ORIGINAL ARTICLE

Dissociating the Neural Basis of Conceptual Self-Awareness from Perceptual Awareness and Unaware Self-Processing

Pawel Tacikowski, Christopher C. Berger and H. Henrik Ehrsson

Department of Neuroscience, Karolinska Institute, SE-17177 Stockholm, Sweden

Address correspondence to Pawel Tacikowski, Department of Neuroscience, Karolinska Institute, Retzius väg 8, SE-17177 Stockholm, Sweden.
Email: pawel.tacikowski@ki.se

Abstract

Conceptual self-awareness is a mental state in which the content of one's consciousness refers to a particular aspect of semantic knowledge about oneself. This form of consciousness plays a crucial role in shaping human behavior; however, little is known about its neural basis. Here, we use functional magnetic resonance imaging (fMRI) and a visual masked priming paradigm to dissociate the neural responses related to the awareness of semantic autobiographical information (one's own name, surname, etc.) from the awareness of information related to any visual stimulus (perceptual awareness), as well as from the unaware processing of self-relevant stimuli. To detect brain activity that is highly selective for self-relevant information, we used the blood-oxygen-level-dependent (BOLD) adaptation approach, which goes beyond the spatial limitations of conventional fMRI. We found that self-awareness was associated with BOLD adaptation in the medial frontopolar-retrosplenial areas, whereas perceptual awareness and unaware self-processing were associated with BOLD adaptation in the lateral fronto-parietal areas and the inferior temporal cortex, respectively. Thus, using a direct manipulation of conscious awareness we demonstrate for the first time that the neural basis of conceptual self-awareness is neuroanatomically distinct from the network mediating perceptual awareness of the sensory environment or unaware processing of self-related stimuli.

Key words: BOLD adaptation, default mode network, fMRI, lateral fronto-parietal network, visual masked priming

Introduction

Who am I? Humans have asked themselves this question since the dawn of man. Apart from the existential relevance for each individual, the very ability to pose this question is intriguing: it requires conceptual self-awareness, that is, a mental state in which the content of one's consciousness refers to a given aspect of knowledge about oneself, for example, when we introduce ourselves to other people or reflect about our personality and identity (Duval and Wicklund 1972; Baumeister 1998; Zeman 2001; Schooler 2002; Carver 2012; Lou et al. 2016).

This type of consciousness is considered the hallmark of the human mind, because both ontogenetically and phylogenetically, it develops as one of the last cognitive functions (Gallup 1997; Zelazo 2004). Awareness of one's own personality and identity, with associated thoughts and feelings, allows us to self-reflect, develop as individuals in a social context, and generate complex behaviors (Baumeister 1998; Zeman 2001; Lou et al. 2016). Despite its importance in everyday life, surprisingly little is known about the neural substrate of conceptual self-awareness.

In contrast, we have a relatively good understanding of the neurocognitive basis of perceptual awareness, that is, the awareness of events and objects in the external environment. By using paradigms in which perceptual awareness is explicitly manipulated to enable the dissociation of aware and unaware processing (e.g., by using a visual masking procedure) of the same sensory stimulus, perceptual awareness has been associated with activity in the anatomically interconnected network of areas located in the lateral fronto-parietal association cortices (Rees et al. 2002; Dehaene et al. 2006; Dehaene and Changeux 2011; but see Koch et al. 2016; Tononi et al. 2016). Specifically, the awareness of external sensory stimuli is considered to arise when information about these stimuli becomes globally distributed within the fronto-parietal network, which in turn facilitates the flexible control of behavior depending on the task at hand (Dehaene et al. 2006; Dehaene and Changeux 2011). However, these theories do not explicitly address the problem of self-awareness, and in fact, it is a common assumption in the field that regardless of what type of information we become aware of—self-relevant or not—it is supported by the same set of neural mechanisms (Crick and Koch 1998) within the lateral fronto-parietal network (Dehaene and Changeux 2011). To the best of our knowledge, no published neuroimaging study has directly tested this assumption by comparing masked and unmasked stimuli that are either self-related or not self-related within a single experimental design. Thus, it remains unknown whether conceptual self-awareness and perceptual awareness rely on common or separate neural substrates.

Apart from being dissociated from the neural correlates of perceptual awareness, the neural basis of conceptual self-awareness should also be differentiated from the brain mechanisms of self-related information processing without awareness, or "unaware self-processing." The conceptual self is a multicomponent and multilevel mental construct, which could be accessed consciously or unconsciously (Baumeister 1998; Greenwald et al. 2002; Morin 2006; Carver 2012; Oyserman et al. 2012). For example, when we reflect about our personality or identity, we access the aware aspects of self-concept. In contrast, turning our head automatically before even consciously realizing that someone has called our name is an example of unaware self-processing. Behavioral evidence suggests that self-relevant information is indeed detected automatically already at unconscious stages of stimulus processing hierarchy (Bargh 1982; Alexopoulos et al. 2012; Pfister et al. 2012). Previous neuroimaging studies have associated the processing of conceptual self-related information with activations in the medial and lateral prefrontal cortex, posterior cingulate cortices (Northoff et al. 2006; Qin and Northoff 2011), hippocampus, temporo-parietal junction, and inferior occipito-temporal cortex (Martinelli et al. 2013). However, as described above, no published study has experimentally manipulated the "awareness" of self-processing; therefore, it remains unknown whether conscious and unconscious components of self-concept have overlapping or distinct brain bases.

Here, we combine advanced functional magnetic resonance imaging (fMRI) and a classical visual masking paradigm to test the hypothesis that conceptual self-awareness is associated with activity in a distinct network of areas including the medial prefrontal cortex (MPFC) and the posterior cingulate cortex (PCC). This hypothesis is based on the idea that information related to the self is processed in a very different way from that of information that is self-irrelevant; hearing our own name called out, for example, leads to much stronger attentional, emotional, and behavioral reactions than does hearing another person's name. Decades of studies in psychology and cognitive neuroscience have shown that self-relevant information is used to construct a multilevel and up-to-date representation of the self (Churchland 2002). These bodily, perceptual, and conceptual self-constructs serve as "reference points" in a variety of physiological, cognitive, and social processes (Schacter et al. 2003; Mitchell et al. 2006; Ehrsson 2007; Bergouignan et al. 2014; Sui and Humphreys 2015). The medial prefrontal and posterior cingulate cortices likely contain a higher order representation of self-concept because these regions are more active during the processing of various types of self-related stimuli (Northoff et al. 2006; Sugiura et al. 2008; Qin and Northoff 2011) across different modalities (Tacikowski et al. 2013) and across different situational contexts, for example, during self-focused daydreaming, imagining one's future, or autobiographical remembering (Cabeza and St Jacques 2007; Buckner et al. 2008; Schooler et al. 2011). Anatomically, these midline prefronto-cingulate structures are interconnected (Greicius et al. 2009) and they show a high degree of functional connectivity during "rest," constituting the so-called "default mode network" (DMN) (Raichle et al. 2001; Buckner et al. 2008). Given the above, activity in these midline fronto-cingulate areas constitutes an ideal candidate system to mediate human conceptual self-awareness.

During fMRI, 26 healthy participants discriminated whether visually presented names, surnames, dates of birth, and nationality codes were self- or other-related. We used these semantic autobiographical stimuli because they are at the core of one's social identity (Watson 1986; Gray et al. 2004; Martinelli et al. 2013) and at the same time they can be presented very briefly allowing conservative manipulations of awareness. Specifically, in half of the trials, the visual masked priming procedure was used to drastically reduce awareness of the stimuli (unaware context), whereas in the other half, there were no masks, and the stimuli were fully visible (aware context). As a result, we had 4 conditions in a 2×2 design with "person" (self vs. other) and "awareness" (aware vs. unaware) as the factors (Fig. 1A). Importantly, this design allowed us to disambiguate conceptual self-awareness from perceptual awareness and unaware self-processing by identifying neural responses that are specific to the aware processing of self-relevant information (i.e., "person" × "awareness" interaction).

Moreover, to ensure the identification of genuine self-selective cortical responses, we used the blood-oxygen-level-dependent (BOLD) adaptation approach. This approach goes beyond the spatial limitations of conventional fMRI and allows detection of overlapping neuronal populations (Grill-Spector et al. 2006). The BOLD-adaptation effect is based on a well-established electrophysiological finding that neurons responding to a certain stimulus feature, for example, shape or color, show reduced firing rates when stimuli possessing this feature are presented repetitively, that is, the so-called "repetition suppression effect" (Li et al. 1993). Thus, the BOLD adaptation corresponds to reduced signal amplitude in congruent trials (i.e., subsequent presentation of stimuli sharing a given feature) than in incongruent trials (i.e., subsequent presentations of stimuli not sharing this feature) (Grill-Spector et al. 2006) and this reduction is robust for both unconscious and conscious cognitive processes (Dehaene et al. 2001; Grill-Spector et al. 2006; Nakamura et al. 2007; Kouider et al. 2010). Here, we presented congruent or incongruent pairs of self-related or nonself-related stimuli (Fig. 1A) to look for areas that show BOLD adaptation only to congruent pairs of self-relevant
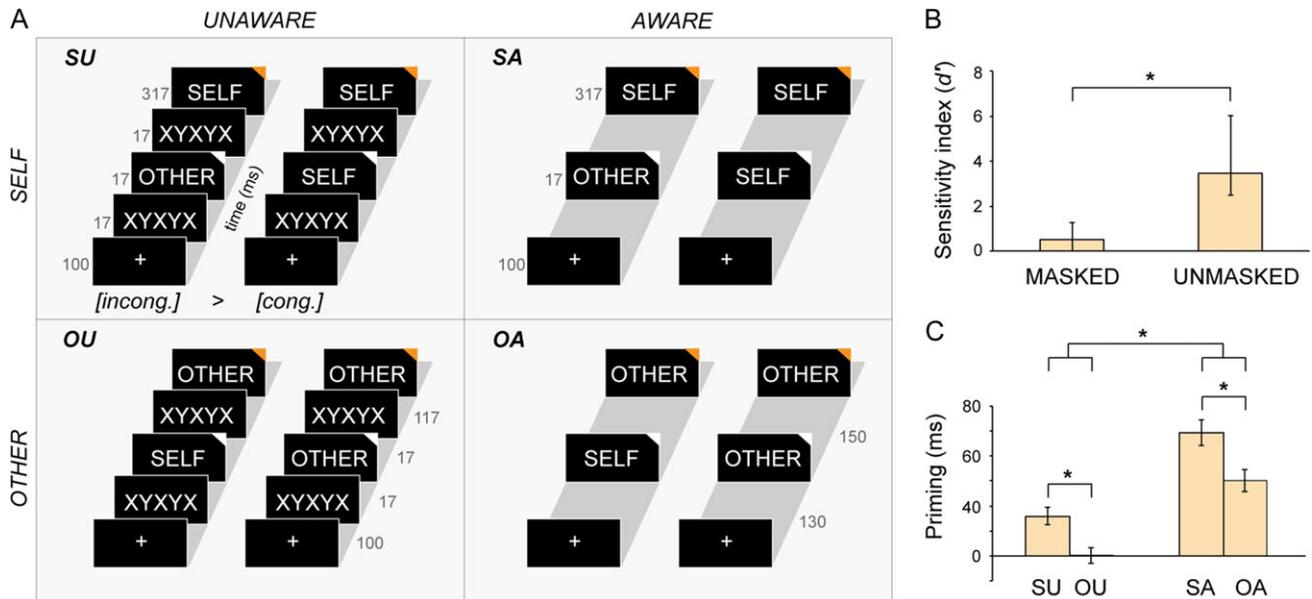
Figure 1. Experimental design and behavioral results. (A) Self- and other-related stimuli (names, surnames, dates of birth, and nationality codes) were presented visually, and the participants' task was to decide (by pressing one of the two buttons) whether stimuli occurring at the end of each trial (the targets; indicated by orange triangles) were related to themselves or to another person. In half of the trials, stimuli appearing before the targets (the primes; white triangles) were immediately preceded and followed by visual masks, thus making the aware processing of these primes highly demanding. In the other half of the trials, there were no masks, which made the aware processing easy. Stimulus durations and intervals between stimuli are indicated in milliseconds, next to the time axis. (B) The results of the visibility test showed that subjects were much worse at discriminating masked primes than unmasked primes (zero indicates the chance level; bar plots indicate medians ± interquartile ranges). (C) The priming effect (RTs incongruent > congruent trials) was significantly stronger in the "self" than in the "other" condition and stronger in the "aware" than in the "unaware" condition (bars indicate means ± standard errors). cong., congruent trials; incong., incongruent trials.

stimuli, meaning that these areas contain neuronal populations that process self-related information in a highly selective fashion. By combining this approach with the experimental manipulation of awareness (in the factorial design as described above), we could provide the first conclusive neuroimaging evidence for the neural basis of conceptual self-awareness.

In a purely descriptive approach, we also examined the potential overlap between areas involved in conceptual self-awareness and DMN. Overlap between areas associated with the processing of self-related information and DMN maps obtained during a "resting state" has been demonstrated before (Gusnard et al. 2001; D'Argembeau et al. 2005; Buckner et al. 2008; Qin and Northoff 2011; Whitfield-Gabrieli et al. 2011). Therefore, we thought that it would also be valuable to examine possible anatomical correspondence for areas specifically related to the awareness of self-related information processing. To that end, we collected resting-state data from all our participants prior to the main fMRI experiment and could therefore directly compare DMN and BOLD-adaptation responses related to conceptual self-awareness within the same group of subjects.

Materials and Methods

Participants

Twenty-six right-handed volunteers (mean age: 25 ± 4, 11 females) participated in the study; two additional participants were excluded due to excessive head movement. All participants were healthy, reported no history of psychiatric illness or neurological disorder, and had normal or corrected-to-normal vision. All participants gave their written informed consent before the start of the experiment. The Regional Ethical Review Board of Stockholm approved the study, and the study was

performed according to the guidelines of the Declaration of Helsinki 1975.

Design and Procedures

During the fMRI-BOLD-adaptation session, we presented self-related stimuli (i.e., each participant's own name, surname, year of birth, and nationality) and other-related stimuli (i.e., another person's name, surname, year of birth, and nationality). Nationality was indicated by a 3-letter international code (e.g., "FRA" for France) according to the ISO 3166 norm. All words were written in white capital letters (Arial font) and were presented centrally on a black background. Self- and other-related stimuli were the same size (names and surnames were matched to have the same number of letters), so did the dates and nationality codes. The participant's own and other persons' names were also matched in terms of gender. Each trial began with a fixation point (i.e., a white "+"), followed by a prime, followed by a target. In half of the trials, the primes were immediately preceded and immediately followed by forward and backward masks, respectively. In all trials, participants had 1500 ms to respond (i.e., the time from the onset of the target to the end of the trial). Figure 1A provides specific information about the onsets, durations, and intervals between events in different types of trials. Recordings using a high-speed camera (1000 Hz, model Exilim EX-ZR100, Casio Computer Ltd) confirmed that the forward masks, primes, and backward masks were displayed for 17 ± 1 ms each and that the intervals between them were 17 ± 2 ms long. In summary, apart from factors that were explicitly manipulated—the occurrence or absence of masks, the person that primes and targets referred to, and the prime-target congruency—all other aspects of stimulus presentation were matched between trial types.

There were 8 trial types: self-masked incongruent; self-masked congruent; other-masked incongruent; other-masked congruent; self-unmasked incongruent; self-unmasked congruent; other-unmasked incongruent; and other-unmasked congruent (Fig. 1A). Each trial type was repeated 24 times. The 192 trials were presented in pseudorandom order (not more than 3 consecutive repetitions of the same trial type) in 2 runs (each run lasted ca. 10 min, during which 96 trials were presented). The break between 2 runs was approximately 2 min. Participants were instructed to ignore the primes and respond only to the targets. In particular, the participants' task was to discriminate whether each target was self-related (e.g., the participant's own name or surname) or other-related (e.g., someone else's name or surname) by pressing one of two buttons on a hand-held MR-compatible response pad (model HHSC-1x4-D, Current Designs, Inc.). For both types of responses, participants used their right thumbs. The instructions emphasized that both the speed and the accuracy of responses were important. Before the actual study, the participants practiced the task on a computer located outside the scanner (16 trials; 2 repetitions of each trial type). The key assignment (i.e., left key for self and right key for other or vice versa) was counterbalanced across participants.

Each name, surname, date, and nationality code was repeated the same number of times within one trial type (6 repetitions resulting in 24 trials per trial type). In all trials, the primes and targets belonged to the same stimulus category; that is, a name prime was always followed by a name target, a surname prime by a surname target, and so on. In the congruent trials, the primes and targets referred to the same person, whereas in the incongruent trials, they referred to different people (Fig. 1A). The intertrial intervals (ITIs) were 3, 4, or 5 s, and they were randomly distributed across the 192 trials. Different ITIs were used to optimize the BOLD responses and signal discrimination between trial types. The stimuli were displayed via an LCD projector (NEC MT1050; NEC Display Solutions Ltd) on a screen that the volunteers viewed through a mirror mounted on the head coil of the scanner. Presentation software (version 16.4, Neurobehavioral Systems, Inc.) was used to control the timing of stimulus presentation and to record participants' responses.

After the main fMRI-BOLD-adaptation session, while still inside the scanner, the participants performed a visibility test. Compared with the BOLD-adaptation session, this test differed only in 2 ways: 1) each trial type was repeated 12 times instead of 24, so that the 96 trials were presented nonstop for approximately 10 min and 2) the participants were asked to decide whether the primes, not the targets, were self- or other-related. The instructions for this part emphasized that if the participants were unsure, they should attempt to guess the correct answer (forced choice). The key assignment in the "visibility test" was always the same as in the BOLD-adaptation session. Crucially, the purpose of the visibility test was to validate the experimental manipulations from the main fMRI-BOLD-adaptation session. Because the stimuli were presented in exactly the same way during both sessions, the observation that the participants could not recognize masked primes above chance level during the visibility test indicates that masked primes were processed without awareness during the fMRI-BOLD-adaptation session as well. Analogously, showing that recognition of unmasked primes was above chance level during the visibility test indicates that unmasked primes were also processed consciously in the fMRI-BOLD-adaptation session. Similar logic has been applied in many previous studies (Dehaene et al. 2001; Grill-Spector et al. 2006; Nakamura et al. 2007; Kouider et al. 2010; Alexopoulos et al. 2012; Pfister et al. 2012).

MRI Data Acquisition

We used a General Electric 3T Discovery MR750 scanner equipped with an 8-channel head coil. First, we collected resting-state data; during this session, the participants were asked to look at a white "+" centered on a black screen, relax, and try not to engage in any specific mental activity. We acquired 150 echo-planar image (EPI) volumes for each participant, using the following scanning parameters: time repetition (TR) = 2200 ms; time echo (TE) = 30 ms, matrix size 72 × 72, field-of-view (FOV) = 220 × 220 mm; flip angle (FA) = 70 degrees; 47 axial slices, voxel size: 3 × 3 × 3 mm. Next, 2 fMRI scan runs were conducted; this functional data consisted of 580 EPI volumes per subject, acquired with the following parameters: TR = 2000 ms; TE = 30 ms; matrix size 72 × 72; FOV = 220 × 220 mm; FA = 70 degrees; number of axial slices: 42; voxel size = 3 × 3 × 3 mm. At the end of the experiment, when the visibility test was conducted, we acquired a high-resolution anatomical scan, using the 3D Cube T2 FLAIR sequence with the following parameters: TR = 8000 ms; FOV = 270 × 270 mm; number of axial slices: 176; matrix size: 256 × 256; voxel size 1 × 1 × 1 mm.

fMRI First-Level Analysis

The fMRI data were analyzed using the Statistical Parametric Mapping software, version 8 (SPM8; http//:www.fil.ion.ucl.ac.uk/spm; Wellcome Department of Cognitive Neurology, London, UK). Preprocessing included the following steps: 1) slice-timing correction; 2) realignment of all functional volumes to the first volume; 3) coregistration of the functional and the anatomical scans; 4) segmentation of the structural scan into the white matter, gray matter, and cerebrospinal fluid partitions; 5) normalization of these partitions to the Montréal Neurological Institute (MNI) standard brain; 6) application of the same transformations to all functional volumes; 7) resampling of the functional volumes to 1.5 × 1.5 × 1.5 mm voxel size; and 8) spatial smoothing of functional images with an 8-mm full-width-at-half-maximum (FWHM) isotropic Gaussian kernel.

Next, a linear regression model was fitted to each participant's data with regressors defined for each trial type. Event timings were based on the onsets of target stimuli. We used only trials in which the subjects responded correctly; all incorrect trials were included in a common nuisance regressor. Each regressor was then convolved with the canonical hemodynamic response function provided in SPM8. Linear contrasts were defined within the general linear model (GLM). In particular, we calculated the difference between incongruent and congruent trials for the self-unaware (SU), other-unaware (OU), self-aware (SA), and other-aware (OA) conditions. In addition to taking advantage of the BOLD-adaptation approach (see Introduction), the calculation of incongruent > congruent contrasts for each condition excluded possible confounding factors due to the presence or absence of masks or due to different motor-response requirements for self- and other-related targets (these procedural aspects were identical in the congruent and incongruent trials within a given condition; therefore, they canceled each other out when calculating the BOLD-adaptation contrasts).

The subject-specific regression models also included movement outliers, 6 movement regressors, and 1 block regressor. The movement outliers were identified using the Artifact

Detection Toolbox (ART, www.nitrc.org/projects/artifact_detect). Specifically, an image was defined as an outlier if the head displacement in the x, y, or z direction was greater than 0.5 mm from the previous frame, if the rotational displacement was greater than 0.02 radians from the previous frame, or if the global mean intensity in the image was greater than 3 standard deviations from the mean image intensity for the entire functional scan. Outlier images were not deleted from the time series but modeled in the GLM thus the temporal structure of the data was not disrupted. Each outlier was represented by a single regressor in the GLM, with a "1" for outlier time points and "0" elsewhere. Two participants had outliers in more than 10% of the volumes; therefore, we excluded those subjects from further analysis (see the "Participants" section). The data quality for the remaining 26 subjects was very high; artifacts were detected in only $2 \pm 1\%$ of all volumes.

fMRI Second-Level Data Analysis

Contrast images that represented the strength of the BOLD-adaptation effect in SU, OU, SA, and OA conditions for each participant were entered into the random-effects group analysis, specified as a two-way repeated-measures ANOVA, with "awareness" (aware vs. unaware) and "person" (self vs. other) as the factors. The effects of interest were identified by the following t-contrasts: 1) active regions related to conceptual self-awareness: (SA − OA) − (SU − OU), masked inclusively by the SA − OA and SA − SU contrasts; 2) active regions related to perceptual awareness: (SA + OA) − (SU + OU), masked inclusively by the SA − SU and OA − OU contrasts; 3) active regions involved in unaware processing of self-related information: (SA + SU) − (OA + OU), masked inclusively by the SA − OA and SU − OU contrasts; and 4) active regions involved in unaware processing of any perceptual information: (SU + OU + SA + OA) > 0, masked inclusively by the (SU + OU) > 0 contrast. Using inclusive masks allowed us to identify brain regions that showed

the directed patterns of the BOLD-adaptation responses of interest, for example, voxels that showed a significant interaction with strong effects in the SA condition as opposed to the contrast being driven by differences between less important conditions. The threshold for all masks was set to P < 0.05 (uncorrected) in a purely descriptive fashion. Note that inclusive masks do not change the inferential descriptive statistics of the main contrasts, and thus in the current study, the masks simply excluded the voxels that showed a pattern of responses that we did not consider psychologically meaningful. In Table 1, we report only activation peaks that surpassed the P < 0.05 threshold corrected for multiple comparisons using the family-wise error (FWE) method, either at the whole-brain cluster level or at the level of small volumes based on a priori neuroanatomical hypotheses. In Supplementary Table 1, we additionally report all major clusters (k > 10) that survived the threshold of P < 0.001 (uncorrected).

The small volumes in the present study (18 mm radius spheres) were centered on the activation peaks reported by previous studies on self-related information processing and on perceptual awareness. In a recent meta-analysis of 13 neuroimaging papers, Martinelli et al. (2013) have shown that the bilateral medial temporal cortices, medial prefrontal, and posterior cingulate cortices, as well as inferior temporal and occipital regions are involved in the processing of visual semantic autobiographical information, that is, exactly the type of stimuli that we have used in the current study. We hypothesized that within this network the medial prefrontal and posterior cingulate cortices are related to the aware processing of self-related information, whereas the inferior occipital and temporal regions are related to the unaware aspects of self-processing. As a result, the small-volume correction for the conceptual self-awareness contrast and the unaware self-processing contrast were based on the findings of Martinelli and colleagues (2013, Table II); all coordinates reported by Martinelli and colleagues were converted from the Talairach to the MNI space using the GingerALE

Table 1 Results of fMRI analysis

Contrast/anatomical region	x	y	z	t value	Voxels	P value[a]
Conceptual self-awareness						
R. medial frontopolar gyrus (frontal pole)[b]	4	56	−11	4.81	1243	0.008[c]
R. superior temporal gyrus (temporal pole)	32	6	−26	3.96	78	0.016
L. superior temporal gyrus (temporal pole)	−30	2	−26	4	117	0.034
L. hippocampus[b]	−21	−21	−15	3.69	296	0.016
L. cingulate gyrus (retrosplenial cortex)[b]	−8	−54	7	3.55	238	0.049
Perceptual awareness						
L. fusiform gyrus	−38	−45	−23	4.38	81	0.004
L. superior frontal gyrus (SMA)	−9	−9	51	3.61	241	0.032
R. superior frontal gyrus (pre-SMA)	9	9	51	3.71	32	0.042
L. superior frontal gyrus (pre-SMA)	−8	12	48	3.55	72	0.049
L. supramarginal gyrus/intraparietal sulcus	−45	−52	51	3.66	197	0.036
L. intraparietal sulcus	−34	−61	57	3.55	57	0.048
L. precentral sulcus, superior part	−38	5	37	3.85	151	0.021
Self-unaware						
L. inferior temporal sulcus, anterior part	−45	−7	−29	3.76	45	0.028
Perceptual-unaware						
R. cingulate gyrus, middle part	9	−16	27	4.31	944	0.024[c]
L. fusiform gyrus	−32	−52	−17	3.61	119	0.042

Note: Labels of functional regions provided in brackets follow labels of anatomical regions. SMA, supplementary motor area. For details on how "conceptual self-awareness," "perceptual awareness," etc., were operationalized, see the "fMRI Second-Level Data Analysis" section.
[a]All brain regions (peaks) with significant activation (P < 0.05, FWE correction) at a voxel level, using small volumes as search spaces.
[b]Specific localizations of these peaks are displayed in Figure 3.
[c]Brain regions (peaks) with significant activation (P < 0.05, FWE correction) at a whole-brain cluster level.

software (http://brainmap.org). With regard to brain regions involved in perceptual awareness, we expected to see activations mainly in the lateral fronto-parietal regions. The occipital and inferior temporal regions, in turn, were hypothesized to be involved in the unaware processing of perceptual information in our paradigm. For these 2 contrasts, the small-volume correction was based on the findings of Dehaene et al. (2001), that is, activations associated with unmasked and masked words, respectively (Experiment I in their study). The anatomical localization of activations in the present study was based on the major sulci and gyri (Duvernoy 1991), distinguishable on a mean MRI generated from the standardized anatomical MRIs of the 26 participants.

Resting-State Data Analysis

Resting-state functional connectivity was analyzed using the CONN-toolbox (http://www.nitrc.org/projects/conn/). Spatial preprocessing of functional volumes included slice-timing correction, realignment, normalization, and smoothing (8-mm FWHM Gaussian filter), using SPM8 default parameters. Temporal time series that characterized estimated motion (3 rotation, 3 translation parameters, and movement-related artifacts identified by ART; see the earlier description), as well as the BOLD time series within the subject-specific white matter and cerebrospinal fluid masks, were used as temporal covariates and were removed using linear regression (please note that the "CompCor" method of noise reduction implemented in the CONN-toolbox does not use global signal regression). The resulting residual BOLD time series were then band-pass filtered ($0.01 < f < 0.1$ Hz) (Whitfield-Gabrieli and Nieto-Castanon 2012).

To analyze resting-state functional connectivity, we applied the "seed-to-voxel" approach. Four spheres (radius 10 mm) were centered on connectivity peaks reported by Fox et al. (2005). These 4 peaks were localized in the MPFC, the PCC, as well as in the right and left lateral parietal cortices (RLP and LLP, respectively). Correlation maps were produced by extracting the BOLD time course from a given seed region and computing bivariate correlation coefficients between that time course and the time course from every other voxel in the brain. As a result, we obtained a voxel-wise map of z-score coefficients (Fisher-transformed correlation coefficients) for each seed region and for each participant.

These maps were then entered into the random-effects analysis (one-way ANOVA), with "seed region" as a single factor at 4 levels: MPFC, PCC, RLP, and LLP. Maps of positive functional connectivity for each seed region were computed as linear contrasts, with "1" for a given seed region and "0" for all other seed regions. Finally, DMN was identified using a conservative conjunction procedure; voxels were included in the DMN mask only if they showed significant positive correlations for all 4 seed regions (Fox et al. 2005). The resulting map of DMN functional connectivity ($P < 0.001$ uncorrected; $k_e > 5$ voxels) was superimposed on the mean MRI generated from the standardized anatomical MRIs of the 26 participants, and overlaid with BOLD-adaptation responses selective to conceptual self-awareness.

Results

Behavioral

We first checked whether the masking procedure was effective in manipulating conscious awareness of prime stimuli by analyzing the data from the visibility test, which was performed immediately after the fMRI sessions while the participants were still laying inside the scanner. To that end, we calculated sensitivity indexes, d' (Wickens 2002), separately for masked (mean = 0.5; standard error of the mean [SEM] = 0.22; median = 0.48; first quartile = 0; third quartile = 1.25) and unmasked primes (mean = 4.5; SEM = 0.52; median = 3.46; first quartile = 2.48; third quartile = 6.03) (Fig. 1B). We found that the d' for unmasked primes was significantly higher than chance level ($P < 0.0005$; one-sample Wilcoxon signed-rank test), and so was the d' for masked primes ($t_{25} = 2.24$; $P = 0.034$; one-sample t-test, two-tailed). We also found that participants were significantly worse at discriminating masked than unmasked primes ($P < 0.0005$; related-samples Wilcoxon signed-rank test). Taken together, these results suggest that our procedure was highly effective in manipulating conscious awareness, even if we cannot prove that the masked primes were processed completely unconsciously based on the very conservative criterion of chance-level behavior (i.e., d' being not significantly higher than chance level).

It should be emphasized that with respect to the aims of the present study, it is most important to demonstrate that priming in the SU condition was unconscious; the status of priming in the OU condition is less relevant because our main goal was to dissociate conceptual self-awareness from perceptual awareness and unaware self-processing. By definition, d' is a joint measure of discrimination sensitivity between 2 types of stimuli (Wickens 2002), whereas hit rates can be extracted from one type of stimulus alone. Thus, to provide more direct evidence on priming in the SU condition, we analyzed the hit rate for self-related masked primes (again using data from the visibility session). We found that this hit rate was very close to chance level (mean = 48.9%; SEM ±4.8%; Supplementary Fig. 1A) and that it did not differ significantly from chance level ($t_{25} = -0.23$; $P = 0.82$; one-sample t-test, two-tailed). This result suggests that correct recognition of masked self-related primes in a forced-choice visibility task was very difficult and, therefore, that priming in the SU condition during the fMRI-BOLD-adaptation session was largely independent of conscious awareness.

To further corroborate that the priming effect in SU was unrelated to conscious awareness, we ran an additional regression analysis with the degree of priming in the SU condition as an outcome variable (BOLD-adaptation session) and hit rate for masked self-related primes as a predictor (visibility test). This predictor was centered on the 50% value (chance level) to make the intercept meaningful with respect to unconscious processing. The outcome and the predictor variables consisted of single average values calculated for each participant. We reasoned that if the intercept in this model was significantly above zero, this would suggest that there was significant priming in the SU condition even when recognition accuracy was exactly at chance level; such a result would demonstrate that priming in the SU condition cannot be explained by conscious perception of self-related primes (Greenwald et al. 1996). Indeed, we found that the intercept was above zero ($b_0 = 32.3$ ms; SEM ± 6.6 ms; $t_{25} = 4.9$; $P < 0.005$; Supplementary Fig. 1B), which, together with the results presented above, suggests that the SU condition was associated with genuinely unconscious self-related information processing.

We next turned to behavioral data collected specifically during the fMRI-BOLD-adaptation sessions. Accuracy of responses was very high; it ranged between 95% and 98% for target stimuli

in all trial types. The reaction time (RT) data analysis began by calculating priming coefficients ($RT_{incongruent} - RT_{congruent}$) separately for each experimental condition: SU, OU, SA, and OA. To test whether stimuli repetition was related to an expected facilitation in RTs (i.e., behavioral priming), we performed a series of one-sample t-tests (two-tailed) against zero, separately for each condition. We found significant priming effects for the SU (mean = 31.6 ms; SEM = 7 ms; t_{25} = 4.49; $P < 0.0005$), the SA (mean = 99 ms; SEM = 10 ms; t_{25} = 9.59; $P < 0.0005$), and the OA (mean = 60 ms; SEM = 9 ms; t_{25} = 6.79; $P < 0.0005$) conditions, but not for the OU condition (mean = 0 ms; SEM = 6 ms; t_{25} = 0.01; $P = 0.99$). To test whether the degree of priming differed between conditions, priming coefficients from each participant were entered into the repeated-measures ANOVA with "awareness" (aware vs. unaware) and "person" (self vs. other) as the factors. This analysis showed the main effect of "awareness" ($F_{1,25}$ = 55.36; $P < 0.0005$; η_p^2 = 0.69) and the main effect of "person" ($F_{1,25}$ = 16.16; $P < 0.0005$; η_p^2 = 0.39) but no interaction effect ($F_{1,25}$ = 0.33; $P = 0.57$; η_p^2 = 0.013). Post hoc paired t-tests (two-tailed) showed that priming was 1) stronger ($P < 0.0005$) in the aware context (mean = 80 ms; SEM = 8 ms) than in the unaware context (mean = 16 ms; SEM = 4 ms) and 2) stronger ($P < 0.0005$) for self-related information (mean = 65 ms; SEM = 6 ms) than for other-related information (mean = 30 ms; SEM = 7 ms).

In sum, our behavioral results from the visibility test and scanning sessions confirmed that the experimental manipulations worked as expected. That is, the masking procedure was highly effective in manipulating the level of awareness (Fig. 1B) and the congruency of stimulus pairs (self or other) led to a typical facilitation of RTs (Fig. 1C).

Functional magnetic resonance imaging

We first looked for the neural responses related to conceptual self-awareness, operationalized as brain regions showing higher BOLD adaptation in the self compared with the other condition, specifically in the aware context ("person" × "awareness" interaction). Such a pattern was present in the medial frontopolar prefrontal cortex and retrosplenial cortex in the cingulate gyrus, as well as in the left hippocampus and bilateral temporal poles ($Ps_{corrected} < 0.05$; Figs 2A and 3; Table 1; Supplementary Table 1).

The neural responses related to perceptual awareness, in contrast, were operationalized as regions showing higher BOLD adaptation in the aware than in the unaware context, regardless of self- or other-related content (main effect of "awareness"). This pattern was found in the lateral fronto-parietal cortex, most notably in the left supramarginal gyrus and cortices lining the left intraparietal and the left precentral sulci; such responses were also present in the left fusiform gyrus and the superior frontal gyrus ($Ps_{corrected} < 0.05$; Fig. 2B; Table 1; Supplementary Table 1).

Next, we investigated neural responses that reflected unconscious self-specific processes. Such processes should be engaged irrespectively of whether the participant becomes aware of the stimulus or not, and hence correspond to the main effect of "self" in our factorial design. We found this pattern of responses in the anterior part of the left inferior temporal sulcus ($x = -45$; $y = -7$; $z = -29$; $P_{corrected} < 0.05$; Fig. 2C; Table 1; Supplementary Table 1). Unaware self-processing was also related to more widespread neural responses in the inferior temporal and occipital cortices (Fig. 2C; Supplementary Table 1); however, these clusters did not surpass the correction
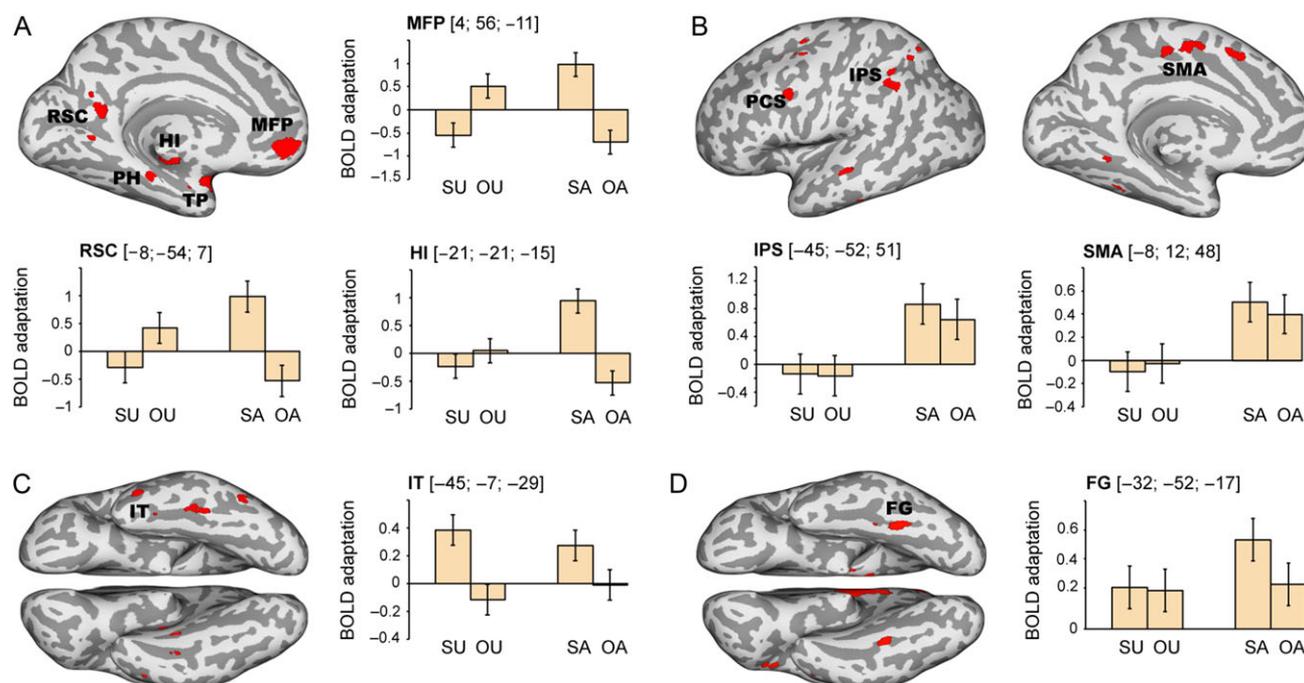
Figure 2. Results of fMRI analysis. Brain regions related to (A) conceptual self-awareness, (B) perceptual awareness, (C) unaware processing of self-related information, and (D) unaware processing of all types of stimuli. Areas showing significant BOLD-adaptation effects are highlighted in red (peaks $P < 0.05$ corrected; cluster maps thresholded at $P < 0.001$, uncorrected for display purposes) and superimposed on an "inflated" MNI template brain. The bar plots show the BOLD-adaptation effect size (incongruent > congruent trials) and they are added for purely descriptive purposes; error bars denote standard errors. FG, fusiform gyrus; HI, hippocampus; IPS, intraparietal sulcus; IT, inferior temporal cortex; MFP, medial frontal pole; PCS, precentral sulcus; PH, parahippocampal gyrus; RSC, retrosplenial cortex; SMA, supplementary motor area; TP, temporal pole.
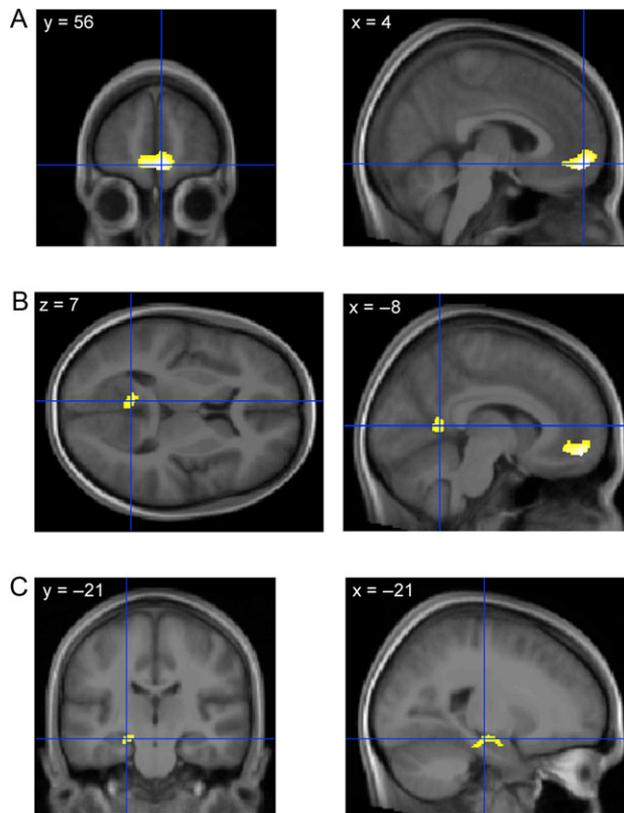
Figure 3. Anatomical localization of BOLD-adaptation responses related to conceptual self-awareness. The peak activations from the interaction contrast (also shown in Fig. 2A) are here superimposed on a mean MRI generated from the standardized anatomical MRIs of the 26 participants. As can be seen, the peaks (indicted by the crossing of the blue lines) are located in (A) the medial frontopolar prefrontal cortex, (B) the hippocampus, and (C) the retrosplenial cortex. Voxels showing a significant BOLD-adaptation effect are highlighted in yellow (peaks $P < 0.05$ corrected; cluster maps thresholded at $P < 0.001$, uncorrected for display purposes).

for multiple comparisons in the whole-brain space and they lay outside our a priori defined small volumes of interest.

In a complementary descriptive approach, we also examined the simple SU − OU contrast. We observed a BOLD-adaptation peak in the exact same area of the inferior temporal sulcus ($x = -45$; $y = -7$; $z = -29$; $P_{uncorrected} = 0.001$) as we observed for the main effect of "self" (Supplementary Table 1). The same result obtained in a different analysis further supports that the anterior part of the left inferior temporal cortex was involved in the unaware processing of self-related visual information.

Finally, we identified brain responses related to unaware processing of all types of stimuli, that is, brain regions where the degree of BOLD adaptation was similar in all conditions and thus independent of self-relevance or awareness. This pattern was found in the posterior part of the left fusiform gyrus in the occipital lobe and in the middle part of the cingulate gyrus ($Ps_{corrected} < 0.05$; Fig. 2D; Table 1; Supplementary Table 1). The occipital region is likely to be related to basic perceptual analysis of the visual stimulus, which occurs before conscious access (Dehaene et al. 2006).

Our additional analysis of the resting-state data showed that BOLD-adaptation effects associated with conceptual self-awareness considerably overlapped with DMN in the medial frontopolar cortex and in the retrosplenial cortex (Fig. 4). These
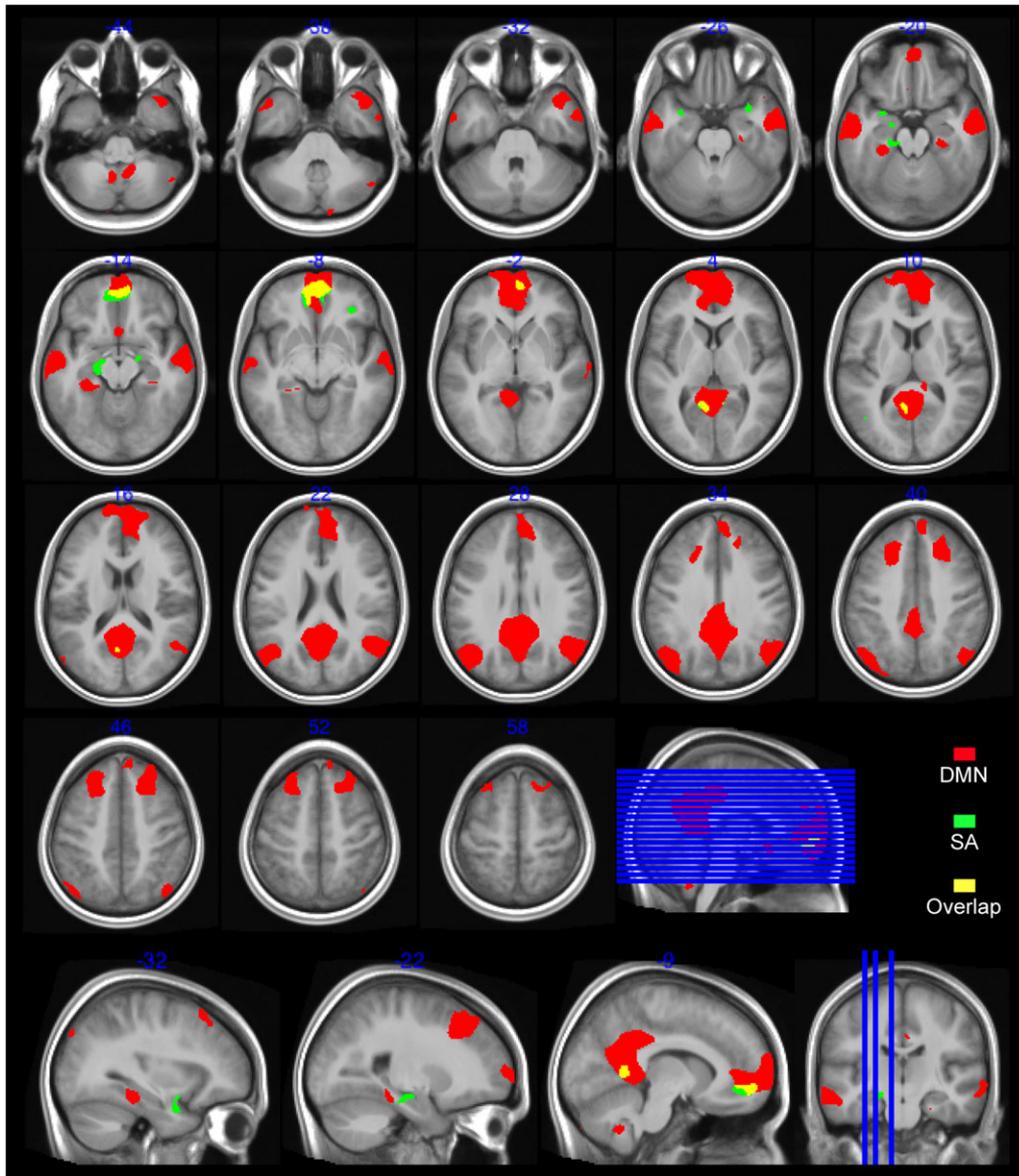
results suggest that the key nodes of DMN are shared with the "conceptual self-awareness network."

Discussion

The aim of this study was to dissociate the neural correlates of conceptual self-awareness from the neural correlates of perceptual awareness and unaware self-processing, using controlled procedures to manipulate the conscious awareness of self- versus other-related stimuli. Our main finding is that the aware processing of self-related stimuli involves selective neural responses in the medial frontopolar prefrontal cortex, retrosplenial cortex, and medial temporal cortices.

The observation that medial frontopolar cortex contains a conscious representation of self-relevant conceptual information is especially interesting. This region corresponds to the most anterior part of the prefrontal cortex (area 10), which is an area that is particularly developed in humans (Semendeferi et al. 2001). Although little is known about its cognitive functions, neuroimaging studies have associated activation of the frontopolar prefrontal cortex to the coordination of internal and external information processing (Burgess et al. 2007), the combination of 2 or more cognitive processes to guide complex behaviors (Koechlin et al. 1999; Ramnani and Owen 2004), and counterfactual and exploratory decision-making (Rushworth et al. 2011). More specifically, the medial part of the frontal pole (frontopolar area 2; Bludau et al. 2013) has been implicated in social cognitive and self-referential tasks, such as assigning personal significance to self-related contents (D'Argembeau et al. 2012), evaluation of self-generated decisions (Tsujimoto et al. 2010), episodic memory recall (Daselaar et al. 2008), and "theory of mind" (Amodio and Frith 2006); this area is also part of DMN (see further below). Our data demonstrating that neuronal populations in the medial frontopolar cortex show selectivity for self-relevant conceptual information in the context of awareness could explain the involvement of this region in a wide range of social cognitive and self-referential paradigms; tasks that all heavily depend on conscious access to self-concept. We speculate that in the course of human evolution a rich conceptual representation of the self developed in the frontopolar prefrontal cortex in response to increasing demands to generate complex social behaviors, requiring the abilities to consciously self-reflect, such as when we recall and reflect upon important life events from our past or when we contemplate about our future at the time of critical behavioral decisions.

Conceptual self-awareness in our fMRI data was also associated with selective neural representations in the retrosplenial cortex, the hippocampus, and the temporal pole cortex. The hippocampus and the temporal pole are anatomically connected with the medial frontopolar prefrontal cortex (Neubert et al. 2015), and these temporal regions are also directly connected with the retrosplenial cortex in the posterior cingulate gyrus (Kravitz et al. 2011). Thus, these structures are likely to work in concert to support awareness of the conceptual self. The retrosplenial cortex and the hippocampus have previously been associated with cognitive functions such as autobiographical episodic memory (Nyberg et al. 1996; Bergouignan et al. 2014), declarative memory (Squire and Zola 1996; Eichenbaum 2000), spatial memory (Bird and Burgess 2008), spatial navigation (Maguire et al. 1998; Doeller et al. 2010), sense of bodily self-location (Guterstam et al. 2015b); and the temporal polar cortex has been implicated in semantic memory and social cognitive processes (Maguire et al. 2000; Olson et al. 2007). The common denominator in these tasks is the self-centric nature

Figure 4. Overlap between the "conceptual self-awareness network" and the "DMN". Neural responses selective for conceptual self-awareness (SA; marked in green) are based on the BOLD-adaptation analysis. A map of DMN's functional connectivity (marked in red) is based on resting-state data that were collected separately for the same group of participants. The overlap is present especially in the medial prefrontal and posterior cingulate regions (marked in yellow). The cluster map thresholds are P < 0.001 (uncorrected) and are superimposed on axial and sagittal slices of a mean MRI generated from the standardized anatomical MRIs of the 26 participants.

of perceptual, spatial, or semantic information that builds up multilevel and multisensory constructs of the self. The fact that areas previously implicated in declarative and autobiographical memory contain neuronal populations with high selectivity for aware processing of self-relevant conceptual information is consistent with the fact that people have more conceptual and factual knowledge about themselves than about any other person, and that this kind of memory requires conscious awareness for recall and verbal report.

The conceptual self-awareness network from our study overlapped with DMN (Raichle et al. 2001; Buckner et al. 2008) in the medial frontopolar and retrosplenial cortices (Fig. 4). The link between DMN and self-processing has been proposed previously (Gusnard et al. 2001; D'Argembeau et al. 2005; Buckner et al. 2008; Qin and Northoff 2011; Whitfield-Gabrieli

et al. 2011); however, experimental evidence, as is presented here, where awareness is explicitly manipulated, has been lacking. DMN shows strong functional connectivity during "rest," a situation in which participants are likely to be engaged in introspective thoughts. But it is unclear to what extent activity seen during an uncontrolled condition such as "rest" reflects self-consciousness per se (Konishi and Smallwood 2016). Furthermore, it has been shown that the DMN's resting-state connectivity is preserved even in the absence of self-awareness, for example, in monkeys under general anesthesia (Vincent et al. 2007). Moreover, the involvement of DMN regions has been linked to domains that are not directly related to the awareness of self-related conceptual information, for example, spatial navigation (Doeller et al. 2010) or the sense of bodily self-location (Guterstam

et al. 2015b). Thus, the present results should be particularly valuable for the ongoing discussion about the link between DMN and the self (Davey et al. 2016; Northoff 2016). From a broader perspective, the existing evidence suggests that DMN is a multifaceted system that is involved in a variety of functions; our study suggests that key nodes of DMN are not associated with the processing of self-relevant content generally but instead are associated selectively with the conscious processing of self-relevant information.

In the anterior part of the inferior temporal cortex, we identified self-selective responses that did not depend on awareness (Fig. 2C). This result suggests that self-related information is processed preferentially already at early stages of information processing. Our interpretation is that, owing to the high salience and familiarity of self-related stimuli (Sui et al. 2012; Sui and Humphreys 2015), the brain automatically detects these stimuli even before they enter awareness. The sound of one's own name, for example, usually signals an important event, for example, a warning, a reward, or a social interaction. Thus, through associative learning, one's own name begins to be prioritized over less behaviorally relevant stimuli. As this mechanism is strengthened by life-long experience, the reaction to one's own name becomes fast, involuntary, and unaware (Bargh 1982; Schneider and Chein 2003). It has been shown that the processing of biologically important stimuli (e.g., faces, emotional stimuli) involves highly specialized brain modules and that this processing does not require awareness (Whalen et al. 1998; Kouider et al. 2009). We believe that certain aspects of self-related processing involve similar specialized and unaware mechanisms. Such unconscious self-biases could play an important role in implicit social cognition, as with ingroup–outgroup discrimination or stereotypes (Greenwald et al. 2002). To the best of our knowledge, our findings provide the first evidence that the neural representation of self-related information has 2 levels. One is the higher level, which is accessed consciously and involves the medial frontopolar prefrontal, medial temporal, and retrosplenial cortices; and the second one is the lower level, which is automatically accessed without awareness and involves more posterior sensory areas.

Furthermore, we found a clear distinction between BOLD-adaptation responses in the "conceptual self-awareness network" and the "perceptual awareness network." As mentioned earlier, the "Global Neuronal Workspace" theory proposes that consciousness is associated with the broadcasting of information within a large network of fronto-parietal regions (Dehaene et al. 2006; Dehaene and Changeux 2011). Our findings are generally consistent with this view; importantly, however, they go further to indicate that conceptual self-awareness and perceptual awareness correspond to very different states of this workspace, or even that these 2 types of awareness engage separate workspaces (Fox et al. 2005; Boly et al. 2007; Vanhaudenhuyse et al. 2011). The neuroanatomical dissociation that we observed could explain a number of functional dissociations, for example, that perceptual awareness and conceptual self-awareness emerge at different times in the child's development (Zelazo 2004); that being focused on the self often makes people less attentive to the external environment, and vice versa (Dixon et al. 2014); or that in disorders like autism, self-awareness is more impaired than perceptual awareness (Frith and Happé 1994). In general, our results show that conceptual self-awareness and perceptual awareness have different neural correlates, suggesting that these 2 types of consciousness are partly independent.

We also found that the SA condition was associated with increased BOLD adaptation both in the medial prefronto-cingulate-temporal network and in the lateral fronto-parietal network (Fig. 2A,B); it was the pattern of responses in the remaining conditions that distinguished between activity profiles of these 2 networks rather than the degree of SA-related BOLD adaptation per se. In other words, conceptual self-awareness in our paradigm engaged neural populations in both the lateral and the medial networks while perceptual awareness only engaged the lateral network. This makes sense because, in order to become aware of the conceptual meaning of autobiographical stimuli (names, date of birth, nationality codes), one must first be able to consciously see the presented letters and words, which requires perceptual awareness. In apparent contrast to this finding, previous resting-state studies reported an anticorrelated pattern of activity between these medial and lateral networks, which was interpreted in terms of an ongoing competition between internally and externally oriented mental processes over limited capacity of cognitive resources (Fox et al. 2005; Fransson 2005; Boly et al. 2007; Vanhaudenhuyse et al. 2011). The current findings show an important exception from this competition-based interpretation and suggest that the "DMN" and the "task-positive network" (TPN) could be activated simultaneously in situations when conceptual self-awareness is triggered by external self-relevant stimuli (for additional examples of situations in which the competition between DMN and TPN is diminished, see Wotruba et al. 2013; Dixon et al. 2014; Spreng et al. 2014; Konishi et al. 2015; Vatansever et al. 2015; Krieger-Redwood et al. 2016).

Additionally, our findings are relevant to the ongoing debate about the role of familiarity in the preferential processing of self-relevant information. Self-related information is by its very nature familiar; and therefore, a general concern in this literature is that effects attributed to self-processing might in fact be explained simply by familiarity (Qin et al. 2012; Sui et al. 2012; Tacikowski et al. 2013; Humphreys and Sui 2016). Our results are more protected against this concern because in our factorial design aware self-processing was always compared with another "self" condition, when contrasting the masked and unmasked trials, and not only to the unfamiliar "other" condition. This means that familiarity was matched in the key contrast identifying BOLD-adaptation responses selective for conscious processing of self-related conceptual information. However, according to this line of reasoning we cannot rule out that familiarity could have contributed to neural responses in the inferior temporal cortex reflecting the main effect of "self." More studies are needed to see whether it is possible to further disambiguate neural responses that reflect self-relevance and semantic familiarity. In the meantime, our findings clearly indicate that either way, conscious processing of self-relevant conceptual information is associated with selective neural responses in the medial frontopolar, posterior cingulate, and medial temporal brain areas.

When looking at the bar plots in Figure 2A, a critical reader might ask whether the interaction effect in the medial frontopolar region could not have been driven, at least in part, by a reversed BOLD-adaptation effect or "repetition enhancement" (Segaert et al. 2013) present in the SU and OA conditions. However, neither the SU < 0 nor the OA < 0 contrasts showed any significant activations in the "conceptual self-awareness network" ($P_{uncorrected} < 0.001$) when we examined this issue in a purely descriptive post hoc analysis. In turn, the SA > 0 contrast overlapped with the "conceptual self-awareness network"

in all key brain regions ($P_{uncorrected} < 0.001$). Thus, there is no evidence in our data to support the notion that the interaction effect was driven by repetition enhancement in the SU or OA conditions, but there is evidence suggesting that this interaction was driven by the actual BOLD-adaptation effect in the SA condition. It should also be noted that the statistical inference in this study is made at the level of the interaction term and that no further conclusions should be based on the plotted effect sizes, which were included for purely descriptive purposes. It is well known that in fast event-related designs without null events, such as our design, there might be high variability from baseline associated with the estimate of each event type on its own, even if there is low variability for the estimate of the differences (Henson 2006); we are basing our statistical inferences upon the latter. Thus, the bar plots in Figure 2 should be interpreted with great caution; as was noted, they were added for illustrative purposes only.

With regard to generalization of our findings, self-consciousness is a broad term that is used to describe a pre-reflective feeling of being a subject of conscious experience ("I"), as well as a mental state in which the content of one's consciousness refers to own behavior, mental states, or self-knowledge ("me"). Being an "active perceiver" is intimately linked to the body; felt as one's own, occupying a given location in space, seen from the first person perspective, and experienced through interoception. The feeling of ownership of one's body has been related to mechanisms of multisensory integration in the premotor and posterior parietal cortices (Ehrsson et al. 2004; Petkova et al. 2011; Ehrsson 2012; Blanke et al. 2015), the sense of bodily self-location to the PCC (Guterstam et al. 2015a, 2015b), and bodily self-consciousness to the temporo-parietal junction (Ionta et al. 2011; Blanke 2012); although these studies did not explicitly manipulate the level of awareness. Furthermore, interoceptive awareness has been linked to neural responses in the insular cortex (Critchley et al. 2004; Craig 2011). Both in terms of operationalization and principal findings, our study clearly diverges from these previous studies investigating bodily-spatial experiences of the self; we focused on higher order cognitive aspects of self-awareness. The location of neural responses showing selectivity for conceptual self-awareness in the present study seems to overlap with brain activity observed during autobiographical memory retrieval (Cabeza and St Jacques 2007), self-reflection (Johnson et al. 2002; Murray et al. 2012), and meta-awareness (Schooler et al. 2011); the latter is a special case of self-awareness, where the aspect of self that one becomes aware of is one's own mental states (Schooler 2002). Taken together, the neural underpinnings of self-consciousness seem to have at least 2 basic features, the bodily-spatial one and the cognitive-conceptual one, that generally correspond to the two main components of the human self characterized by William James; the "I" and the "me" (James 1890; Damasio 2010).

In sum, using the BOLD-adaptation approach and the visual masking paradigm, for the first time we could pinpoint the neural basis of conceptual self-awareness to the medial frontopolar-retrosplenial areas and show that it is neuroanatomically distinct from the lateral fronto-parietal network that mediates perceptual awareness and from the inferior temporal cortex related to the unaware processing of self-relevant information. These novel findings suggest that 1) the neural representation of the human self-concept has a hierarchical structure, with separate conscious and unconscious components; 2) that the brain automatically detects semantic self-related stimuli even before they enter awareness; and 3) that the neural correlates of conscious awareness are not confined to the lateral fronto-parietal network but extend to specific medial prefronto-cingulate-temporal regions when the content of information refers to self-concept. Together, our study opens up a new venue of research aiming to reveal the neurocognitive architecture of the human self-representation and self-awareness using well-controlled behavioral manipulations. Crucially, identifying the neural substrate of self-awareness in the healthy brain has strong bearings on clinical research into psychiatric and neurological disorders, where the sense of self is impaired, for example, in schizophrenia, depression, autism, or in personality disorders.

Supplementary Material

Supplementary material are available at *Cerebral Cortex* online.

Funding

Notes

References

Alexopoulos T, Muller D, Ric F, Marendaz C. 2012. I, me, mine: automatic attentional capture by self-related stimuli. Eur J Soc Psychol. 42:770–779.

Amodio DM, Frith CD. 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci. 7: 268–277.

Bargh JA. 1982. Attention and automaticity in the processing of self-relevant information. J Pers Soc Psychol. 43:425–436.

Baumeister RF. 1998. The self. In: Lindzey G, Gilbert D, Fiske S, editors. The handbook of social psychology. New York: Oxford University Press. p. 680–740.

Bergouignan L, Nyberg L, Ehrsson HH. 2014. Out-of-body-induced hippocampal amnesia. Proc Natl Acad Sci USA. 111: 4421–4426.

Bird CM, Burgess N. 2008. The hippocampus and memory: insights from spatial processing. Nat Rev Neurosci. 9: 182–194.

Blanke O. 2012. Multisensory brain mechanisms of bodily self-consciousness. Nat Rev Neurosci. 13:556–571.

Blanke O, Slater M, Serino A. 2015. Behavioral, neural, and computational principles of bodily self-consciousness. Neuron. 88:145–166.

Bludau S, Eickhoff SB, Mohlberg H, Caspers S, Laird AR, Fox PT, Schleicher A, Zilles K, Amunts K. 2013. Cytoarchitecture, probability maps and functions of the human frontal pole. Neuroimage. 93:260–275.

Boly M, Balteau E, Schnakers C, Degueldre C, Moonen G, Luxen A, Phillips C, Peigneux P, Maquet P, Laureys S. 2007. Baseline brain activity fluctuations predict somatosensory perception in humans. Proc Natl Acad Sci USA. 104:12187–12192.

Buckner RL, Andrews-Hanna JR, Schacter DL. 2008. The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci. 1124:1–38.

Burgess PW, Dumontheil I, Gilbert SJ. 2007. The gateway hypothesis of rostral prefrontal cortex (area 10) function. Trends Cogn Sci. 11:290–298.

Cabeza R, St Jacques P. 2007. Functional neuroimaging of autobiographical memory. Trends Cogn Sci. 11:219–227.

Carver CS. 2012. Self-awareness. In: Leary MR, Tangney JP, editors. Handbook of self and identity. 2nd ed. New York: The Guilford Press.

Churchland P. 2002. Self-representation in nervous systems. Science. 296:308–310.

Craig AD. 2011. Significance of the insula for the evolution of human awareness of feelings from the body. Ann N Y Acad Sci. 1225:72–82.

Crick F, Koch C. 1998. Consciousness and neuroscience. Cereb Cortex. 8:97–107.

Critchley HD, Wiens S, Rotshtein P, Ohman A, Dolan RJ. 2004. Neural systems supporting interoceptive awareness. Nat Neurosci. 7:189–195.

D'Argembeau A, Collette F, Van der Linden M, Laureys S, Del Fiore G, Degueldre C, Luxen A, Salmon E. 2005. Self-referential reflective activity and its relationship with rest: a PET study. Neuroimage. 25:616–624.

D'Argembeau A, Jedidi H, Balteau E, Bahri M, Phillips C, Salmon E. 2012. Valuing one's self: medial prefrontal involvement in epistemic and emotive investments in self-views. Cereb Cortex. 22:659–667.

Damasio AR. 2010. Self comes to mind: constructing the conscious brain. London: William Heinemann.

Daselaar SM, Rice HJ, Greenberg DL, Cabeza R, LaBar KS, Rubin DC. 2008. The spatiotemporal dynamics of autobiographical memory: neural correlates of recall, emotional intensity, and reliving. Cereb Cortex. 18:217–229.

Davey CG, Pujol J, Harrison BJ. 2016. Mapping the self in the brain's default mode network. Neuroimage. 132:390–397.

Dehaene S, Changeux J-P, Naccache L, Sackur J, Sergent C. 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends Cogn Sci. 10:204–211.

Dehaene S, Changeux JP. 2011. Experimental and theoretical approaches to conscious processing. Neuron. 70:200–227.

Dehaene S, Naccache L, Cohen L, Bihan DL, Mangin JF, Poline JB, Rivière D. 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat Neurosci. 4:752–758.

Dixon ML, Fox KCR, Christoff K. 2014. A framework for understanding the relationship between externally and internally directed cognition. Neuropsychologia. 62:321–330.

Doeller CF, Barry C, Burgess N. 2010. Evidence for grid cells in a human memory network. Nature. 463:657–661.

Duval S, Wicklund RA. 1972. A theory of objective self-awareness. Oxford: Academic Press.

Duvernoy HM. 1991. The human brain: surface, blood supply, and three-dimensional sectional anatomy. New York: Springer.

Ehrsson HH. 2007. The experimental induction of out-of-body experiences. Science. 317:1048.

Ehrsson HH. 2012. The concept of body ownership and its relation to multisensory integration. In: Stein BE, editor. The new handbook of multisensory processes. Cambridge (MA): MIT Press. p. 775–792.

Ehrsson HH, Spence C, Passingham RE. 2004. That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. Science. 305:875–877.

Eichenbaum H. 2000. A cortical-hippocampal system for declarative memory. Nat Rev Neurosci. 1:41–50.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci USA. 102:9673–9678.

Fransson P. 2005. Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis. Hum Brain Mapp. 26: 15–29.

Frith U, Happé F. 1994. Autism: beyond "theory of mind". Cognition. 50:115–132.

Gallup GG. 1997. On the rise and fall of self-conception in primates. Ann N Y Acad Sci. 818:73–82.

Gray HM, Ambady N, Lowenthal WT, Deldin P. 2004. P300 as an index of attention to self-relevant stimuli. J Exp Soc Psychol. 40:216–224.

Greenwald AG, Banaji M, Rudman LA, Farnham SD, Nosek B, Mellott DS. 2002. A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. Psychol Rev. 109: 3–25.

Greenwald AG, Draine SC, Abrams RL. 1996. Three cognitive markers of unconscious semantic activation. Science. 273: 1699–1702.

Greicius MD, Supekar K, Menon V, Dougherty RF. 2009. Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb Cortex. 19: 72–78.

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci. 10:14–23.

Gusnard DA, Akbudak E, Shulman GL, Raichle ME. 2001. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. Proc Natl Acad Sci USA. 98:4259–4264.

Guterstam A, Björnsdotter M, Bergouignan L, Gentile G, Li TQ, Ehrsson HH. 2015a. Decoding illusory self-location from activity in the human hippocampus. Front Hum Neurosci. 9:412.

Guterstam A, Björnsdotter M, Gentile G, Ehrsson H. 2015b. Posterior cingulate cortex integrates the senses of self-location and body ownership. Curr Biol. 25:1416–1425.

Henson R. 2006. Efficient experimental design for fMRI. In: Friston K, Ashburner J, Kiebel S, Nichols T, Penny W, editors. Statistical parametric mapping: the analysis of functional brain images. London: Elsevier. p. 193–210.

Humphreys GW, Sui J. 2016. Attentional control and the self: the Self Attention Network (SAN). Cogn Neurosci. 7:5–17.

Ionta S, Heydrich L, Lenggenhager B, Mouthon M, Fornari E, Chapuis D, Gassert R, Blanke O. 2011. Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. Neuron. 70:363–374.

James W. 1890. The principles of psychology. New York: Halt.

Johnson SC, Baxter LC, Wilder LS, Pipe JG, Heiserman JE, Prigatano GP. 2002. Neural correlates of self-reflection. Brain. 125:1808–1814.

Koch C, Massimini M, Boly M, Tononi G. 2016. The neural correlates of consciousness: progress and problems. Nat Rev Neurosci. 307–321.

Koechlin E, Basso G, Pietrini P, Panzer S, Grafman JH. 1999. The role of the anterior prefrontal cortex in human cognition. Nature. 399:148–151.

Konishi M, McLaren DG, Engen H, Smallwood J. 2015. Shaped by the past: The default mode network supports cognition that is independent of immediate perceptual input. PLoS One. 10: e0132209.

Konishi M, Smallwood, J. 2016. Shadowing the wandering mind: how understanding the mind-wandering state can inform our appreciation of conscious experienceWiley Interdiscip Rev Cogn Sci. 7:4233–246.

Kouider S, de Gardelle V, Dehaene S, Dupoux E, Pallier C. 2010. Cerebral bases of subliminal speech priming. Neuroimage. 49:922–929.

Kouider S, Eger E, Dolan R, Henson RN. 2009. Activity in face-responsive brain regions is modulated by invisible, attended faces: evidence from masked priming. Cereb Cortex. 19: 13–23.

Kravitz DJ, Saleem KS, Baker CI, Mishkin M. 2011. A new neural framework for visuospatial processing. Nat Rev Neurosci. 12:217–230.

Krieger-Redwood K, Jefferies E, Karapanagiotidis T, Seymour R, Nunes A, Ang JWA, Majernikova V, Mollo G, Smallwood J. 2016. Down but not out in posterior cingulate cortex: deactivation yet functional coupling with prefrontal cortex during demanding semantic cognition. Neuroimage. 141:366–377.

Li L, Miller EK, Desimone R. 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. J Neurophysiol. 69:1918–1929.

Lou HC, Changeux JP, Rosenstand A. Forthcoming 2016. Towards a cognitive neuroscience of self-awareness. Neurosci Biobehav Rev. doi: 10.1016/j.neubiorev.2016.04.004.

Maguire EA, Burgess N, Donnett JG, Frackowiak RS, Frith CD, O'Keefe J. 1998. Knowing where and getting there: a human navigation network. Science. 280:921–924.

Maguire EA, Mummery CJ, Büchel C. 2000. Patterns of hippocampal-cortical interaction dissociate temporal lobe memory subsystems. Hippocampus. 10:475–482.

Martinelli P, Sperduti M, Piolino P. 2013. Neural substrates of the self-memory system: new insights from a meta-analysis. Hum Brain Mapp. 34:1515–1529.

Mitchell JP, Macrae CN, Banaji MR. 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron. 50:655–663.

Morin A. 2006. Levels of consciousness and self-awareness: a comparison and integration of various neurocognitive views. Conscious Cogn. 15:358–371.

Murray RJ, Schaer M, Debbane M, Debbané M. 2012. Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. Neurosci Biobehav Rev. 36:1043–1059.

Nakamura K, Dehaene S, Jobert A, Le Bihan D, Kouider S. 2007. Task-specific change of unconscious neural priming in the cerebral language network. Proc Natl Acad Sci USA. 104: 19643–19648.

Neubert F-X, Mars RB, Sallet J, Rushworth MFS. 2015. Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. Proc Natl Acad Sci. 112:E2695–704.

Northoff G. 2016. Is the self a higher-order or fundamental function of the brain? The "basis model of self-specificity" and its encoding by the brain's spontaneous activity. Cogn Neurosci. 7:203–222.

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. 31:440–457.

Nyberg L, McIntosh AR, Houle S, Nilsson L-G, Tulving E. 1996. Activation of medial temporal structures during episodic memory retrieval. Nature. 380:715–717.

Olson IR, Plotzker A, Ezzyat Y. 2007. The enigmatic temporal pole: a review of findings on social and emotional processing. Brain. 130:1718–1731.

Oyserman D, Elmore K, Smith G. 2012. Self, self-concept, and identity. In: Leary MR, Tangney JP, editors. Handbook of self and identity. 2nd ed. New York: The Guilford Press. p. 69–104.

Petkova VI, Bjornsdotter M, Gentile G, Jonsson T, Li TQ, Ehrsson HH. 2011. From part- to whole-body ownership in the multisensory brain. Curr Biol. 21:1118–1122.

Pfister R, Pohl C, Kiesel A, Kunde W. 2012. Your unconscious knows your name. PLoS One. 7:e32402.

Qin P, Liu Y, Shi J, Wang Y, Duncan N, Gong Q, Weng X, Northoff G. 2012. Dissociation between anterior and posterior cortical regions during self-specificity and familiarity: a combined fMRI-meta-analytic study. Hum Brain Mapp. 33: 154–164.

Qin P, Northoff G. 2011. How is our self related to midline regions and the default-mode network? Neuroimage. 57: 1221–1233.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. 2001. A default mode of brain function. Proc Natl Acad Sci USA. 98:676–682.

Ramnani N, Owen AM. 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat Rev Neurosci. 5:184–194.

Rees G, Kreiman G, Koch C. 2002. Neural correlates of consciousness in humans. Nat Rev Neurosci. 3:261–270.

Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE. 2011. Frontal cortex and reward-guided learning and decision-making. Neuron. 70:1054–1069.

Schacter DL, Chiao JY, Mitchell JP. 2003. The seven sins of memory: implications for self. Ann N Y Acad Sci. 1001:226–239.

Schneider W, Chein JM. 2003. Controlled & automatic processing: behavior, theory, and biological mechanisms. Cogn Sci. 27:525–559.

Schooler JW. 2002. Re-representing consciousness: dissociations between experience and meta-consciousness. Trends Cogn Sci. 6:339–344.

Schooler JW, Smallwood J, Christoff K, Handy TC, Reichle ED, Sayette MA. 2011. Meta-awareness, perceptual decoupling and the wandering mind. Trends Cogn Sci. 15:319–326.

Segaert K, Weber K, de Lange FP, Petersson KM, Hagoort P. 2013. The suppression of repetition enhancement: a review of fMRI studies. Neuropsychologia. 51:59–66.

Semendeferi K, Armstrong E, Schleicher A, Zilles K, Van Hoesen GW. 2001. Prefrontal cortex in humans and apes: a comparative study of area 10. Am J Phys Anthropol. 114: 224–241.

Spreng RN, DuPre E, Selarka D, Garcia J, Gojkovic S, Mildner J, Luh WM, Turner GR. 2014. Goal-congruent default network activity facilitates cognitive control. J Neurosci. 34: 14108–14114.

Squire LR, Zola SM. 1996. Structure and function of declarative and nondeclarative memory systems. Proc Natl Acad Sci USA. 93:13515–13522.

Sugiura M, Sassa Y, Jeong H, Horie K, Sato S, Kawashima R. 2008. Face-specific and domain-general characteristics of cortical responses during self-recognition. Neuroimage. 42: 414–422.

Sui J, He X, Humphreys GW. 2012. Perceptual effects of social salience: evidence from self-prioritization effects on perceptual matching. J Exp Psychol Hum Percept Perform. 38: 1105–1117.

Sui J, Humphreys GW. 2015. The integrative self: how self-reference integrates perception and memory. Trends Cogn Sci. 19:719–728.

Tacikowski P, Brechmann A, Nowicka A. 2013. Cross-modal pattern of brain activations associated with the processing of self- and significant other's name. Hum Brain Mapp. 34:2069–2077.

Tononi G, Boly M, Massimini M, Koch C. 2016. Integrated Information Theory: from consciousness to its physical substrates. Nat Rev Neurosci. 17:450–461.

Tsujimoto S, Genovesio A, Wise SP. 2010. Evaluating self-generated decisions in frontal pole cortex of monkeys. Nat Neurosci. 13:120–126.

Vanhaudenhuyse A, Demertzi A, Schabus M, Noirhomme Q, Bredart S, Boly M, Phillips C, Soddu A, Luxen A, Moonen G, et al. 2011. Two distinct neuronal networks mediate the awareness of environment and of self. J Cogn Neurosci. 23:570–578.

Vatansever D, Menon XDK, Manktelow AE, Sahakian BJ, Stamatakis EA. 2015. Default mode dynamics for global functional integration. J Neurosci. 35:15254–15262.

Vincent JL, Patel GH, Fox MD, Snyder AZ, Baker JT, Van Essen DC, Zempel JM, Snyder LH, Corbetta M, Raichle ME. 2007. Intrinsic functional architecture in the anaesthetized monkey brain. Nature. 447:83–86.

Watson RS. 1986. The named and the nameless: gender and person in Chinese society. Am Ethnol. 13:619–631.

Whalen PJ, Rauch SL, Etcoff NL, McInerney SC, Lee MB, Jenike MA. 1998. Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J Neurosci. 18:411–418.

Whitfield-Gabrieli S, Moran JM, Nieto-Castañón A, Triantafyllou C, Saxe R, Gabrieli JDE. 2011. Associations and dissociations between default and self-reference networks in the human brain. Neuroimage. 55:225–232.

Whitfield-Gabrieli S, Nieto-Castanon A. 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect. 2:125–141.

Wickens TD. 2002. Elementary signal detection theory. New York: Oxford University Press.

Wotruba D, Michels L, Buechler R, Metzler S, Theodoridou A, Gerstenberg M, Walitza S, Kollias S, Rossler W, Heekeren K. 2013. Aberrant coupling within and across the default mode, task-positive, and salience network in subjects at risk for psychosis. Schizophr Bull. 40:1095–1104.

Zelazo PD. 2004. The development of conscious control in childhood. Trends Cogn Sci. 8:12–17.

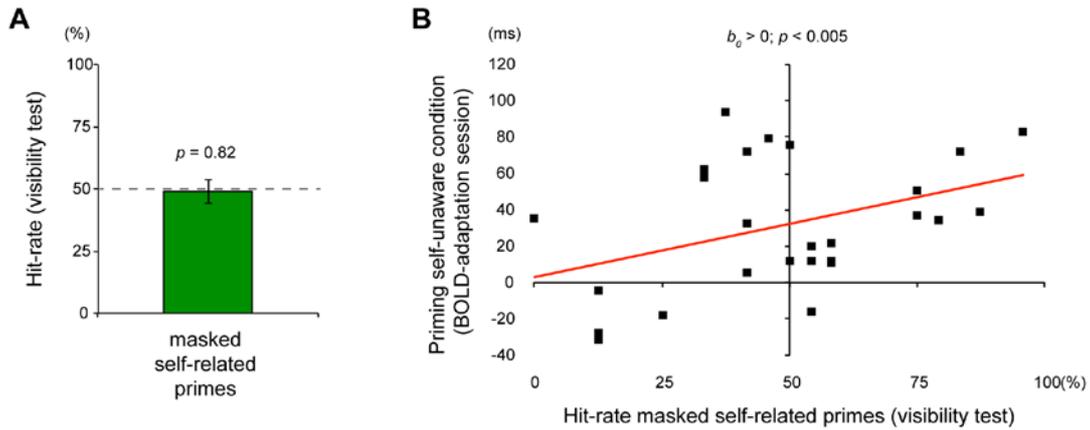Zeman A. 2001. Consciousness. Brain. 124:1263–1289.

Title: "Dissociating the neural basis of conceptual self-awareness from perceptual awareness and unaware self-processing"

Authors: P. Tacikowski, C.C. Berger, and H.H. Ehrsson

Supplementary Online Materials (SOM):

    Supplementary Figure S1

    Supplementary Table S1

Supplementary Figure S1. Additional evidence that the SU condition was associated with unconscious processing. (**A**) In the forced-choice visibility test recognition of masked self-related primes was very close to chance-level (dashed line) and it did not differ significantly from chance-level. The bar plot represents the mean hit-rate from the whole group of 26 participants ± SEM. (**B**) We used simple linear regression to predict the degree of priming in the self-unaware condition during fMRI-BOLD-adaptation session (y-axis) based on hit-rates for masked self-related primes in the visibility test (x-axis). Each dot represents one participant and the line represents the best fit using the least squares approach. We found that the model's intercept was significantly above zero, which indicates that priming was present even when visibility was at chance-level. This result suggests that priming in the SU condition cannot be explained simply by the conscious perception of self-related primes.

Supplementary Table S1. Additional descriptive fMRI results (for significant results see Table 1).

	x	y	z	t-val.	vox.	p-val.
Conceptual self-awareness						
R. posterior orbital gyrus	33	36	-8	3.78	69	0.0001
R. hippocampus	20	-9	-11	3.40	40	0.0005
Perceptual awareness						
R. calcarine sulcus (striate cortex)	24	-68	7	4.53	90	0.0000
R. superior frontal gyrus (pre-SMA)	17	14	48	4.19	59	0.0000
L. lingual sulcus	-26	-53	-2	4.03	23	0.0001
L. superior temporal sulcus	-48	-32	-9	3.92	166	0.0001
R. supramarginal gyrus	68	-42	27	3.73	19	0.0002
R. superior frontal gyrus	29	-15	54	3.69	31	0.0002
L. cerebellum	-15	-47	-24	3.56	26	0.0003
L. middle frontal gyrus	-29	5	51	3.51	17	0.0003
R. claustrum	27	18	3	3.49	12	0.0004
R. cerebellum	12	-63	-27	3.39	18	0.0005
R. inferior parietal gyrus	42	-51	41	3.35	12	0.0006
Self-unaware						
R. collateral sulcus	38	-21	-23	4.16	79	0.0000
L. fusiform gyrus (T4)	-42	-38	-20	4.08	105	0.0000
R. superior temporal sulcus	48	-9	-24	4.08	33	0.0000
R. postcentral gyrus	27	-42	66	3.79	155	0.0001
L. superior parietal gyrus	-20	-78	50	3.68	33	0.0002
L. precentral gyrus	-30	2	38	3.60	14	0.0003
L. inferior temporal gyrus (T3)	-48	-71	-15	3.54	75	0.0003
L. cingulate gyrus	-9	-18	36	3.40	35	0.0005
Perceptual-unaware						
R. superior parietal gyrus	9	-69	42	3.66	126	0.0002
R. inferior frontal gyrus	48	24	9	3.65	129	0.0002
R. intraparietal sulcus	23	-60	39	3.65	125	0.0002
R. cerebellum	15	-42	-32	3.64	93	0.0002
L. superior parietal gyrus	-9	-72	33	3.62	21	0.0002
L. superior parietal gyrus	-12	-81	47	3.60	39	0.0003
R. intraparietal sulcus	36	-78	36	3.58	65	0.0003
R. superior temporal sulcus	53	2	-21	3.56	44	0.0003
R. superior frontal gyrus	14	48	50	3.54	16	0.0003
R. fusiform gyrus	33	-42	-20	3.51	41	0.0003
R. inferior lingual gyrus	-21	-65	24	3.46	36	0.0004
L. cingulate gyrus	-12	-45	41	3.41	35	0.0005
SU > OU						
L. inferior temporal sulcus, anterior part	-45	-7	-29	3.12	31	0.0012

All brain regions (peaks) with significant activation at $p < 0.001$ (uncorrected) and $k > 10$ voxels are listed for purely descriptive purposes. For the significant activations that are the

basis of the conclusions of this study, see Table 1. Labels of functional regions provided in brackets follow labels of anatomical regions. **Abbreviations:** SMA – supplementary motor area. For details on how "conceptual self-awareness", "perceptual-awareness", etc., were operationalized, see the "fMRI second-level data analysis" section in the main text.