Self-affirmation activates brain systems associated with self-related processing and reward and is reinforced by future orientation

Christopher N. Cascio[1], Matthew Brook O'Donnell[1], Francis J. Tinney, Jr.[2], Matthew D. Lieberman[3], Shelley E. Taylor[3], Victor J. Strecher[2], & Emily B. Falk[1]

University of Pennsylvania[1], University of Michigan[2], University of California, Los Angeles[3]

Correspondence:
Christopher N. Cascio
Annenberg School for Communication
University of Pennsylvania
3620 Walnut St.
Philadelphia, PA 19104
ccascio@asc.upenn.edu

Emily B. Falk
Annenberg School for Communication
University of Pennsylvania
3620 Walnut St.
Philadelphia, PA 19104
falk@asc.upenn.edu

Total words in main text <5332>

Acknowledgements: We thank Holly Derry, Ian Moore, and Michele Demers for assistance in developing intervention materials, and the staff of the University of Michigan fMRI Center for support and assistance with fMRI data acquisition. We thank Angela Fagerlin, Thad Polk, Lawrence An, Kennith Resnicow and the Michigan CECCR for support in realizing this project and Sonya Dal Cin and Sara Konrath for helpful discussions. This research was supported by The Michigan Center of Excellence in Cancer Communication Research/NIH-P50 CA101451 (PI Strecher), a NIH New Innovator Award/1DP2DA03515601 (PI Falk) and NIH/NCI 1R01CA180015-01 (PI Falk). The authors thank Kristin Shumaker, Nicolette Gregor, Alison Sagon for assistance with data collection, and Jonathan Mitchell for advice regarding processing of the accelerometer data.

For Peer Review

Abstract

Self-affirmation theory posits that people are motivated to maintain a positive self-view and that threats to perceived self-competence are met with resistance. When threatened, self-affirmations can restore self-competence by allowing individuals to reflect on sources of self-worth, such as core values. Many questions exist, however, about the underlying mechanisms associated with self-affirmation. We examined the neural mechanisms of self-affirmation with a task developed for use in a functional magnetic resonance imaging (fMRI) environment. Results of a region of interest analysis demonstrated that participants who were affirmed (compared to unaffirmed participants) showed increased activity in key regions of the brain's self-processing (MPFC+PCC) and valuation (VS+VMPFC) systems when reflecting on future oriented core values (compared to everyday activities). Furthermore, this neural activity went on to predict changes in sedentary behavior consistent with successful affirmation in response to a separate physical activity intervention. These results highlight neural processes associated with successful self-affirmation, and further suggest that key pathways may be amplified in conjunction with prospection.

Keywords: self-affirmation, fMRI, reward, positive valuation, emotion regulation

For Peer Review

Introduction

It is well documented that people seek to maintain a positive self-view and that threats to perceived self-competence across many domains are met with resistance (Sherman & Cohen, 2006). A large body of literature, however, demonstrates that a class of interventions called self-affirmations have benefits across threatening situations; affirmations can decrease stress, increase well being, improve academic performance and make people more open to behavior change (for a review, see (Cohen & Sherman, 2014)). Self-affirmations are acts that affirm one's self-worth, often by having individuals reflect on core values, which may give individuals a broader view of the self. This in turn can allow individuals to move beyond specific threats to self-integrity or self-competence (Cohen & Sherman, 2014; Steele, 1988).

Effects associated with self-affirmation interventions often occur without explicit awareness (Sherman et al., 2009). This lack of awareness makes it difficult for individuals to introspect on their experience and makes it difficult for researchers to examine specific underlying mechanisms that lead from the affirmation experience to behavioral change. Neuroimaging methods offer one way to examine a set of processes underlying self-affirmation interventions at the point of actual affirmation exposure, without the need for individuals to reflect on their experience (Falk et al., 2015); however, the neural mechanisms that underpin acts of self-affirmation have not been studied (Cohen & Sherman, 2014). Understanding the underlying neural mechanisms associated with self-affirmation will help to further expand our theoretical understanding of the processes at play during self-affirmation and may contribute to the development of more effective interventions. Thus, our first research question centers on the neurocognitive processes associated with the act of self-affirmation. Furthermore, the core brain systems involved in self-related processing and reward, that we hypothesize to be involved in

For Peer Review

affirmation, overlap with past studies of temporal orientation (i.e., considering events in the past and future; (D'Argembeau et al., 2010; D'Argembeau, Xue, Lu, Van der Linden, & Bechara, 2008)). Thus, our second research question focuses on whether the neural pathways to self-affirmation might be amplified in conjunction with specific temporal orientations.

Potential pathways to self-affirmation

One account of why self-affirmations are successful is attributed to their ability to broaden a person's overall perspective and reduce the effect of negative emotions (Sherman, 2013; Cohen & Sherman, 2014). For example, researchers have suggested that self-affirmations remind individuals of psychosocial resources that extend beyond a specific threat, which allows them to focus on sources of positive self-worth that transcends the threat. This in turn is thought to reduce reactivity to the threat and protect overall psychological wellbeing (Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009; Cook, Purdie-Vaughns, Garcia, & Cohen, 2012; Koole, Smeets, Van Knippenberg, & Dijksterhuis, 1999; Sherman et al., 2013).

Such effects might arise through several different pathways. First, affirmations may increase focus on sources of positive value to individuals. Self-affirmation interventions often rely on having participants reflect on personal core values and rewarding experiences. This pathway would engage neural mechanisms associated with reward and positive valuation. A recent meta-analysis demonstrates that brain regions most prominently involved in reward and positive valuation include the ventral striatum (VS) and ventral medial prefrontal cortex (VMPFC; (Bartra, McGuire, & Kable, 2013)).

Related to the pathway described above, affirmations could also work by focusing people on sources of positive self-worth, such as personal successes. This may also involve specific reflection on personal attributes outside of the threat (Sherman & Hartson, 2011; Sherman,

For Peer Review

2013). Meta-analyses across a variety of tasks find that self-related processing is most often associated with increased activity in the medial prefrontal cortex (MPFC) and posterior cingulate cortex (PCC; (Denny, Kober, Wager, & Ochsner, 2012; Northoff et al., 2006)). Thus, if self-affirmations succeed due to a boost in self-related processing prior to threat exposure, activity in the MPFC and PCC should increase during affirmation.

Furthermore, self-affirmations may allow for more efficient use of psychological resources needed to deal with the incoming threat (Sherman, 2013). This has been demonstrated in studies that examine the success of self-affirmation interventions in counteracting manipulations that reduce available cognitive and psychological resources (e.g., cognitive load and ego-depletion manipulations; (Schmeichel & Vohs, 2009; Vohs & Faber, 2007; Burson et al., 2012; Logel & Cohen, 2012)). Although these studies find evidence that affirmation interventions can reduce threat, it is unclear which psychological resources are actually involved in this process. One possible source of regulatory resources include the ventrolateral prefrontal cortex (VLPFC) and anterior cingulate cortex (ACC), which have been implicated in regulation of emotion and facilitating difficult choices (Marsh, 2007; Oschner et al., 2004; Wager et al., 2008). Self-affirmations may work by priming these regions to regulate emotions.

Affirmation and temporal orientation

Activity within several of our key self-related processing and reward regions of interest (ROIs) changes with manipulations of temporal focus. Although self-affirmation interventions have been successfully carried out using manipulations that focus both on past experiences as well as future goals (for a review, see; (McQueen & Klein, 2006)), temporal orientation has not been a core focus of affirmation research. Given the overlap between brain systems hypothesized to support affirmation effects and to support temporal orientation effects, however, we examined

For Peer Review

whether neural responses in brain systems associated with successful self-affirmation might change with or be reinforced by temporal focus.

For example, it has been found that imagining future personally relevant, emotionally positive, and rewarding events is associated with changes in VMPFC, striatum, MPFC and PCC (Benoit, Gilbert, & Burgess, 2011; Benoit, Szpunar, & Schacter, 2014; D'Argembeau et al., 2010, 2008). Increased activity in the MPFC has also been shown to positively correlate with imagining positive (vs. negative) future episodes (D'Argembeau et al., 2008) and such activation is further associated with projected reward value of the imagined future (Benoit et al., 2011, 2014). In addition, a recent meta-analysis found that increased activity in the MPFC and PCC, among other regions, was associated with thinking about hypothetical (e.g., future) compared to past episodes (Benoit & Schacter, 2015). Furthermore, another recent meta-analysis examining neural correlates associated with personal goals, future thinking, and mind wandering found that the MPFC is consistently activated in all three domains (Stawarczyk & D'Argembeau, 2015). These studies support the idea that mentally simulating future events, especially those relevant to personal goals, involves key regions hypothesized to be involved in self-affirmation interventions, including the VMPFC, MPFC, and PCC. Thus, if both future oriented thought and self-affirmation rely on similar neural mechanisms, they may mutually reinforce one another.

Importantly, these differences are not limited to neural activity. For example participants have better memory recall when encoding new information coupled with imagined scenarios that plan for the future, in comparison to remembering past events or events that are considered without a time relationship (Klein, Robertson, & Delton, 2010). In addition, mental simulations focusing on future events have been shown to benefit goal planning and one's psychological wellbeing (for a review, see; Schacter, 2012). Taken together, all of these studies reinforce the

For Peer Review

hypothesis that engagement of our key ROIs may differ by temporal focus, and that future orientation may reinforce the effects of reflecting on personally relevant core values. Thus, we examined temporal focus as a potential moderator of neural responses in our key ROIs during affirmation.

The current study

In sum, the current study aims to elucidate the underlying mechanisms associated with self-affirmation by examining participants' neural activity during a self-affirmation task specifically designed for functional magnetic resonance imaging (fMRI). We tested the extent to which exposure to self-affirmation produced increases in brain systems associated with positive valuation (VS+VMPFC), self-related processing (MPFC+ PCC), and emotion regulation (rACC+rVLPFC). In addition, we examined whether the neural effects of affirmation are moderated by temporal orientation (past vs. future). We validated our fMRI-compatible self-affirmation intervention in relation to its ability to increase receptivity to a subsequent set of health messages designed to reduce sedentary behavior in sedentary adults (Falk et al., 2015).

Methods

Participants

Participants (N=67; self-affirmed=33; unaffirmed=28) were adults between the ages of 18-64 (41 females; mean age=33.42 years, *SD*=13.04; 44 White, 12 Black, 3 Asian, 1 Hispanic, 7 Other), recruited as part of a study examining neural correlates of exposure to health messages that encouraged physical activity behavior in sedentary adults. All participants were sedentary (participants self-reported an estimated < 195 minutes of combined walking, moderate or vigorous activity per week at the time of recruitment). This was defined by mean activity on short-form International Physical Activity Questionnaire (IPAQ) at the time of recruitment. On

For Peer Review

average, recruited participants reported 23.5 minutes/week of activity, *SD*=49.5; mean BMI = 27.99, *SD*=6.84, indicating that on average participants were in the overweight category. Participants were right-handed, did not suffer from claustrophobia, were not currently taking any psychoactive medications, had normal (or corrected to normal) vision, and did not have metal in their body that was contraindicated for MRI (see supplemental materials for additional sample details).

Study design

Participants completed a three-part study (see Figure 1). At baseline, participants ranked a list of 8 personal values, completed self-report questionnaires and were fitted with an accelerometer to measure physical activity behavior. One week later participants completed an fMRI appointment in which they underwent the fMRI-compatible self-affirmation (or control) intervention. All participants then saw potentially-threatening messages encouraging physical activity and the success of the affirmation manipulation was validated based on objectively measured physical activity/sedentary behavior change attributable to self-affirmation in the subsequent month. Additional details on the sample and task session can be found in Falk et al., 2015, however, the neural processes associated with the actual affirmation task have not been previously examined.

Self-affirmation task

During the initial baseline appointment participants were asked to rank a list of 8 values from least to most valued, "*Please order the following values according to how important they are to you*". The list of 8 values included, creativity, relations with family and friends, sense of humor, independence, business or earning money, politics, religious values, and spontaneity or living life in the moment. These values were then used in the MRI portion of the self-affirmation

For Peer Review

task, such that participants in the affirmed condition reflected on their top ranked value and participants in the control condition reflected on their lowest ranked value.

Although there are many approaches to self-affirmation, two of the most prominent approaches ask individuals to write about a highly ranked personal value or to respond to questionnaires containing questions relevant to a highly ranked personal value (Cohen & Sherman, 2014; McQueen & Klein, 2006; Sherman, 2013). Typically self-affirmation writing tasks instruct affirmed participants to write for a period of time on one of their core values; control groups typically write on a topic that is not valued (McQueen & Klein, 2006; Napper, Harris, & Epton, 2009). Similarly, value scales involve the completion of questionnaires that allow participants to express their identification with the core value and why their core value is important to them; control participants complete questionnaires about topics of lower personal value and importance (Sherman, 2013). As in other widely used affirmation manipulations (see Cohen & Sherman, 2015; McQueen & Klein, 2006 for reviews), there were some differences in the values most consistently ranked as top and bottom values in the current study, however, there is also substantial overlap in values used in the affirmation and control conditions (for distribution, see; table 1).

To save time and standardize instructions, participants received task instructions for the affirmation task during the structural scan, directly prior to the task. To start the task preparation, participants were initially instructed during the structural scan to "*Please think about an experience you had involving* [VALUE]", where [VALUE] was replaced with their assigned value. This was followed by instructions to "*Try and visualize yourself in the experience and remember as many specific details as possible*". Participants were then prompted with phrases to help keep them prepare for the main affirmation task. Example statements

For Peer Review

included "*Think about when the experience occurred*" and "*Think about how you currently feel about this experience".* Once participants had come up with scenarios relevant to their top (or bottom) ranked value during the structural scan, they completed the main self-affirmation task during functional scanning.

To test the interaction between affirmation and temporal orientation, the main fMRI affirmation task instructed participants on different trials to think about a time when value-relevant scenarios had occurred (past) and when parallel scenarios could occur (future). Participants in both affirmation and control conditions were presented with prompts for scenarios focused on value statements as well as everyday activities (as a within subjects control condition). All participants were presented with the same control (everyday) activity scenarios. Example statements relating to experience of a specific value in the past or future condition are as follows [value = friends and family]: "*Think about a time in the past when you had fun with family and friends*", "*Think about a time in the future when you might be having fun with family and friends*". Example everyday statements included: "*Think about a time in the past when you charged your cell phone*", "*Think about a time in the future when you might charge your cell phone*". Importantly it should be noted that all statements (value and control) were focused on oneself and not subject to factual knowledge. This was done in order to have the distinction between high and low values pertain more to the importance placed on the topic rather than on topic knowledge. For example, for those who were assigned to think about what we referred to as "politics", the statements were not about politicians but rather how political values might be manifest in one's life (e.g., *Think about a time in the future when you might read about current events; Think about a time in the future when you might be inspired by people taking political action*). The self-affirmation task used a 2X2 block design, (past vs. future)X(everyday vs.

For Peer Review

value). Each block consisted of exposure to the scenario prompt for 12 seconds in which participants thought about the given statement and responded by pressing a button with their index figure each time they thought of a personally relevant example associated with the given statement. Participants were instructed to think about as many examples as they could for each scenario. Ten different scenarios were given for each condition (past value, past everyday, future value, and future everyday) for a total of 40 blocks in the task. Value-specific scenarios were created based on reflections one may have when engaging in a self-affirmation writing exercise, whereas everyday scenarios were created to represent common events that occur on a daily basis. Participants saw a fixation cross for 2 and 12 (every fifth trial) seconds between each block.

Validation of the fMRI self-affirmation intervention

Following their randomly assigned affirmation or control intervention, all participants were exposed to the same health messages encouraging increased physical activity and decreased sedentary behavior. The success of the affirmation intervention was validated using behavior change effects attributable to the experimental manipulation of self-affirmation. More specifically, aggregate measures of sedentary behavior were created measuring pre and post intervention activity captured for one week prior and one month following the intervention using triaxial accelerometers, and compared by condition. For further details on the health messaging task and accelerometer data collection and analysis, please see (Falk et al., 2015).

fMRI data acquisition and data analysis

Imaging data were acquired using a 3 Tesla GE Signa MRI scanner. One functional run was acquired for each participant (323 volumes total[1]). Functional images were recorded using a

[1] Note: For the first six participants (1 control, 5 affirmed due to the randomizer), a slightly longer (2 run) version of the task was used, in which the blocks were 16 seconds long instead of 12, and the affirmation task was split into two runs of 209 volumes each. These participants were the first to do the study and we initially had longer scan time that included an extra 4 seconds for each block.

For Peer Review

reverse spiral sequence (TR=2000ms, TE=30ms, flip angle=90°, 43 axial slices, FOV=220mm, slice thickness=3mm; voxel size=3.44x3.44x3.0mm). We also acquired in-plane T1-weighted images (43 slices; slice thickness=3mm; voxel size=.86x.86x3.0mm) and high-resolution T1-weighted images (SPGR; 124 slices; slice thickness=1.02x1.02x1.2mm) for use in coregistration and normalization.

Functional data were pre-processed and analyzed using Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK; please see supplemental materials for details of preprocessing stream). Data were modeled using the general linear model as implemented in SPM8. Four trial types were modeled: past value scenarios, future value scenarios, past everyday scenarios, future everyday scenarios; fixation trials were not modeled and constituted an implicit baseline. The six rigid-body translation and rotation parameters derived from spatial realignment were also included as nuisance regressors. Data were high-pass filtered with a cutoff of 128s.

Region of interest analysis

To test the balance of activity within brain networks involved in positive valuation and reward (VS+VMPFC), self-related processing (MPFC+PCC) and regulating emotions (rACC+rVLPFC), we first conducted *a priori* defined ROI analyses on each network of interest independently. Percent signal change scores were extracted from each combined network ROI contrasting the value > everyday scenarios; past value > past everyday scenarios; and future value > future everyday scenarios for each participant (see supplementary materials for ROI definitions and analysis details).

To investigate neural processes associated with self-affirmation that extended beyond our main ROI analyses, we subsequently conducted whole brain analyses examining differences

For Peer Review

between the affirmed and control participants for each of our main target contrasts: value > everyday scenarios; past value > past everyday scenarios; and future value > future everyday scenarios. All analyses are reported with a threshold of p=.005, K=35, corrected for multiple comparisons based on a Monte Carlo simulation using AlphaSim (Ward, 2000). Furthermore, based on *a priori* hypotheses linking valuation activity (VS+VMPFC) to self-affirmation processes, the relatively small size of VS, and positive results from *a priori* planned ROI analyses, additional analyses were run using a threshold of (p=.005, K=19), corrected for multiple comparisons based on a Monte Carlo simulation for the VS+VMPFC mask (949 total voxels) in order to maintain an appropriate balance of type I and type II error risk, given the exploratory nature of the whole-brain analysis (Lieberman & Cunningham, 2009).

Results

Effects of affirmation: Region of interest analysis

Main effects of affirmation. First, we examined whether activity in our *a priori* hypothesized ROIs associated with valuation (VS+VMPFC), self-related processing (MPFC+PCC), and emotion regulation (rACC+rVLPFC) were differentially activated for those in the affirmed versus control group as they reflected on value > everyday scenarios. Overall, affirmed participants displayed significantly greater activity in the valuation/reward network (M=.102) versus control participants (M=.012) when exposed to value versus everyday scenarios (t(57)=2.43, p=.018). Activity in the self-processing network while viewing value versus everyday scenarios was not significantly different for those in the affirmed versus control group, when averaging across temporal orientations (t(57)=.88, p=.382). Activity in the emotion regulation network also did not differ between affirmed and control groups when averaging across temporal orientations (t(57)=.62, p=.540). All ROI results are summarized in table 2.

For Peer Review

Affirmation and temporal orientation. Second, we tested whether affirmation effects differed by temporal orientation within our key ROIs (see Table 2). On average, affirmed participants displayed significantly greater activity in the valuation network (M=.133) than control participants (M=-.029) when viewing future oriented value scenarios versus future oriented everyday scenarios (t(57)=3.26, p=.002); the difference between responses to future and past oriented value scenarios was also significantly different between affirmed and control participants (t(57)=3.09, p=.003). Additionally, affirmed participants displayed significantly greater activity in the self-processing network (M=.100) than control participants (M=.032) when viewing future oriented value scenarios versus future oriented everyday scenarios (t(57)=2.50, p=.015); the difference between responses to future and past oriented value scenarios was also significantly different between affirmed and control participants (t(57)=3.48, p=.001). Participants in the affirmation and control conditions did not differ in their activity in the emotion regulation network when reflecting on future oriented value and everyday scenarios (t(57)=1.30, p=.200), however the difference between responses to future and past oriented value scenarios was significantly different between affirmed and control participants (t(57)=2.39, p=.02).

Next, we examined whether affirmation effects differed by past orientation within our key ROIs. No significant differences were observed between those in the affirmation versus control condition for activity in regions associated with valuation (t(57)=.34, p=.738), self-related processing (t(57)=-.97, p=.337) or emotion regulation (t(57)=1.77, p=.540) when reflecting on past oriented value versus everyday scenarios.

Finally, within the affirmation group paired samples t-tests were run to examine whether neural activity within our hypothesized ROIs were differently activated depending on temporal

For Peer Review

orientation (past versus future). Neural activity within the valuation network (VS+VMPFC) was significantly greater when viewing future oriented value scenarios (*M*=.108) compared to viewing past oriented value scenarios (*M*=.003), *t*(29)=3.83, *p*<.001. Similarly, neural activity with our self-processing network (MPFC+PCC) was also significantly greater when viewing future oriented value scenarios (*M*=.114) compared to viewing past oriented value scenarios (*M*=.044), *t*(29)=2.79, *p*=.009. Finally, neural activity within our emotion regulation network (rACC+rVLPFC) was not significantly different when viewing future oriented value scenarios (*M*=.056) compared to viewing past oriented value scenarios (*M*=.025), *t*(29)=1.65, *p*=.111.

Whole brain analysis

Following our hypothesis-driven ROI analyses, we ran a series of exploratory whole brain analyses that examined differences in neural activity between the affirmed and control groups for key contrasts of interest to explore regions outside of those covered by our ROI analyses. Results of the whole brain contrast of value > everyday scenarios did not yield significant results; future value > future everyday scenarios are reported in table 3, figure 2; past value > past everyday scenarios did not yield significant results; and future > past value scenarios are reported in table 4. Significant results from the whole brain analysis reinforce effects observed in the ROI analyses. We observed increased activity within VMPFC and VS when affirmed (relative to control) participants reflected on future-oriented (but not past-oriented) value scenarios highlighting the role of activity within the valuation system, particularly during prospection.

Neural activity during affirmation and subsequent behavior change

Validating the downstream effect of our affirmation manipulation on behavior change following exposure to health messages, participants in the self-affirmation condition showed

For Peer Review

steeper declines in their levels of sedentary behavior over time compared to control participants, p=.008 (Falk et al., 2015). Given that the effects of affirmation within the brain during the affirmation task were strongest in our hypothesized valuation and self-processing regions, we next examined whether this activity was related to target behavior change. Increased activity in the VS+VMPFC (figure 3) and MPFC+PCC ROIs during value > everyday scenarios were associated with decreased average post intervention sedentary behavior, controlling for age, gender, education, body mass index (BMI), and pre intervention sedentary behavior (β=-.26, t(33)=-2.27, p=.030; β=-.27, t(33)=-2.15, p=.039; respectively). Next, a follow up analysis was run examining temporal orientation differences. Results indicate that increased neural activity in the valuation network during future value versus future everyday scenarios was marginally associated (β=-.22, t(33)=-1.97, p=.057) and the self-processing network was significantly associated (β=-.25, t(33)=-2.39, p=.023) with decreased sedentary behavior following the affirmation intervention, controlling for age, gender, BMI, and pre intervention sedentary behavior. No significant results were found for past oriented scenarios, p>.05.

Finally, we tested the indirect relationship between group assignment (affirmation versus control) and changes in one's sedentary behavior (post – pre intervention) through neural activity in valuation and self-processing systems. Significant indirect effects were found for both the valuation and self-processing ROIs (average causal mediation effect (ACME); B=-.07, CI=[-.13, -.02], p=.01; B=-.03, CI=[-.07, -.00], p=.04; respectively), such that those in the affirmed condition displayed greater activity in valuation and self-processing networks relative to those in the control condition; in turn participants who displayed greater activation in the valuation and

For Peer Review

self-processing networks also displayed significantly greater decreases in sedentary behavior following the affirmation intervention, controlling for age, gender, years educated, and BMI.[2]

Discussion

Results from the current study provide initial evidence of neural processes associated with the act of self-affirmation. First, our hypotheses regarding the relationship between affirmation and neural reward pathways were supported. ROI analyses revealed that affirmed relative to control participants showed significantly greater activity in the hypothesized positive valuation regions (VS+VMPFC), and that this effect was driven by affirmations focusing on future rather than past experiences. In addition, increased activity in reward/valuation regions during self-affirmation was associated with decreases in sedentary behavior following the affirmation intervention. Furthermore, we observed a significant indirect effect, such that those in the affirmed condition displayed greater activity in the valuation network, which was associated with greater change in sedentary behavior following the affirmation intervention. Thus, our results are consistent with the hypothesis that systems associated with positive valuation play an important role in successful affirmation and are consistent with the broadened value account of why self-affirmation interventions succeed (Cohen et al., 2009; Cook et al., 2012; Koole et al., 1999; Sherman et al., 2013). The VS and VMPFC are brain regions that are most commonly associated with the expectation and receipt of positively valued or rewarding outcomes (Bartra et al., 2013). Importantly, this system encodes not only primary rewards (such as food) but also more abstract rewards (Bartra et al., 2013), of the type that are called to mind by personally meaningful values in self-affirmation.

[2] Note: In addition to examining changes in sedentary behavior using difference scores, a test of indirect effects was also run that examined post intervention sedentary behavior controlling for pre intervention sedentary behavior (ACME; B=-.08, CI=[-.15, -.03], p<.01). Results were consistent for future oriented statements (ACME; B=-.04, CI=[-.09, -.01], p=.01); past oriented statements were not significant (ACME; B=-.02, CI=[-.06, .01], p=.23).

For Peer Review

In addition, our findings suggest that positive affirmations may have especially strong effects within the reward system in conjunction with future orientation. This finding converges with prior studies demonstrating that increased activity in the VMPFC is associated with imagining positive rather than negative future events (D'Argembeau et al., 2008) and increases when anticipating future rewards (Benoit et al., 2011, 2014). This account is also consistent with a role of the reward system in guiding reinforcement learning and future behavioral decisions through computation of the "incentive value of a contemplated behavioral act" (Knutson & Cooper, 2005; McClure, Daw, & Read Montague, 2003).

Furthermore, although not directly addressed by our data, past research suggests that self-transcending values and goals may be particularly powerful. For example, affirmation of self-transcending values is more powerful in reducing behaviors associated with ego depletion than affirmation of self-enhancing values (Burson, Crocker, & Mischkowski, 2012). Our neural data provide a possible link between such behavioral results and research examining neural reward activity in response to prosocial (eudaimonic) versus selfish (hedonic) decisions, which finds that VS activity differentially predicts later mental health outcomes. More specifically, increased activity in the VS in response to potential prosocial rewards, relative to self-focused rewards is associated with later positive outcomes (Telzer, Fuligni, Lieberman, & Galván, 2014). These findings along with the findings from the current study support potential synergy between prospection and value affirmation in eliciting the types of reward response that can prime positive behavior.

Second, we found support for the hypothesis that future-oriented affirmations activated brain regions implicated in self-related processing. In particular, the MPFC is often implicated in reflecting on one's own preferences, motivations and in the process of self-insight (for a review,

For Peer Review

see; Lieberman, 2010). During future oriented affirmation, affirmed, relative to control, participants displayed significantly greater activity in our MPFC and PCC ROIs. Importantly, MPFC and PCC are consistently implicated in both self-related processing (Denny et al., 2012; Northoff et al., 2006) and imagining personally relevant future events (D'Argembeau et al., 2010), as well as remembering past events (for a review, see; Schacter, 2012). Furthermore, increased activity in the MPFC is associated with imagining positive rather than negative future events (D'Argembeau et al., 2008) and increases while anticipating future rewards (Benoit et al., 2011, 2014). Thus, the current data and recent meta-analytic evidence suggests that in addition to a strong role in self-related processing (Denny et al., 2012; Northoff et al., 2006), the MPFC is more active when thinking about future compared to past episodes (Benoit & Schacter, 2015) and when thinking about personal goals, future thinking, and mind wondering (Stawarczyk & D'Argembeau, 2015). Successful self-affirmation interventions bring together several of these components and our neural data suggest a new way in which these paths may mutually reinforce one another. In other words, we find novel evidence that a future frame may act synergistically with value-based self-affirmations to bolster a sense of self prior to threat exposure. This may occur by calling to mind desired future states or motivations, also consistent with the broadened value account of why self-affirmation interventions succeed.

Finally, the current study reports on the successful development of an fMRI-compatible self-affirmation task, which can be used to examine neural mechanisms associated with self-affirmation in other behavioral or theoretical contexts, and combined with other subsequent tasks of interest to affirmation researchers. One strength of this task is that all aspects of the task (including the within subjects control condition and instructions) are identical for affirmed and control participants, differing only in the importance of the focal value to participants. This rules

For Peer Review

out many confounds related to differing tasks. An additional strength is that our objective behavioral results indicate that the self-affirmation manipulation was successful in decreasing sedentary behavior in at-risk (sedentary) adults, which was mediated by activity in the valuation network during affirmation and by activity in both valuation and self-related processing systems during future oriented affirmations.

This adds to our understanding of affirmation from both basic science and applied perspectives. The current results 1) highlight novel pathways to affirmation through neural reward and self-processing pathways; and 2) suggest that these mechanisms may be reinforced or augmented by prospection. It is possible that future oriented affirmations may be more successful than past oriented affirmation, though between subjects follow up studies are needed to test this hypothesis. Finally, the creation of a scanner compatible affirmation task opens future research possibilities to explore the neural effects of affirmation in other contexts.

In addition to the primary strengths of the study addressed above, it should be noted that each of our primary ROIs serves functions that go beyond those hypothesized in the current investigation and thus should be taken as one of several possibilities (Poldrack, 2006). However, in the current study the use of *a priori* hypothesized and theoretically driven ROIs helps reduce problems with reverse inference. Furthermore, there are confines associated with the scanning environment, such that we cannot know the specific scenarios envisioned by each participant in response to our prompts at the time of affirmation exposure, we can only examine neural processing that takes place during that time. Therefore, it is likely that variability in how important the "lowest" ranked value was to participants existed, which may have allowed for affirming benefit to some of those in the control condition, resulting in a conservative test of our hypotheses. In addition, self-affirmation interventions are often confounded with value and

For Peer Review

content making it difficult to distinguish which aspects of the intervention are driving results. Future neuroimaging studies should attempt to untangle these differences in order to better understand the underlying mechanisms associated with self-affirmation interventions.

Conclusion

The current results demonstrate that activity in hypothesized reward/valuation regions (VS+VMPFC (Bartra et al., 2013) are primary pathways associated with self-affirmation. Furthermore, regions associated with self-related processing (MPFC+PCC (Northoff et al., 2006; Denny et al., 2012) and prospection (D'Argembeau et al., 2010, 2008) are associated with self-affirmations that are future oriented. These neural correlates of self-affirmation were further associated with objectively measured behavior change, suggesting the external validity of the affirmation task. Taken together, our results highlight ways in which brain systems implicated in positive valuation and self-related processing may be reinforced by prospection and suggest novel insight into the balance of processes supporting affirmation. These results also introduce a task for understanding the underlying mechanisms associated with self-affirmation and hence provide a tool for future studies to examine effects of self-affirmation interventions across a wide range of potential applications and outcomes.

For Peer Review

References

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of {BOLD} fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*(0), 412 – 427.

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *The Journal of Neuroscience*, *31*(18), 6771–6779.

Benoit, R. G., & Schacter, D. L. (2015). Specifying the core network supporting episodic simulation and episodic memory by activation likelihood estimation. *Neuropsychologia*, *75*, 450–457.

Benoit, R. G., Szpunar, K. K., & Schacter, D. L. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences*, *111*(46), 16550–16555.

Burson, A., Crocker, J., & Mischkowski, D. (2012). Two types of value-affirmation implications for self-control following social exclusion. *Social Psychological and Personality Science*, *3*(4), 510–516.

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), 400–403.

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, *65*, 333–371.

For Peer Review

Cook, J. E., Purdie-Vaughns, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: protective benefits of values affirmation on identity development. *Journal of Personality and Social Psychology*, *102*(3), 479.

D'Argembeau, A., Stawarczyk, D., Majerus, S., Collette, F., Van der Linden, M., Feyers, D., … Salmon, E. (2010). The neural basis of personal goal processing when envisioning future events. *Journal of Cognitive Neuroscience*, *22*(8), 1701–1713.

D'Argembeau, A., Xue, G., Lu, Z.-L., Van der Linden, M., & Bechara, A. (2008). Neural correlates of envisioning emotional events in the near and far future. *NeuroImage*, *40*(1), 398–407.

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*(8), 1742–1752.

Falk, E. B., Cascio, C. N., O'Donnell, M. B., Shumaker, K., Tinney, F., Kang, Y., … Strecher, V. (2015). Self affimration alters the brain's response to health messages and subsequent behavior change. *Proceedings of the National Academy of Sciences*.

Klein, S., Robertson, T., & Delton, A. (2010). Facing the future: Memory as an evolved system for planning future acts. *Memory & Cognition*, *38*(1), 13–22. http://doi.org/10.3758/MC.38.1.13

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*(4), 411–417.

For Peer Review

Koole, S. L., Smeets, K., Van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology*, *77*(1), 111.

Lieberman, M. D. (2010). Social Cognitive Neuroscience. *Handbook of Social Psychology*.

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), 423–428.

McClure, S. M., Daw, N. D., & Read Montague, P. (2003). A computational substrate for incentive salience. *Trends in Neurosciences*, *26*(8), 423–428.

McQueen, A., & Klein, W. M. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, *5*(4), 289–354.

Napper, L., Harris, P. R., & Epton, T. (2009). Developing and testing a self-affirmation manipulation. *Self and Identity*, *8*(1), 45–62.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, *31*(1), 440–457.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*(8), 603.

Sherman, D. K. (2013). Self□Affirmation: Understanding the Effects. *Social and Personality Psychology Compass*, *7*(11), 834–845.

For Peer Review

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, *38*, 183.

Sherman, D. K., Cohen, G. L., Nelson, L. D., Nussbaum, A. D., Bunyan, D. P., & Garcia, J. (2009). Affirmed yet unaware: Exploring the role of awareness in the process of self-affirmation. *Journal of Personality and Social Psychology*, *97*(5), 745.

Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., … Cohen, G. L. (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.

Stawarczyk, D., & D'Argembeau, A. (2015). Neural correlates of personal goal processing during episodic future thinking and mind□wandering: An ALE meta□analysis. *Human Brain Mapping*.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, *21*(2), 261–302.

Telzer, E. H., Fuligni, A. J., Lieberman, M. D., & Galván, A. (2014). Neural sensitivity to eudaimonic and hedonic rewards differentially predict adolescent depressive symptoms over time. *Proceedings of the National Academy of Sciences*, *111*(18), 6600–6605.

Ward, B. D. (2000). Simultaneous inference for fMRI data. *AFNI 3dDeconvolve Documentation, Medical College of Wisconsin*.

For Peer Review

Tables

Table 1. Value rankings.

	Overall Frequency		Group Frequency	
Value	**Highest Value**	**Lowest Value**	**Affirm**	**Control**
Money	2	4	2	3
Creativity	3	4	1	3
Independence	8	1	3	0
Politics	1	34	1	16
Friends_Family	33	0	14	0
Religion	8	20	7	5
Humor	9	1	4	1
Spontaneity	3	3	1	0

Table 2. ROI analysis summary for the contrasts value > control, value future > control future, and past value > past control.

ROI (value > control)	Affirmed Mean	Control Mean	t(57)	p
VS & VMPFC	0.102	0.012	2.43	0.018
MPFC & PCC	0.12	0.094	0.87	0.387
rACC & rVLPFC	0.035	0.018	0.62	0.54

ROI (future value > future control)	Affirmed Mean	Control Mean	t(57)	p
VS & VMPFC	0.133	-0.029	3.26	0.002
MPFC & PCC	0.147	0.048	2.37	0.021
rACC & rVLPFC	0.04	-0.01	1.3	0.2

ROI (past value > past control)	Affirmed Mean	Control Mean	t(57)	p
VS & VMPFC	0.071	0.052	0.34	0.738
MPFC & PCC	0.092	0.14	-0.94	0.353
rACC & rVLPFC	0.03	0.046	-0.45	0.657

For Peer Review

Table 3. Whole brain analysis comparing the contrast (future value > future control scenarios) for the affirmed group subtracted from the control group (p=.005, K=35). *Results based on cluster correction for multiple comparisons using the VS+VMPFC mask (p=.005, K=19).

Region	*x*	*y*	*z*	*k*	*t*
VMPFC (bilateral)	8	56	-11	172	4.19
Posterior Cingulate (left)	-9	-60	4	95	3.98
Thalamus (right)	15	-23	16	37	3.76
Supplimentary Motor Area (left)	-30	8	52	144	3.5
Supplimentary Motor Area (right)	29	15	40	99	3.77
Calcarine (bilateral)	1	-102	-8	39	3.87
Brainstem (bilateral)	-2	-33	-17	46	3.57
Cerebelum (right)	32	-47	-50	44	3.64
*VS (left)	-13	22	1	20	3.52

Table 4. Whole brain analysis comparing the contrast (future value > past value scenarios) for the affirmed > control group (p=.005, K=35).

Region	*x*	*y*	*z*	*k*	*t*
VMPFC (bilateral)	-2	43	-11	444	3.88
Precuneus (bilateral)	-2	-60	67	46	3.38
Precunues/PCC (bilateral)	1	-54	4	1355	5.57
VS (bilateral)	10	0	12		
DLPFC (right)	29	15	43	127	4.60
DLMPFC (left)	-23	26	40	248	4.60
Occipital (right)	39	-81	34	148	4.87
Occipital (left)	-37	-81	37	349	4.83
Cerebelum (right)	11	-50	-47	85	4.28
Cerebelum (right)	46	-71	-38	74	4.21

For Peer Review

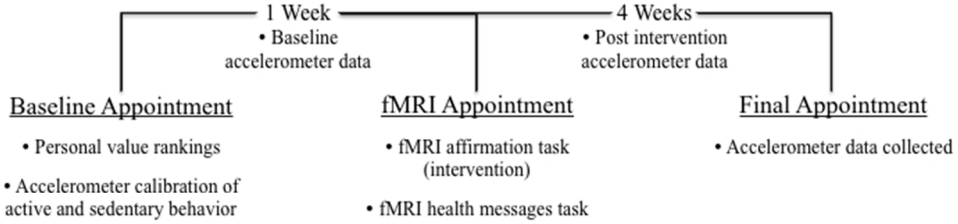
Figures

Figure 1. Study design

Figure 2. Whole brain analysis comparing the contrast (future value > future control scenarios) for the affirmed group > control group.

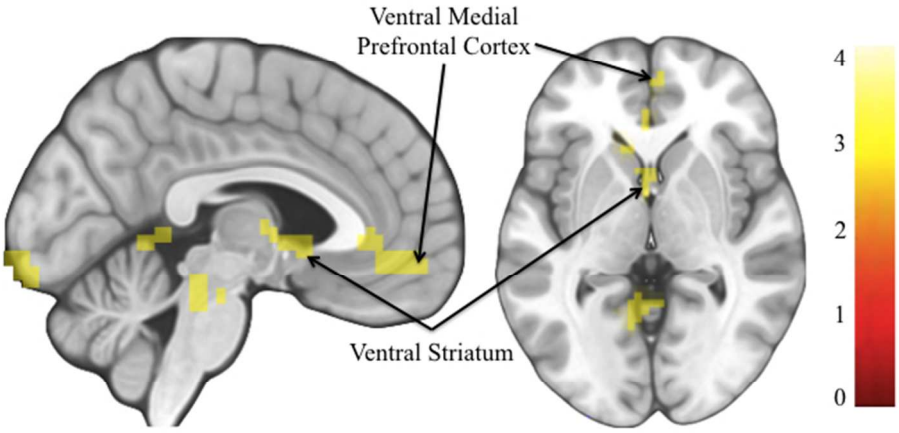
Figure 3. Scatter plot of the residualized percent signal change activity in the valuation network (VS+VMPFC) ROI from the contrast (value > control scenarios) predicting post intervention sedentary behavior, controlling for age, gender, education, BMI, and pre intervention sedentary behavior.

For Peer Review



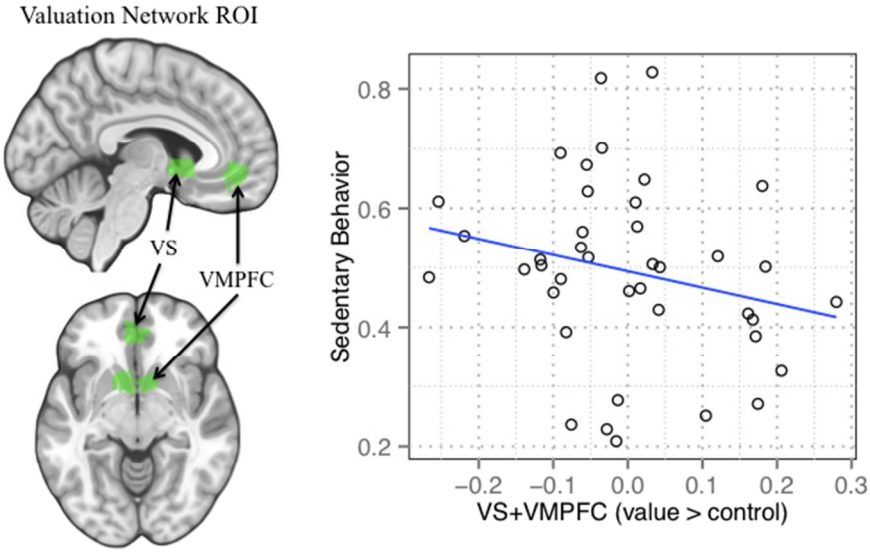
254x190mm (72 x 72 DPI)

review



254x190mm (72 x 72 DPI)

Review



254x190mm (72 x 72 DPI)

Review

Self-affirmation activates brain systems associated with self-related processing and reward and is reinforced by future orientation

Christopher N. Cascio[1], Matthew Brook O'Donnell[1], Francis J. Tinney, Jr.[2], Matthew D. Lieberman[3], Shelley E. Taylor[3], Victor J. Strecher[2], & Emily B. Falk[1]

University of Pennsylvania[1], University of Michigan[2], University of California, Los Angeles[3]

Supplementary Material

Participants

Participants were randomly assigned between subjects to either self-affirmation (n=33; mean age=33.33; *SD*=12.43) or control conditions (n=28; mean age=31.63; *SD*=13.65). One person failed to complete the post-scan portion of the study, accelerometer data were not obtained for an additional 12 participants due to equipment failure and damage, and 3 participants were removed from the study due to excess head movement or technical difficulties with the scanner during the health messages task. Therefore, neural results focusing on the neural correlates of self-affirmation are based on a final sample of 61; results validating the effects of affirmation using accelerometer data are based on a subsample of 45 participants (self-affirmation (n=22; mean age=33.27; *SD*=13.72), control condition (n=23; mean age=30.09; *SD*=13.07)). No significant differences in age were found between the control and affirmation participants in the full or subsample, ($t(58)=-.051$, $p=.615$; $t(43)=-.798$, $p=.429$; respectively). The results of this investigation come from the same sample and task session as those from Falk et al., 2015, however, the neural processes associated with the actual affirmation task have not been previously examined.

Preprocessing

For Peer Review

Functional data were pre-processed and analyzed using Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK; please see supplemental materials for details of preprocessing stream). To allow for the stabilization of the BOLD signal, the first four volumes (eight seconds) of each run were discarded prior to analysis. Functional images were despiked using the 3dDespike program as implemented in the AFNI toolbox. Next, data were corrected for differences in the time of slice acquisition using sinc interpolation; the first slice served as the reference slice. Data were then spatially realigned to the first functional image. We then co-registered the functional and structural images using a two-stage procedure. First, in-plane T1 images were registered to the mean functional image. Next, high-resolution T1 images were registered to the in-plane image. Following coregistration the high-resolution T1 images were segmented into white and gray matter allowing the skull to be removed. Structural and functional images were then normalized to the skull-stripped MNI template provided by FSL ("MNI152_T1_1mm_brain.nii"). Finally, functional images were smoothed using a Gaussian kernel (8mm FWHM).

Region of interest (ROI) definitions

Anatomical ROIs were constructed in the Wake Forest University Pickatlas toolbox within SPM (Maldjian, Laurienti, Kraft, & Burdette, 2003) by combining definitions from the Automated Anatomical Labeling Atlas (Tzourio-Mazoyer et al., 2002), Talairach Daemon database atlas (TD atlas; (Lancaster et al., 2000)), and Brodmann areas, as well as functional meta-analytic findings (Bartra et al., 2013). MarsBar was used to convert these images to ROIs (Brett, Anton, Valabregue, & Poline, 2002). The valuation network ROI was constructed using the functional VS and VMPFC activations found in figure 9 of the Bartra et al. (2013) meta-

For Peer Review

analysis, examining common regions activated across tasks that focus on positive valuation (for additional details, see; Bartra et al., 2013; Figure 9).

The MPFC and PCC, regions found to be more active during self reflection versus reflecting on others and control tasks (for a review, see; (Lieberman, 2010)) were used to create our self-processing network. The network was constructed by taking the union of the MPFC (all voxels within Brodmann area 10 restricted medially by intersecting a box-shaped mask that extends from x=-20 to 20, y=45 to 70, z=-10 to 30, in order to restrict BA10 medially because of its role in self-related processing compared to more dorsal aspects which are likely involved to greater extents in social cognition; see Lieberman, 2010 for a review) and PCC (the union of the left and right posterior cingulate defined by the TD atlas; (Lancaster et al., 2000)).

Finally, the VLPFC and rACC, regions implicated in regulation of emotion and facilitating difficult choices (Marsh, 2007; Oschner et al., 2004; Wager et al., 2008) were combined to create our emotion regulation network. , the rACC was constructed as the intersection of the anterior cingulate and a box that extends from (x=-25 to 25, y=32 to 45, z=-15 to 32 in order to avoid overlap with the subgenual cingulate and dACC) and the rVLPFC was constructed by taking the union of Brodmann areas 44, 45, and 47 in the right hemisphere.

Figure S1. Regions of interest (ROIs) consist of the positive valuation network (VS+VMPFC), self-related processing network (MPFC+PCC), and emotional regulation network (rACC+rVLPFC).

ROI analysis

We examined independent t-tests on neural activity for those in the affirmed condition versus those in the control condition during key task comparisons to examine main effects of

For Peer Review

affirmation, as well as interactions with temporal orientation. In addition, to validate the effects of affirmation on behavior in relation to the neural effects observed, linear regression models were run to examine the relationship between neural activity in regions of interest found to be significantly associated with affirmation (value > everyday scenarios for affirmed > control participants) and sedentary behavior change over the month following the intervention. Finally, we used the mediation package in R in order to test indirect effects of affirmation on sedentary behavior through brain activity.

Results

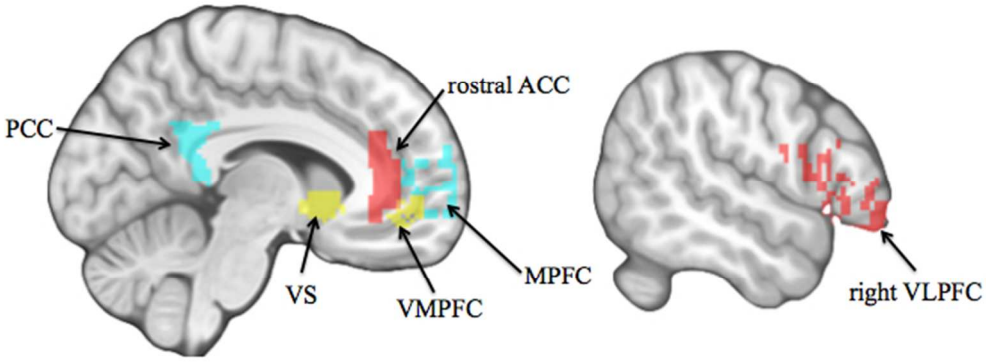
Zero order correlations. Activity in the valuation network (VS+VMPFC) in the future scenario condition (value > control) significantly correlates with changes in sedentary behavior (post – pre), r=-.29, p=.05. Thus, increased activity in the valuation network is associated with greater decreases in sedentary behavior following the intervention. This result was not significant in the past scenario condition, r=.10, p=.51. In addition, activity in the self-processing network (MPFC+PCC) in the future scenario condition (value > control) significantly correlates with changes in sedentary behavior (post – pre), r=-.30, p=.046. Thus, increased activity in the self-processing network is associated with greater decreases in sedentary behavior following the intervention. This result was not significant in the past scenario condition, r=.22, p=.14.

Self-affirmation intervention values versus content. Differences in effects of self-affirmation may be inherently tied to the types of values that humans prioritize (i.e., close others), therefore, additional analyses were run in order to better understand whether the results were driven by the general value one placed on the scenario or by the specific content. Within the current study no significant differences existed within our ROIs (positive valuation, self-related processing, and emotion regulation) for those affirmed with family and friends versus the

For Peer Review

other possible values (positive valuation network: $t(29)=-.46$, $p=.649$; self-related processing network: $t(29)=-.54$, $p=.592$; emotion regulation network: $t(29)=-.01$, $p=.992$; respectively).

For Peer Review



234x91mm (72 x 72 DPI)

r Peer Review