interventions in health neuroscience. *Annals of the New York Academy of Sciences*, **1428**(1), 51–70.

Dutcher, J.M., Creswell, J.D. (2018b). The role of brain reward pathways in stress resilience and health. *Neuroscience & Biobehavioral Reviews*, **95**, 559–67.

Dutcher, J.M., Creswell, J.D., Pacilio, L.E., *et al.* (2016). Self-affirmation activates the ventral striatum: a possible reward-related mechanism for self-affirmation. *Psychological Science*, **27**(4), 455–66. doi: 10.1177/0956797615625989.

Eisenberger, N.I., Cole, S.W. (2012). Social neuroscience and health: neurophysiological mechanisms linking social ties with physical health. *Nature Neuroscience*, **15**(5), 669–74. doi: 10.1038/nn.3086.

Eisenberger, N.I., Master, S.L., Inagaki, T.K., Taylor, S.E., Shirinyan, D., Lieberman, M.D. (2011). Attachment figures activate a safety signal-related neural region and reduce pain experience. *Proceedings of the National Academy of Sciences*, **108**(28), 11721–6. doi: 10.1073/pnas.1108239108.

Epton, T., Harris, P.R., Kane, R., van Koningsbruggen, G.M., Sheeran, P. (2015). The impact of self-affirmation on health-behavior change: a meta-analysis. *Health Psychology*, **34**(3), 187–96. doi: 10.1037/hea0000116.

Ferrer, R.A., Cohen, G.L. (2019). Reconceptualizing self-affirmation with the trigger and channel framework: lessons from the health domain. *Personality and Social Psychology Review*, **23**(3), 285–304.

Gianaros, P.J., Wager, T.D. (2015). Brain-body pathways linking psychological stress and physical health. *Current Directions in Psychological Science*, **24**(4), 313–21.

Gianaros, P.J., Derbtshire, S.W., May, J.C., Siegle, G.J., Gamalo, M.A., Jennings, J.R. (2005). Anterior cingulate activity correlates with blood pressure during stress. *Psychophysiology*, **42**(6), 627–35.

Goyer, J.P., Garcia, J., Purdie-Vaughns, V., *et al.* (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences USA*, **114**(29), 7594–9.

Inagaki, T.K., Bryne Haltom, K.E., Suzuki, S., *et al.* (2016). The neurobiology of giving versus receiving support: the role of stress-related and social reward–related neural activity. *Psychosomatic Medicine*, **78**(4), 443–53. doi: 10.1097/PSY.0000000000000302.

Kang, Y., Cooper, N., Pandey, P., *et al.* (2018). Effects of self-transcendence on neural responses to persuasive messages and health behavior change. *Proceedings of the National Academy of Sciences USA*, **115**(40), 9974–9.

Klein, W.M., Harris, P.R. (2009). Self-affirmation enhances attentional bias toward threatening components of a persuasive message. *Psychological Science*, **20**(12), 1463–7.

Kober, H., Brewer, J.A., Height, K.L., Sinha, R. (2017). Neural stress reactivity relates to smoking outcomes and differentiates between mindfulness and cognitive-behavioral treatments. *NeuroImage*, **151**, 4–13.

Legault, L., Al-Khindi, T., Inzlicht, M. (2012). Preserving integrity in the face of performance threat: self-affirmation enhances neurophysiological responsiveness to errors. *Psychological Science*, **23**(12), 1455–60.

Lieberman, M.D., Cunningham, W.A. (2009). Type I and type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, **4**(4), 423–8.

McLaren, D.G., Ries, M.L., Xu, G., Johnson, S.C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *NeuroImage*, **61**(4), 1277–86.

Muscatell, K.A., Eisenberger, N.I. (2012). A social neuroscience perspective on stress and health. *Social and Personality Psychology Compass*, **6**(12), 890–904.

O'Mara, E.M., Gaertner, L. (2017). Does self-enhancement facilitate task performance? *Journal of Experimental Psychology: General*, **146**(3), 442.

Saper, C.B. (1982). Convergence of autonomic and limbic connections in the insular cortex of the rat. *Journal of Comparative Neurology*, **210**(2), 163–73.

Schmeichel, B.J., Vohs, K. (2009). Self-affirmation and self-control: affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology*, **96**(4), 770–82.

Sherman, D.K., Bunyan, D.P., Creswell, J.D., Jaremka, L.M. (2009). Psychological vulnerability and stress: the effects of self-affirmation on sympathetic nervous system responses to naturalistic stressors. *Health Psychology*, **28**(5), 554.

Sherman, D.K., Hartson, K.A., Binning, K.R., *et al.* (2013). Deflecting the trajectory and changing the narrative: how self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, **104**(4), 591. doi: 10.1037/a0031495.

Slavich, G.M., Way, B.M., Eisenberger, N.I., Taylor, S.E. (2010). Neural sensitivity to social rejection is associated with inflammatory responses to social stress. *Proceedings of the National Academy of Sciences*, **107**(33), 14817–22.

Spicer, J., Shimbo, D., Johnston, N., *et al.* (2016). Prevention of stress-provoked endothelial injury by values affirmation: a proof of principle study. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, **50**(3), 471–9.

Spunt, R.P., Lieberman, M.D., Cohen, J.R., Eisenberger, N.I. (2012). The phenomenology of error processing: the dorsal ACC response to stop-signal errors tracks reports of negative affect. *Journal of Cognitive Neuroscience*, **24**(8), 1753–65.

Steele, C.M. (1988). The psychology of self-affirmation: sustaining the integrity of the self. *Advances in Experimental Social Psychology*, **21**, 261–302.

Steele, C.M., Liu, T.J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology*, **45**(1), 5–19. doi: 10.1037/0022-3514.45.1.5.

Ulrich-Lai, Y.M., Herman, J.P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience*, **10**(6), 397–409.

Vernon, P.E., Allport, G.W. (1931). A test for personal values. *The Journal of Abnormal and Social Psychology*, **26**(3), 231.

Wang, J., Rao, H., Wetmore, G.S., *et al.* (2005). Perfusion functional MRI reveals cerebral blood flow pattern under psychological stress. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(49), 17804–9. doi: 10.1073/pnas.0503082102.

Younger, J., Aron, A., Parke, S., Chatterjee, N., Mackey, S. (2010). Viewing pictures of a romantic partner reduces experimental pain: involvement of neural reward systems. *PLoS One*, **5**(10), e13309. doi: 10.1371/journal.pone.0013309.