

# Salience network engagement with the detection of morally laden information

Gunes Sevinc,<sup>1,2,3</sup> Hakan Gurvit,<sup>3,4</sup> and R. Nathan Spreng<sup>1</sup>

<sup>1</sup>Laboratory of Brain and Cognition, Department of Human Development, Human Neuroscience Institute, Cornell University, Ithaca, NY, USA, <sup>2</sup>Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, <sup>3</sup>Department of Neurosciences, Institute for Medical Research, and <sup>4</sup>Department of Neurology, Behavioral and Movement Disorders Unit, Faculty of Medicine, Istanbul University, Istanbul, Turkey

Correspondence should be addressed to Gunes Sevinc, Division of Psychiatric Neuroscience, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, 120 2nd Ave, Charlestown, MA 02129, USA. E-mail: gsevinc@mgh.harvard.edu, gkayaci@gmail.com.

## Abstract

Moral cognition is associated with activation of the default network, regions implicated in mentalizing about one's own actions or the intentions of others. Yet little is known about the initial detection of moral information. We examined the neural correlates of moral processing during a narrative completion task, which included an implicit moral salience manipulation. During fMRI scanning, participants read a brief vignette and selected the most semantically congruent sentence from two options to complete the narrative. The options were immoral, moral or neutral statements. RT was fastest for the selection of neutral statements and slowest for immoral statements. Neuroimaging analyses revealed that responses involving morally laden content engaged default and executive control network brain regions including medial and rostral prefrontal cortex, and core regions of the salience network, including anterior insula and dorsal anterior cingulate. Immoral vs moral conditions additionally engaged the salience network. These results implicate the salience network in the detection of moral information, which may modulate downstream default and frontal control network interactions in the service of complex moral reasoning and decision-making processes. These findings suggest that moral cognition involves both bottom-up and top-down attentional processes, mediated by discrete large-scale brain networks and their interactions.

**Key words:** morality; moral cognition; salience network; default network; implicit processing

## Introduction

The ability to make moral judgments lies at the centre of human social endeavors. This ability emerges from the interaction of affective and cognitive processes including mentalizing, retrieval of autobiographical memory and mental simulation (Greene and Haidt, 2002). The degree to which each of these processes is involved in those judgments depends on contextual factors (Monin *et al.*, 2007). Within the context of shocking moral violations, moral judgments primarily involve

rapid, affect-driven reactions (Moll *et al.*, 2002a). In contrast, moral dilemmas involving choice or decision-making, require mental simulation to guide deliberative reasoning (Greene *et al.*, 2001, 2004). While discrete cognitive abilities are associated with specific moral contexts, an essential, shared component of moral judgment involves the capacity to detect morally salient content within a given social context.

The neurobiological mechanisms underlying moral behavior have been investigated using structural and functional

Received: 22 November 2016; Revised: 13 February 2017; Accepted: 6 March 2017

© The Author (2017). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

neuroimaging techniques and there is growing agreement with respect to the brain regions involved in moral cognition. These include prefrontal cortex (PFC), particularly its medial aspect, regions of temporal cortex such as posterior superior temporal sulcus (pSTS), temporoparietal junction (TPJ), temporal polar cortices and medial parietal regions including posterior cingulate cortex (PCC) (Heekeren et al., 2003; Moll et al., 2002a,b). Subcortical and limbic structures including ventral striatum, hypothalamus, amygdala and basal forebrain have also been implicated (Moll et al., 2005). Cortical areas involved in moral cognition closely overlap regions of the default network (Harrison et al., 2008; Reniers et al., 2012). The default network supports mental simulation (Spreng and Grady, 2010) and is active when individuals are engaged in internally focused tasks including autobiographical memory retrieval, envisioning the future and discerning the thoughts or intentions of others (Buckner et al., 2008; Andrews-Hanna et al., 2014). Each of these processes has been associated with moral cognition, implicating the default network in moral judgments.

In a recent meta-analysis examining brain activity during moral judgments, we found support for this prediction. Tasks involving active moral judgments, requiring mental simulation to infer the intentions of others, engaged regions of the default network. In contrast, tasks requiring the passive viewing of moral stimuli were associated with activation of emotional and visual association cortices, including the amygdala and ventral temporal brain regions (Sevinc and Spreng, 2014). While the default network is involved in more associative aspects of moral cognition (e.g. mapping actions to the intentions of others), and visual and emotional processing regions are associated with perception of morally-based stimuli, little is known about the neural systems involved in the detection of moral content within the myriad social interactions we encounter in our daily lives.

The ventral-attention, or salience network, is associated with detecting behaviorally-relevant stimuli in the environment (Seeley et al., 2007; Uddin, 2015). Paralimbic components of the salience network, including the dorsal anterior cingulate (dACC) and orbital frontoinsula, have been associated with interoceptive, autonomic processing (Mesulam, 1998; Damasio, 1999; Craig, 2002; Critchley, 2005; Seeley et al., 2007). Limbic aspects of the salience network are associated with emotion, homeostatic regulation, and reward (Ongur and Price, 2000; Menon and Levitin, 2005; Seeley et al., 2007). This network is thought to be important in integrating sensory data with autonomic, visceral and somatic markers, leading to the attribution of personal salience to a stimulus. The salience network responds to behaviorally-relevant events, and activity in these regions has been associated with modulation of activity and interactivity in other brain networks including the default and frontoparietal control networks (or, executive control networks; Menon and Uddin, 2010; Spreng et al., 2013).

This modulatory or switching role of salience network among different neural networks is supported by clinical evidence. Behavioral variant frontotemporal dementia (bvFTD), which has a clinical feature of impaired moral reasoning, is a particular neurodegeneration, in which the earliest involvement is said to occur in a core salience network structure, that is frontoinsula (Seeley, 2010). It has recently been shown that patients with bvFTD have reduced recruitment of the default network compared with healthy control subjects when deliberating about personal moral dilemmas, that might reflect an impaired modulatory role for the salience network in regulating default network activation (Chiong et al., 2013). In another study,

cognitively normal Parkinson's disease subjects, when compared to healthy controls were shown to display decreased salience-executive control network, but increased default-executive control network coupling; moreover, salience network coupling with striatum decreased with increasing disease severity, probably indicating the critical switching role of the salience network between different brain networks (Putcha et al., 2015).

Based on the involvement of the salience network in the detection of behaviorally relevant or salient stimuli, and its purported role in dynamic switching between internally and externally directed cognition (Uddin, 2015), we propose that the salience network may play an important role in the initial detection of moral content, and the subsequent configuration of downstream brain regions to perceive and process this information in the service of goal-directed cognition. In this model, moral information, which possesses high survival value for the organism, may be detected automatically and tagged as salient by regions of the salience network. Initial detection initiates more context-specific and goal-directed processing of the moral content, involving default and frontal-control brain regions, and their interactions, in the service of morally-based judgment and decision-making.

To test this model of moral cognition, here we investigate how the detection of morally salient information is associated with brain network activation. Considering the variety of cognitive processes involved in moral judgments, we adopt the term moral processing to refer to the mental processes associated with moral judgments including its detection. To examine the role of distributed brain networks in detecting moral salience in real world contexts, we use daily life scenarios to provide contextual information, and ask participants to complete each scenario with logical, semantically-related statements that vary in moral content (i.e. moral, immoral or neutral). Critically, moral processing is incidental to the task demands, allowing us to investigate automatic, or implicit, detection of morally salient information. Further, to explore the temporal unfolding of neural activity associated with moral judgment, we examine brain activity over the complete post-decision epoch. We predict greater engagement of brain regions associated with salience detection for judgments involving the processing of moral content, with later engagement of default and control network brain regions, involved in more associative and integrative processes.

## Materials and methods

### Participants

Twenty-two young adults (10 females and 12 males; mean age and SD  $22.1 \pm 3.14$ ) took part in this study. All participants were right-handed, had normal or corrected vision, and were screened for history of psychiatric or neurological problems. They were compensated for their time and all gave informed consent. The study was approved by Cornell University, Institutional Review Board. Two of the participants were excluded from analyses due to high number of incorrect responses (30 and 26% of all the scenarios, respectively), more than three standard deviations above the mean.

### fMRI task design

During a semantic association task, participants read a brief vignette, and then completed the vignette by selecting the semantically-related sentence from two options. Correct and

incorrect sentences involved neutral, moral or immoral content, although this manipulation was not part of the task and participants were not made aware of the sentence categories. Following the fixation screen, a scenario such as the following appears on the screen: 'I was on my way to work. I got on my usual train. The train was very crowded that day, but I got the last free seat. A few people boarded at the next stop. One was a man on crutches, who looked like he was in pain. So,'. When participants finish reading this scenario, they proceed to the next screen by a button press. This next screen involves the two-sentence options: '(a) I gave him my seat, (b) I left the building' and participant's task is to chose the most semantically congruent sentence. For each category, the semantically congruent option includes behaviors corresponding to that category (either moral, immoral or neutral) and the incongruent sentence options include sentences pertaining to another of all the three categories and is counterbalanced across trials and runs.

The tasks were intermixed with fixation intervals randomized from 1 to 6 s. Following each fixation interval, a scenario was presented and participants were asked to press (1) with their index finger to move on to the next screen with the two sentence options. Participants chose the correct response by pressing the button (1) with their index, or pressing the button (2) with their middle finger. Figure 1 reflects the sequence of stimuli participants saw in the scanner and includes an exemplar from each semantic association task condition. The experiment included 162 scenarios (54 to be completed with moral, 54 with immoral and 54 with neutral sentences) and comprised 6 runs. The vignettes were prepared by a team of three people and a vignette was included only when the team agreed about its moral valence. Moral and immoral scenarios are devised from examples in daily life and by peers of the participant. All scenarios and response options are provided in Supplemental Material.

At the beginning of each run, a fixation screen was presented for 20s to account for scanner calibration. Scenarios were presented randomly and the presentation order of the

correct option was counterbalanced. Both the semantically-related correct sentence, and the incorrect sentence incorporated immoral, moral or neutral statements to account for the processing related to passive exposure to moral content. As moral processing was incidental to the semantic categorization task demands, brain response associated with processing of moral salience during decision-making could be assessed.

### Imaging methods

Imaging data were acquired on a 3T GE scanner at the Cornell MRI Facility. High-resolution three-dimensional T1-weighted images were acquired as anatomical scans [repetition time (TR) = 2530 ms; echo time (TE) = 3.4 ms; inversion time (TI) = 1100 ms; flip angle (FA) = 7°; bandwidth = 195 Hz/pixel; 1.0 mm isotropic voxels, 176 slices]. Functional data were acquired using a multi-echo EPI sequence [repetition time (TR) = 2500 ms; echo time (TE) = 14, 30, and 47 ms; flip angle (FA) = 77°; matrix size = 72 x 72; field of view (FOV) = 210 mm; 38 axial slices; slice thickness 3.0 mm]. Head motion was restricted using two padded clamps. Earplugs were provided to attenuate scanner noise and participants held a button box in their right hand. Visual stimuli were projected onto a screen positioned at the head of the magnet bore, which was reflected in a mirror on top of the head coil.

### fMRI preprocessing and analysis

Preprocessing was performed with multi-echo independent components analysis (Kundu *et al.*, 2012) and then data were smoothed using a Gaussian kernel with an isotropic full width at half maximum (FWHM) of 6 mm. Preprocessed data were then submitted to Partial Least Squares (PLS; McIntosh *et al.*, 1996; Krishnan *et al.*, 2011). PLS is a data driven multivariate statistical method, which reveals activity across the entire brain that correlates with the study design. As opposed to the investigation of task related co-activation, this method reveals functionally connected regions, voxels that are synchronized in time which co-vary with the design, and is particularly suited for

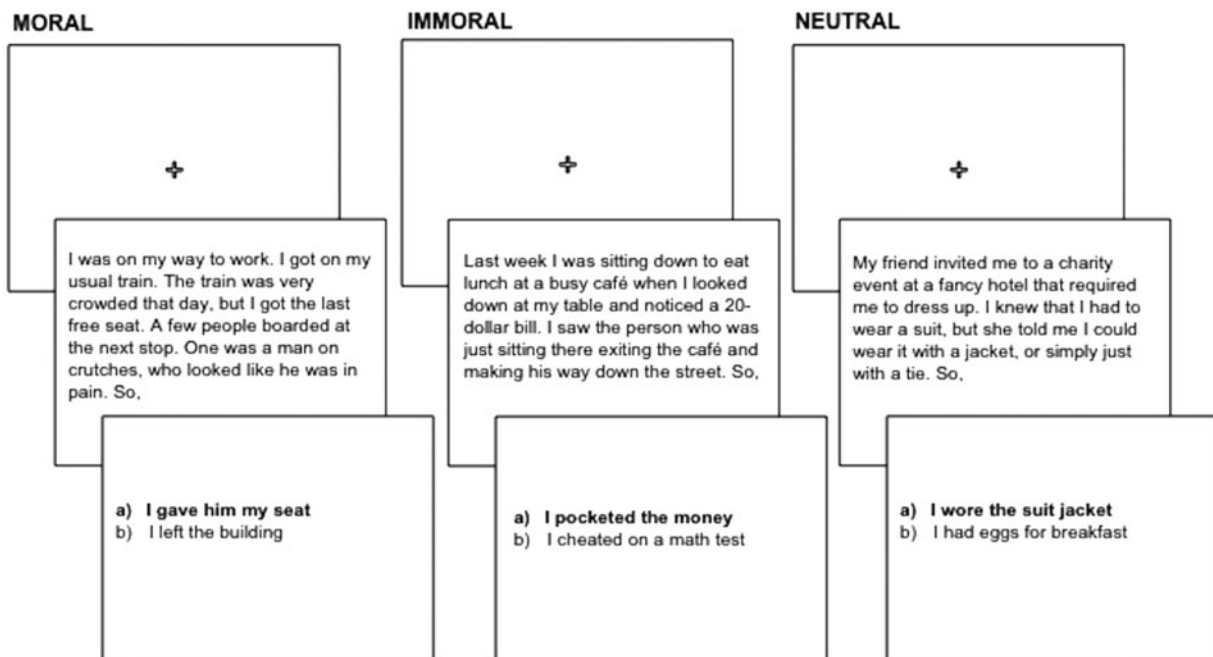


Fig. 1. The sequence of stimuli during the semantic association task with an exemplar from each task condition.

identifying the regions that form a functional network differentially engaged across the experiment.

The covariance between design and brain voxels is decomposed into latent variables (LVs). For each latent variable, each voxel is given a singular value weight that is proportional to the covariance of activity with the task contrast. This value is referred to as salience and the multiplication of salience by the BOLD signal at each voxel, summed across all voxels provides a brain score for each participant on a given latent variable. Significance of LVs is calculated using 500 permutation tests using resample without replacement. The reliability of each voxel's contribution to a LV is determined by bootstrapping. In order to estimate the salience of each voxel, the data was resampled with replacement 100 times. A bootstrap ratio (BSR), the ratio of each salience value to its standard error is then calculated.

To capture the temporal unfolding of brain activity from initial presentation of the options, the analysis window included 5 TRs, beginning with the onset of the choice screen. Task contrast analyses were conducted using non-rotated task PLS with moral + immoral > neutral and moral > immoral task contrasts. Peak voxels with a BSR greater than  $\pm 2.81$  were considered reliable and approximate a probability of  $P < 0.005$ . Clusters containing at least 10 reliable voxels were extracted, and the minimum distance between each cluster peak was set at 20 millimeters.

Network localization was done overlaying demarcations of a large-scale parcellation of the cerebral cortex (Yeo et al., 2011) on the contrast images using the Connectome Workbench visualization software (Marcus et al., 2011; see Figures 1 and 2, Supplementary Material). Network membership of the peak coordinates of each contrasts are presented in Tables 1 and 2.

## Results

### Behavioral results

Percentage of incorrect responses for moral, immoral and neutral conditions were 1.1%, 1.5% and 0.7%, respectively. Reaction times (RT) were calculated such that for each subject, incorrect

responses were replaced by the condition mean, outliers were Winsorised, and subject RTs were converted to z-scores. Mean RT are reported in Figure 2. A repeated measures analysis of variance revealed that RT was significantly different between the conditions ( $F(2,57) = 44.3, P < 0.001$ ). Bonferroni corrected post hoc comparisons ( $\alpha = 0.05$ ) revealed significant differences in RT between all three conditions, with neutral < moral < immoral (Figure 2).

### Neuroimaging results

A significant pattern of brain activity differentiated moral-content (moral + immoral) vs neutral selections ( $P = 0.002$ ) and involved lateral (rostral and dorsal aspects) as well as medial prefrontal cortex (MPFC), dACC, PCC, anterior insula (aINS), fusiform gyrus, right TPJ and inferior parietal lobule. This pattern was associated with moral-content selections and highly overlapped executive control, salience and default networks (Table 1; Figure 3).

A direct contrast of brain activation for immoral vs moral sentence completions revealed a significant pattern of differentiated brain activity ( $P = 0.012$ ). In the immoral > moral sentence selections, regions included lateral and rostral PFC, aINS, left TPJ and dACC, regions closely overlapping salience and executive control networks. The reverse contrast (moral > immoral) revealed activity in retrosplenial cortex, PCC, superior parietal lobule and other regions (Table 2; Figure 4).

Examining the temporal unfolding of brain activity revealed a different temporal pattern for each contrast. Brain scores for the moral + immoral vs neutral contrast peaked at lag 3 (7.5 s; Figure 3B) and was sustained over 1 TR before declining. Additionally, default and salience network brain regions were coactive early in the epoch, followed by increasing activation in executive control brain regions (e.g. lateral PFC, Figure 5). For the moral vs immoral contrast, early engagement of posterior default network regions (PCC), MPFC and superior parietal lobule were prominent for the moral selection condition. In contrast, for the immoral condition, there was a clear temporal unfolding of brain activity across the epoch (Figure 6). While

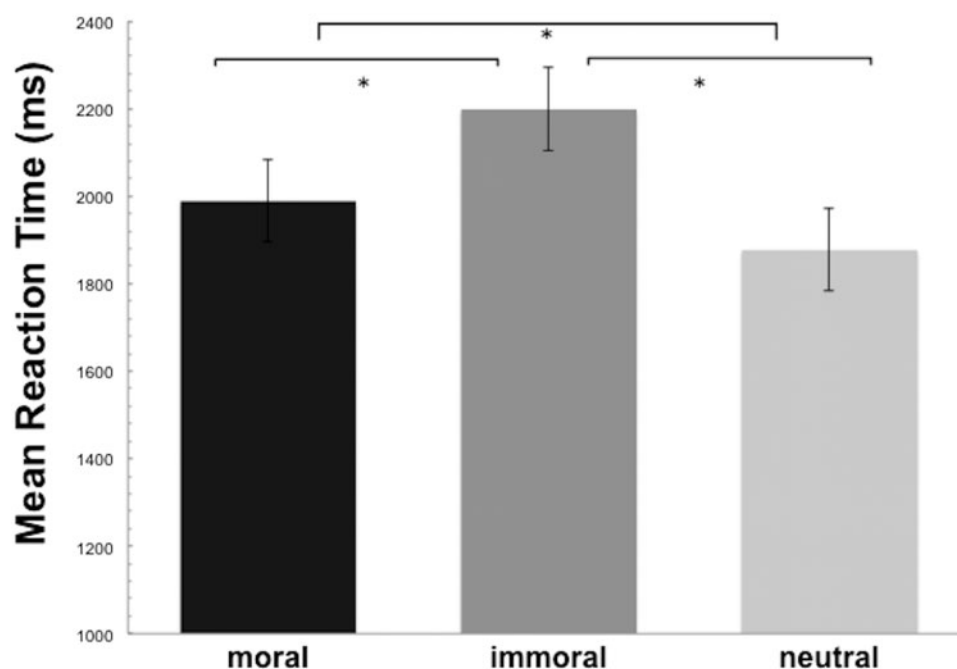


Fig. 2. Mean reaction times for neutral, moral and immoral task conditions. Reaction time differences were statistically different: \* $P < 0.01$ .

Table 1. Activation peaks for the Moral &amp; Immoral &gt; Neutral contrast

Region	MNI coordinates			BSR	Network
	X	Y	Z		
<b>Moral &amp; Immoral &gt; Neutral</b>					
Dorsal anterior cingulate	0	38	22	9.09	DN
Medial prefrontal cortex	0	60	18	8.80	DN
Middle frontal gyrus	-18	60	12	8.47	DN
Posterior middle frontal gyrus	-36	18	36	8.11	DN/FP
Head of caudate	12	10	12	7.70	Subcortical
Lateral occipital cortex	38	-70	2	7.51	Visual
Inferior frontal gyrus	46	16	4	7.38	SN
Head of caudate	-12	14	10	7.30	Subcortical
Ventromedial prefrontal cortex	6	36	-6	7.27	DN
Posterior superior temporal sulcus	-42	-32	-4	6.52	DN
Putamen	-24	6	-10	6.40	Subcortical
Inferior frontal gyrus	-40	34	-10	6.21	DN
Posterior cingulate cortex	0	-44	36	6.15	DN
Amygdala	-26	-6	-18	6.10	Subcortical
Posterior hippocampus	-28	-30	-8	5.94	DN
Superior temporal sulcus	60	-22	-10	5.81	DN
Inferior parietal lobule	-46	-52	24	5.69	DN
Temporal pole	52	-10	-28	5.19	DN
Anterior insula	36	18	-10	5.05	SN
Intraparietal sulcus	36	-50	48	4.99	DA
Middle cingulate cortex	2	-12	34	4.72	DN
Anterior insula	-34	22	-6	4.69	DN/FP
Middle frontal gyrus	40	16	52	4.56	FP
Temporal pole	48	8	-26	4.43	DN
<b>Neutral &gt; Moral &amp; Immoral</b>					
Medial occipital cortex	-8	-94	6	-3.91	Visual

Note: DN, default network; SN, salience network; FP, frontoparietal network; DA, dorsal attention network.

Network localization was done overlaying demarcations of a large-scale parcellation of the cerebral cortex (Yeo et al., 2011) on the contrast images using the Connectome Workbench visualization software (Marcus et al., 2011, see Figure 1, Supplementary Material).

Table 2. Activation peaks for the Immoral &gt; Moral contrast

Region	MNI coordinates			BSR	Network
	X	Y	Z		
<b>Immoral &gt; Moral</b>					
Dorsal anterior cingulate	-4	50	28	5.90	DN/SN
Anterior insula	-36	18	-10	5.00	SN/DN
Anterior insula	44	24	-8	4.86	DN/SN
<b>Moral &gt; Immoral</b>					
Retrosplenial cortex	-6	-36	20	-7.47	DN
Occipital cortex	24	-84	14	-5.32	Visual
Temporal pole	50	-6	-18	-5.20	DN
Temporal pole	-54	-6	-16	-4.73	DN
Superior parietal lobule	-24	-52	68	-4.69	DA
Somatomotor cortex	42	0	46	-4.13	DA/FP

Note: DN, default network; SN, salience network; FP, frontoparietal network; DA, dorsal attention network.

Network localization was done overlaying demarcations of a large-scale parcellation of the cerebral cortex (Yeo et al., 2011) on the contrast images using the Connectome Workbench visualization software (Marcus et al., 2011, see Figure 2, Supplementary Material).

right TPJ) was active early in the epoch, salience network regions (aINS, dACC) regions persisted throughout the epoch.

## Discussion

We examined behavioral performance and brain activation during a semantic judgment task involving implicit processing of moral, immoral and neutral information. Participants were slower in selecting morally laden vs neutral responses, with the greatest slowing observed during selections involving immoral content. Analyses of brain activity revealed that semantic judgments involving moral content (moral or immoral) were associated with default network activity, as well as engagement of executive control and salience network regions. In a direct contrast of semantic judgments involving moral vs immoral content, immoral content selections were associated with greater activation of salience network brain regions. These regions were active early in the post-decision-epoch, reflecting their purported role in detecting behaviorally-relevant stimuli (Corbetta et al., 2008). Consistent with predictions, salience network activity was also sustained, co-varying with frontal (lateral PFC) and default (medial PFC) later in the post-decision epoch.

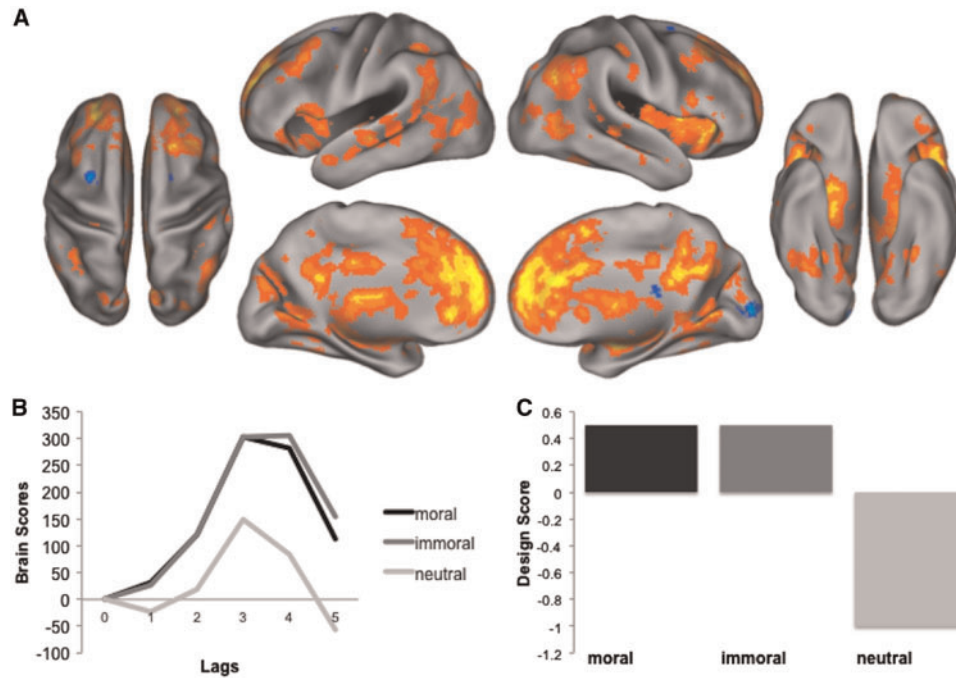


Fig. 3. Multivariate patterns dissociating moral and immoral from a neutral control condition. (A) Neural activity associated with the moral + immoral (warm colours) > neutral (cool colours), with bootstrap ratio  $\pm 2.81$ ,  $P < 0.005$ . Neural activity at lag 3 mapped onto the flattened Human PALS-B12 surface-based atlas using Caret5 (Van Essen, 2001, 2005). (B) Temporal brain scores plot. (C) Condition design scores convey the contrast.

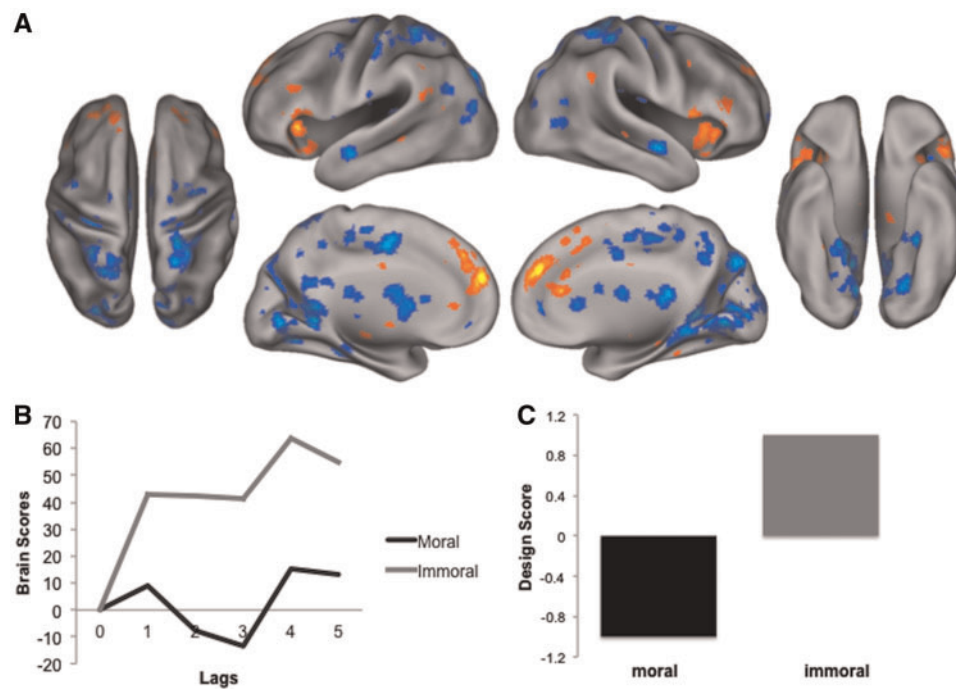


Fig. 4. Multivariate patterns dissociating moral from immoral condition. (A) Neural activity associated with the moral (cool colours) > immoral (warm colours), with bootstrap ratio  $\pm 2.81$ ,  $P < 0.005$ . Neural activity at lag 2 mapped onto the flattened Human PALS-B12 surface-based atlas using Caret5 (Van Essen, 2001, 2005). (B) Temporal brain scores plot. (C) Condition design scores convey the contrast.

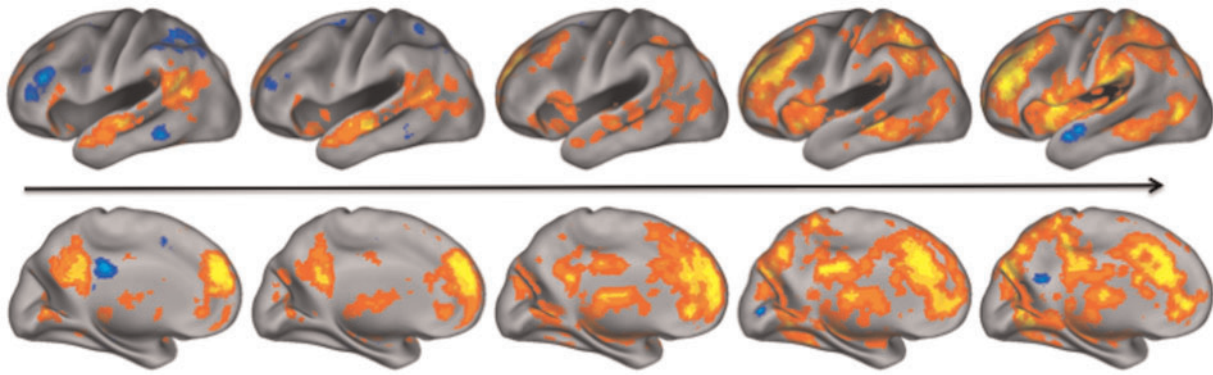


Fig. 5. Moral and Immoral (warm colours) > Neutral (cool colours),  $BSR \pm 2.81$ , Temporal unfolding from lag 1 to lag 5 mapped onto the flattened left hemisphere Human PALS-B12 surface-based atlas using Caret5 (Van Essen, 2001, 2005).

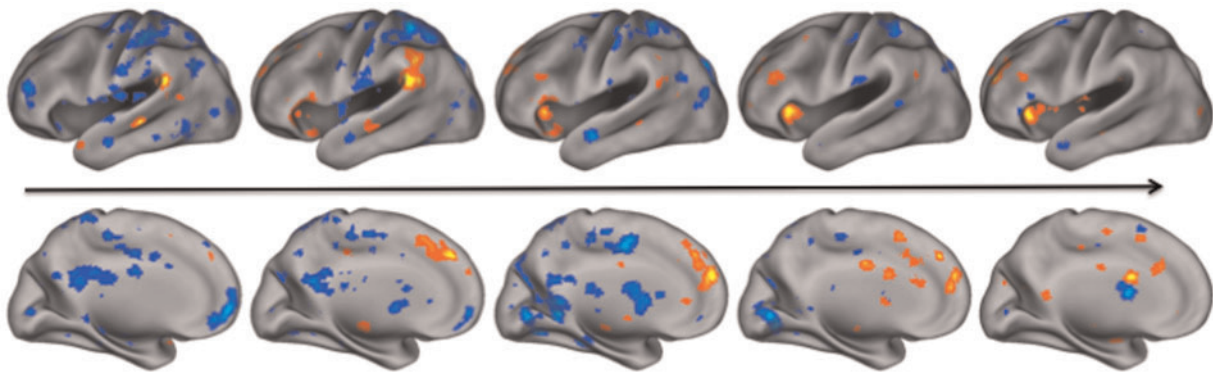


Fig. 6. Immoral (warm colours) > Moral (cool colours),  $BSR \pm 2.81$ , Temporal unfolding from lag 1 to lag 5 mapped onto the flattened left hemisphere Human PALS-B12 surface-based atlas using Caret5 (Van Essen, 2001, 2005).

Choices involving moral (immoral and moral) vs neutral content were reliably associated with activity in the regions of the default network as well as salience and executive control networks. The default network has been implicated in moral cognition previously (Harrison et al., 2008; Bzdok et al., 2012; Sevinc and Spreng, 2014). Moral cognition involves both mental simulation as well as theory of mind (Greene et al., 2004; Young et al., 2007). In line with current results, a NeuroSynth analysis using the term ‘moral’ yields neural activity in default network regions such as medial frontal cortex, TPJ, PCC, temporal pole and regions of the lateral temporal cortex. However, it is important to note that activity in default network regions may be associated with active moral judgment paradigms using a window of analysis usually time-locked to the subject’s decisions. Schaich-Borg and colleagues (2011) compared deliberation-locked (time-locked to stimulus onset modeled with no specified duration as variable length boxcar function) and verdict-locked activity (time-locked to participant response) and found that the two models were associated with differential activity patterns. In line with this, engagement of the salience network during morally-based tasks has been less frequently reported. Our findings suggest that salience network brain regions may play a critical role in early detection of information that is considered to be immoral, and thus may have greater behavioral relevance in social contexts.

Another important finding is the differential engagement of right TPJ for immoral content, especially early in the epoch. The rTPJ has been robustly associated with mentalizing (Saxe and Wexler, 2005), which underlies the ability to attribute mental

states to other individuals in order to explain, predict, and manipulate their behavior (van Overwalle and Baetens, 2009). With respect to the moral judgments, rTPJ has been demonstrated to be associated with encoding of other people’s beliefs, as well as with the use of those beliefs to achieve a moral judgment (Young and Saxe, 2009). Moreover, enhancement of rTPJ activation was associated more with negative moral judgments compared to positive judgments, suggesting that the participants engaged in mental state reasoning more robustly when confronted with immoral behaviors (Young et al., 2011). In line with our interpretation that rTPJ is involved in the automatic detection and evaluation of morally salient information, it has been suggested that the enhanced activation in rTPJ reflected greater attention to or deeper processing of mental state information especially when it supports a negative moral judgment, indicating neural evidence for ‘intuitive prosecution’ (Young et al., 2011). Further, salience brain regions were coactive with frontal and default network brain regions later in the post-decision epoch, suggesting that this detection of moral salience engaged downstream control resources, particularly for judgments involving immoral content.

Behaviorally, we observed a slowing of responses during judgments involving moral content. As the semantic association task did not involve explicit moral processing, these slowed RTs suggest that our manipulation was successful, with implicit, morally-laden choices leading to longer response times. We argue that this behavioral slowing for moral-content vs neutral choices reflects engagement of additional cognitive resources, supported by recruitment of default and executive

control brain regions. Morally-based choices likely evoke deeper associative processes necessary to integrate affective and cognitive experiences and guide decision-making in these conditions. Consistent with this interpretation, associative and elaborative processing has been associated with default (e.g. Bar, 2007) and default-executive network interactions (e.g. Spreng et al., 2010).

Our observation of sustained engagement of salience network regions, and their co-activation with default and executive regions, supports the recent suggestion that moral judgments, particularly those involving immoral content, engage arousal, motivation and cognitive re-appraisal processes (Christensen et al., 2014). We argue that the salience network, which has been implicated in the integration of affective and motivational processes during decision-making (Samanez-Larkin and Knutson, 2015), as well as coordination of executive control and associative processes (mediated by default and lateral frontal regions respectively, Uddin, 2015), are necessary for the arousal, motivational and cognitive appraisal processes required for moral, and particularly immoral, decision-making.

Detection of immoral behavior is an important facet of social interaction, and is necessary for survival. Identifying and endorsing immoral sentence completions, even outside of conscious awareness, is likely both behaviorally salient and cognitively challenging, involving early recruitment of the salience network and later engagement of default and cognitive control brain regions. The salience network is particularly involved in responding to behaviorally salient events and in the initiation of cognitive control (Seeley et al., 2007; Menon and Uddin, 2010). Nodes of the salience network are activated in response to a wide range of attention and cognitive control tasks such as the 'oddball' task, which involves detection of deviant stimuli embedded in a stream of standard stimuli (Crottaz-Herbette and Menon, 2006), and the stop signal and go-no-go tasks which assess inhibitory control (Swick et al., 2011). It is also involved in internally signaled errors in attentionally-demanding tasks (Ham et al., 2013). During the endorsement of immoral sentences, neural activity in salience network regions might therefore reflect the early detection of a potential moral violation—a behaviorally salient event, which creates cognitive conflict and signals recruitment of the default network for further contextual processing.

The salience network has recently been associated with impaired moral evaluations in psychopathy (Yoder et al., 2015) and also with perception of pain in others (Lamm et al., 2011). As mentioned earlier, the salience network is also implicated in bvFTD, which is characterized by social and moral impairments (Mendez and Shapira, 2009; Zhou et al., 2010). Salience network dysfunction has also been reported in other affective disorders such as psychosis (Kapur, 2003), schizophrenia (White et al., 2010) and autism (von dem Hagen, 2013). Aberrant salience detection and mapping of external stimuli and internal mental events has been suggested as a common mechanism underlying these disorders (Menon, 2011). In terms of moral behavior, the salience network might contribute to the moral processing, especially through early perception of antisocial actions and modulate the reappraisal of these actions, by diverting additional attentional resources towards associative processing.

The semantic association task used in the current study involved semantic judgments about real-world scenarios. Significant slowing of behavior during the immoral condition suggests that the selection of immoral vs moral or neural responses was more cognitively demanding, likely due to increased demand for associative and controlled processing (Bar, 2007). Since moral processing was incidental to the task

demands, it could also be argued that salience network engagement may reflect early, automatic components of moral processing. The early rise in temporal brain scores present in the moral-immoral contrast (Figure 4), is consistent with moral judgment involving early, bottom up attentional capture and later, top-down associative and appraisal processes. Salience network activity is associated with detection of behaviorally relevant stimuli in the environment (Corbetta and Shulman, 2008). Engagement of the salience network during the processing of immoral stimuli signals early detection and subsequent engagement of associated and control mechanisms to process more arousing, and arguably more behaviorally-salient, immoral content. This interpretation is consistent with recent research suggesting a complex integration of both early automatic and later controlled components during moral judgments (Yoder and Decety, 2014). Also, since moral and immoral content is present in all of three conditions as a control condition; we believe that the task specifically isolates neural activity associated with active engagement with the morally-laden content. This feature of the task might add to the behavioral saliency of the stimulus and might have contributed to the automatic processing of moral content in the service of goal-directed cognition.

It should also be noted that since the sample in the study is limited to young adults, further work is necessary to examine how current framework relates to other populations such as discrepant age groups where network interactions shift with healthy advancing age (Spreng and Schacter, 2012). Given the vulnerability of salience network to certain neurodegenerative diseases such as bvFTD, the modulatory role of the salience network for the processing of moral information may be associated with altered emotional moral responses in this population group (Mendez and Shapira, 2009).

Taken together, these results emphasize that salience detection is a crucial component of moral behavior, and suggest that detection of morally-relevant information in the environment is associated with salience detection systems in the human brain. These findings extend current descriptions of the neural basis of moral cognition by proposing that initial detection of morally-salient information involves discrete brain regions from those engaged by more contextual or associative processing of moral information during judgment and decision-making. This model of moral judgment, implicating the salience network in early detection of behaviorally-relevant, moral information, and later coordination of associative and appraisal processes, suggests that both bottom-up and top-down attentional influences are involved in everyday moral judgments. Further investigation, however, is necessary for a more conclusive network engagement and coupling. Different experimental designs, optimized for investigation of interregional functional connectivity, are required to provide more information on functional coupling of neural networks during moral processing as well as for the temporal co-evaluation of these network interactions (c.f. Dixon et al., 2017).

## Acknowledgements

This project was supported in part by NIH grant 1S10RR025145. We thank Wen-Ming Luh, Emily Qualls, Elizabeth DuPre, Juliana Garcia, Erica Miller and Jennifer Lieberman for their assistance with this project.



## Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

## References

- Andrews-Hanna, J.R., Smallwood, J., Spreng, R.N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, **1316**, 29–52.
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, **11**(7), 280–9.
- Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, **1124**(1), 1–38.
- Bzdok, D., Schilbach, L., Vogeley, K., et al. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, **217**(4), 783–96.
- Chiong, W., Wilson, S.M., D'Esposito, M., et al. (2013). The salience network causally influences default mode network activity during moral reasoning. *Brain*, **136**, 1929–41.
- Christensen, J.F., Flexas, A., Calabrese, M., Gut, N.K., Gomila, A. (2014). Moral judgment reloaded: a moral dilemma validation study. *Frontiers in Psychology*, **5**, 607.
- Corbetta, M., Patel, G., Shulman, G.L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, **58**(3), 306–24.
- Craig, A.D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, **3**(8), 655–66.
- Critchley, H.D. (2005). Neural mechanisms of autonomic, affective, and cognitive integration. *Journal of Comparative Neurology*, **493**(1), 154–66.
- Crottaz-Herbette, S., Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response: combined fMRI and ERP evidence. *Journal of Cognitive Neuroscience*, **18**(5), 766–80.
- Damasio, A.R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Houghton Mifflin Harcourt.
- Dixon, M.L., Andrews-Hanna, J.R., Spreng, R.N., Irving, Z.C., Christoff, K. (2017). Anticorrelations between default and dorsal attention networks vary by default subsystem and across cognitive states. *NeuroImage*, **147**, 632–49.
- Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, **293**(5537), 2105–8.
- Greene, J., Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, **6**(12), 517–23.
- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, **44**(2), 389–400.
- Ham, T., Leff, A., de Boissezon, X., Joffe, A., Sharp, D.J. (2013). Cognitive control and the salience network: an investigation of error processing and effective connectivity. *The Journal of Neuroscience*, **33**(16), 7091–8.
- Harrison, B.J., Pujol, J., López-Solà, M., et al. (2008). Consistency and functional specialization in the default mode brain network. *Proceedings of the National Academy of Sciences*, **105**(28), 9781–6.
- Heekeren, H.R., Wartenburger, I., Schmidt, H., Schwintowski, H.P., Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport*, **14**(9), 1215–9.
- Krishnan, A., Williams, L.J., McIntosh, A.R., Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage*, **56**(2), 455–75.
- Kundu, P., Inati, S.J., Evans, J.W., Luh, W.-M., Bandettini, P.A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, **60**, 1759–70.
- Lamm, C., Decety, J., Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, **54**(3), 2492–502.
- Marcus, D., Harwell, J., Olsen, T., et al. (2011). Informatics and data mining tools and strategies for the human connectome project. *Frontiers in Neuroinformatics*, **5**, 4.
- McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage*, **3**(3), 143–57.
- Mendez, M.F., Shapira, J.S. (2009). Altered emotional morality in frontotemporal dementia. *Cognitive Neuropsychiatry*, **14**(3), 165–79.
- Menon, V., Levitin, D.J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage*, **28**(1), 175–84.
- Menon, V., Uddin, L.Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, **214**(5–6), 655–67.
- Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in Cognitive Sciences*, **15**(10), 483–506.
- Mesulam, M.M. (1998). From sensation to cognition. *Brain*, **121**(6), 1013–52.
- Moll, J., de Oliveira-Souza, R., Eslinger, P.J., et al. (2002a). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience*, **22**(7), 2730–6.
- Moll, J., de Oliveira-Souza, R., Bramati, I.E., Grafman, J. (2002b). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage*, **16**(3), 696–703.
- Moll, J., de Oliveira-Souza, R., Moll, F.T., et al. (2005). The moral affiliations of disgust: a functional MRI study. *Cognitive and Behavioral Neurology*, **18**(1), 68–78.
- Monin, B., Pizarro, D.A., Beer, J.S. (2007). Deciding versus reacting: conceptions of moral judgment and the reason-affect debate. *Review of General Psychology*, **11**(2), 99.
- Ongur, D., Price, J.L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, **10**(3), 206–19.
- Putchá, D., Ross, R.S., Cronin-Golomb, A., Janes, A.C., Stern, C.E. (2015). Altered intrinsic functional coupling between core neurocognitive networks in Parkinson's disease. *NeuroImage: Clinical*, **7**, 449–55.
- Reniers, R.L., Corcoran, R., Völlm, B.A., Mashru, A., Howard, R., Liddle, P.F. (2012). Moral decision-making, ToM, empathy and the default mode network. *Biological Psychology*, **90**(3), 202–10.
- Samanez-Larkin, G.R., Knutson, B. (2015). Decision making in the ageing brain: changes in affective and motivational circuits. *Nature Reviews Neuroscience*, **16**(5), 278–89.
- Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, **43**(10), 1391–9.

- Schaich-Borg, J., Sinnott-Armstrong, W., Calhoun, V.D., Kiehl, K.A. (2011). Neural basis of moral verdict and moral deliberation. *Social Neuroscience*, *6*(4), 398–413.
- Seeley, W.W., Menon, V., Schatzberg, A.F., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience*, *27*(9), 2349–56.
- Seeley, W.W. (2010). Anterior insula degeneration in frontotemporal dementia. *Brain Structure and Function*, *214*(5–6), 465–75.
- Sevinc, G., Spreng, R.N. (2014). Contextual and perceptual brain processes underlying moral cognition: a quantitative meta-analysis of moral reasoning and moral emotions. *PLoS One*, *9*(2), e87427.
- Spreng, R.N., Grady, C.L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience*, *22*(6), 1112–23.
- Spreng, R.N., Schacter, D.L. (2012). Default network modulation and large-scale network interactivity in healthy young and old adults. *Cerebral Cortex*, *22*(11), 2610–21.
- Spreng, R.N., Sepulcre, J., Turner, G.R., Stevens, W.D., Schacter, D.L. (2013). Intrinsic architecture underlying the relations among the default, dorsal attention, and frontoparietal control networks of the human brain. *Journal of Cognitive Neuroscience*, *25*(1), 74–86.
- Swick, D., Ashley, V., Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *Neuroimage*, *56*(3), 1655–65.
- Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry*, *160*(1), 13–23.
- Uddin, L.Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, *16*(1), 55–61.
- Van Essen, D.C. (2005). A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. *Neuroimage*, *28*, 635–62.
- Van Essen, D.C., Dickson, J., Harwell, J., Hanlon, D., Anderson, C.H., Drury, H.A. (2001). An integrated software system for surface-based analyses of cerebral cortex. *Journal of American Medical Informatics Association*, *8*(5), 443–59.
- Van Overwalle, F., Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564–84.
- von dem Hagen, E.A., Stoyanova, R.S., Baron-Cohen, S., Calder, A.J. (2012). Reduced functional connectivity within and between 'social' resting state networks in autism spectrum conditions. *Social Cognitive and Affective Neuroscience*, *8*, 694–701.
- White, T.P., Joseph, V., Francis, S.T., Liddle, P.F. (2010). Aberrant salience network (bilateral insula and anterior cingulate cortex) connectivity during information processing in schizophrenia. *Schizophrenia Research*, *123*(2), 105–15.
- Yeo, B.T., Krienen, F.M., Sepulcre, J., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–65.
- Young, L., Cushman, F., Hauser, M., Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–40.
- Young, L., Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–405.
- Young, L., Scholz, J., Saxe, R. (2011). Neural evidence for "intuitive prosecution": the use of mental state information for negative moral verdicts. *Social Neuroscience*, *6*(3), 302–15.
- Yoder, K.J., Decety, J. (2014). Spatiotemporal neural dynamics of moral judgment: a high-density ERP study. *Neuropsychologia*, *60*, 39–45.
- Yoder, K.J., Harenski, C., Kiehl, K.A., Decety, J. (2015). Neural networks underlying implicit and explicit moral evaluations in psychopathy. *Translational Psychiatry*, *5*, e625; doi:10.1038/tp.2015.117.
- Zhou, J., Greicius, M.D., Gennatas, E.D., et al. (2010). Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain*, *133*(5), 1352–67.