Social Cognitive and Affective Neuroscience, 2017, 1845–1858

doi: 10.1093/scan/nsx116
Advance Access Publication Date: 4 October 2017
Original article

Oxytocin effects on self-referential processing: behavioral and neuroimaging evidence

Yi Liu,[1,2,3] Bing Wu,[4] Xuena Wang,[1,2,3] Wenxin Li,[5,6] Ting Zhang,[1,2,3] Xinhuai Wu,[4] and Shihui Han[1,2,3]

[1]School of Psychological and Cognitive Sciences, [2]PKU-IDG/McGovern Institute for Brain Research, and [3]Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China, [4]Department of Radiology, PLA Army General Hospital, Beijing, China, [5]Academy for Advanced Interdisciplinary Studies, and [6]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

Yi Liu and Bing Wu contributed equally to this work.
Correspondence should be addressed to Shihui Han, School of Psychological and Cognitive Sciences, Peking University, 52 Haidian Road, Beijing 100080, China. E-mail: shan@pku.edu.cn.

Abstract

Oxytocin (OT) influences other-oriented mental processes (e.g. trust and empathy) and the underlying neural substrates. However, whether and how OT modulates self-oriented processes and the underlying brain activity remains unclear. Using a double-blind, placebo-controlled between-subjects design, we manipulated memory encoding and retrieval of trait adjectives related to the self, a friend and a celebrity in a self-referential task in male adults. Experiment 1 ($N = 51$) found that OT *vs* placebo treatments reduced response times during encoding self-related trait adjectives but increased recognition scores of self-related information during memory retrieval. Experiment 2 ($N = 50$) showed similar OT effects on response times during encoding self-related trait adjectives. Moreover, functional magnetic resonance imaging (fMRI) results revealed that OT *vs* placebo treatments decreased the activity in the medial prefrontal cortex (MPFC) involved in encoding of self-related trait adjectives and weakened the coupling between the MPFC activity and a cultural trait (i.e. interdependence). Experiment 3 ($N = 52$) revealed that OT *vs* placebo treatments increased the right superior frontal activity during memory retrieval of self-related information. The results provide behavioral and fMRI evidence for OT effects on self-referential processing and suggest distinct patterns of OT modulations of brain activities engaged in encoding and retrieval of self-related information.

Key words: oxytocin; self-referential processing; fMRI, encoding; medial prefrontal cortex

Introduction

Oxytocin (OT) is a neuropeptide synthesized in the hypothalamus, functions as both hormone and neural transmitter, and plays a key role in social behaviors in both animals and humans (see Insel, 1992; Donaldson and Young, 2008; De Dreu, 2012 for review). Because social cognition—the processes of self- and other-related information—underlies human social behavior, there has been increasing interest in how OT influences other-oriented mental processes (e.g. trust, empathy) and the underlying neural substrates (see Bartz *et al.*, 2011; Meyer-Lindenberg *et al.*, 2011; Shamay-Tsoory and Abu-Akel, 2016; Ma *et al.*, 2016a for review). However, despite of the tremendous influence of self-concept and self-reflection on human motivation and behavior (Triandis, 1989; Markus and Kitayama, 1991), it is surprising that we have known little about OT effects on the processing of self-related information and the underlying neural mechanisms.

Received: 15 May 2017; **Revised:** 9 September 2017; **Accepted:** 2 October 2017

Early animal research has revealed that oxytocin facilitated maternal behavior (Pedersen *et al.*, 1982; Keverne and Kendrick, 1992) and enhanced social attachments with a mate (Young and Wang, 2004). Recent studies of humans have demonstrated extensive OT influences on cognition and emotion involved in social behavior. For example, intranasal administration of OT compared with placebo in human adults enhanced identification of facial expressions (Domes *et al.*, 2007), increased trust to others (Kosfeld *et al.*, 2005; Van IJzendoorn *et al.*, 2012; but see Nave *et al.*, 2015), facilitated positive social communication (Ditzen *et al.*, 2009), increased generosity in the ultimatum game but not in the dictator game (Zak *et al.*, 2007), enhanced interpersonal coordination (Arueti *et al.*, 2013; Mu *et al.*, 2016), and regulated ingroup favoritism in economic decisions (De Dreu *et al.*, 2010; Ma *et al.*, 2015). While these findings uncovered positive OT effects on social behavior, the findings of other studies also suggest negative OT effects on social interactions. For instance, in an economic game, OT *vs* placebo administration increased feelings of both envy, when a participant gained less money than others, and gloating, when a participant gained more money than others (Shamay-Tsoory *et al.*, 2009). OT enhanced cooperative behavior after prior contact with a game partner but exacerbated intrinsic self-interested behavior in anonymous conditions (Declerck *et al.*, 2014). OT *vs* placebo administration also made adults to lie more to benefit ingroup's outcome in a simple coin-toss prediction task in which participants could dishonestly report their performance levels (Shalvi and De Dreu, 2014). These behavioral findings indicate that OT effects on social cognition and behavior are strongly independent of social contexts.

Similarly, brain imaging research has shown evidence for context-dependent variations of OT effects on brain activity underlying social cognition and behavior. For example, OT compared with placebo treatments down-regulated amygdala responses to social and nonsocial threats (Kirsch *et al.*, 2005; Kanat *et al.*, 2015), reduced the amygdala activity when experiencing social trust betrayal and social evaluative threats (Baumgartner *et al.*, 2008; Grimm *et al.*, 2014), but increased amygdala activity during positive social-affective processes during cooperation (Rilling *et al.*, 2012) and in response to social feedback (Hu *et al.*, 2015) and happy faces (Gamer *et al.*, 2010). OT *vs* placebo administration increased neural responses in the reward system in response to happy faces (Gamer *et al.*, 2010) and during anticipation of social reward (Groppe *et al.*, 2013) but decreased the reward-related activity during nonsocial judgments (Gordon *et al.*, 2013). OT *vs* placebo treatments enhanced empathic neural responses to perceived pain expressions of racial ingroup members but not of racial outgroup members (Sheng *et al.*, 2013) and increased pleasantness and neural responses in the insula, anterior cingulate, and orbital frontal cortex in heterosexual males when they believed to be touched by a woman but not by a man (Scheele *et al.*, 2014).

The OT effects on human behavior and brain activities have been understood in the framework of increasing the salience of and sensitivity to social signals (Shamay-Tsoory and Abu-Akel, 2016), promoting motivation for social interactions (Stavropoulos and Carver, 2013), and facilitating social adaptation (Ma *et al.*, 2016a). For example, the social salience hypothesis assumes that a key functional role of OT is to modulate attention orienting responses to external contextual social cues and to increasing the salience of competitive social cues (Bartz *et al.*, 2011; Shamay-Tsoory and Abu-Akel, 2016). The current theories and models of OT function focus on how OT modulates behavior and brain activity toward others that are important for social interaction and adaption. However, appropriate and efficient social interactions require not only other-oriented emotion/motivation but also appropriate cognitive/affective processes of oneself (Banaji and Prentice, 1994; Cross and Vick, 2001; Gardner *et al.*, 2002). Thus it is essential to examine the functional role of OT in the process of self-related information and the underlying neural mechanisms.

The previous brain imaging findings suggest distinct neural underpinnings of the processes of self- and other-related information (Lieberman, 2007). For example, reflection on one's own attributes activated the default mode network including the ventral medial prefrontal cortex (vMPFC) and posterior cingulate cortex (PCC) (Kelley *et al.*, 2002; Macrae *et al.*, 2004; Zhu *et al.*, 2007; Han *et al.*, 2008) whereas inference of others' mental states engaged the dorsal MPFC (dMPFC) and temporoparietal junction (TPJ) (Amodio and Frith, 2006; Frith and Frith, 2006; Saxe *et al.*, 2006). The neural correlates of self-reflection and inference of others' mental states are modulated by sensory experiences in different ways such that absence of visual experience decreased the vMPFC activity underlying self-reflection (Ma and Han, 2011) but did not change the dMPFC and TPJ activity involved in reasoning of others' mental states (Bedny *et al.*, 2009). Other neuroimaging findings suggest a teeterboard relationship between the processes of self- and other-related information. For instance, relative to young adults, older adults showed increased neural activity underlying remembering of self-related information but decreased neural activity related to remembering of other-related information (Gutchess *et al.*, 2010). Healthy adults showed reduced activity in the default mode network during cognitively demanding tasks (Greicius *et al.*, 2003), whereas individuals with autism, who are diagnosed on the basis of impaired social communication with others, showed increased activation in the default model network even when being involved in a cognitive task (Kennedy *et al.*, 2006).

Given the possible teeterboard relationship between the processes of self- and other-related information and the related separate neural underpinnings, one may expect that OT *vs* placebo administration may produce opposite effects on the processes of self-related and other-related information. Because the main stream of current findings indicate that OT *vs* placebo administration modulates other-oriented processing, it is likely that OT *vs* placebo administration may weaken the process of self-related information by decreasing the neural activity underlying the processing of self-related information. Liu *et al.* (2013) first tested this hypothesis by recording event-related potentials (ERPs) in a self-referential task (Rogers *et al.*, 1977). Typically, during the encoding phase of the self-referential task participants are required to judge whether a number of trait adjectives can describe the self or a familiar other (e.g. a celebrity). In the following retrieval phase, participants are presented with the old words used during trait judgments and new words, and are asked to recall as many of the old words as they can. Behavioral performance of the self-referential task is characterized by the self-reference effect, i.e. self-descriptive words are better remembered than those descriptive of others (Rogers *et al.*, 1977; Symons and Johnson, 1997). ERP studies found that, relative to judgments of word valence, trait judgments of oneself induced faster responses and increased the amplitude of a frontal positive activity at 220–280 ms (P2) (Mu and Han, 2010; Liu *et al.*, 2013). Moreover, Liu *et al.* (2013) found that OT *vs* placebo administration significantly decreased the P2 effect associated with the self-referential processing. In contrast, OT *vs* placebo administration tended to increase the amplitude of a late positive potential at 520–1000 ms (LPP) during the processing of personality traits of a celebrity.

A recent functional magnetic resonance imaging (fMRI) study employed the same paradigm to examine OT effects on behavioral performance and brain activities during trait judgments of oneself and others (e.g. one's mother, classmate and stranger; Zhao *et al.*, 2016). Whole brain analyses of the fMRI data with a strict threshold revealed increased modulations of activities in the MPFC and PCC by trait judgments of different target persons, replicating the previous findings (Kelley *et al.*, 2002; Macrae *et al.*, 2004; Northoff *et al.*, 2006; Zhu *et al.*, 2007; Ma and Han, 2011; Ma *et al.*, 2014). However, the whole brain analyses of the fMRI data with the same threshold did not find robust effect of OT administration or differential OT effects on brain activities underlying trait judgments of different target persons. Only when using a lenient threshold (small volume correction with $P < 0.05$) did Zhao *et al.* (2016) find that OT *vs* placebo treatments tended to decrease the MPFC activity in response to trait judgments, but this effect was not specific to trait judgments of the self.

The aforementioned ERP and fMRI studies initiated a brain imaging approach to OT effects on self-referential processing, but left open a few questions. First, neither Liu *et al.* (2013) nor Zhao *et al.* (2016) found behavioral evidence for self-specific OT effects on memory encoding and retrieval, which, however, is pivotal for building a conceptual framework of the functional role of OT in self-referential processing. Second, although the previous ERP results suggest a self-specific OT effect on the neural activity engaged in self-referential processing (Liu *et al.*, 2013), the fMRI study did not find self-specific OT effect on the neural correlates of self-referential processing (Zhao *et al.*, 2016). Thus it remains unclear whether the activity in the key brain regions underlying self-referential processing (e.g. the MPFC) is modulated by OT *vs* placebo treatments. Finally, the previous research (Liu *et al.*, 2013; Zhao *et al.*, 2016) only tested OT effects on the neural activities engaged in encoding of self-relevant trait adjectives but did not examine OT effects on the neural correlates of memory retrieval of self-related information.

Here we conducted three experiments to examine the effects of OT (*vs* placebo) treatments on behavioral performance and brain activity in the self-referential task (Rogers *et al.*, 1977) using a double-blind, placebo-controlled between-subjects design. To avoid potential influences of scanner noise and body position on behavioral performance during encoding and retrieval of self-related information, Experiment 1 measured behavioral performance in the self-referential task in a quiet testing room from two groups of participants who had been treated with nasal spray of OT or placebo. Reaction times (RTs) to trait judgments were calculated to index behavioral performance during the encoding phase and recognition scores of self-/other-related items during the retrieval phase were calculated to estimate participants' memory retrieval. Because Experiment 1 found evidence that OT *vs* placebo treatments reduced response times during encoding of self-related trait adjectives but improved recognition of self-related items during memory retrieval, Experiment 2 tested whether OT (*vs* placebo) treatments decrease neural activities in the brain regions (e.g. the MPFC) involved in coding self-related information by scanning two independent groups of participants using fMRI during trait judgment tasks after OT or placebo treatment. Experiment 3 further investigated self-specific OT effects on neural correlates of memory retrieval of self-related information by first asking two independent subject groups to perform trait judgments of oneself and others. After the encoding phase participants were treated with OT or placebo and were then scanned during memory retrieval of self-related and other-related information. The behavioral results of Experiment 1 would predict OT enhancement of brain activity related to memory retrieval of self-related information.

Materials and methods

Participants

Experiment 1 recruited 51 male adults (mean age = 23.16, s.d. = 2.39 years). Participants were randomly assigned to OT ($n = 26$) or placebo ($n = 25$) treatment. Experiment 2 recruited 50 male adults (mean age = 22.70, s.d. = 2.53 years) who were randomly assigned to OT or placebo treatment ($n = 25$ for each group). Experiment 3 recruited 52 male adults (mean age = 21.90, s.d. = 2.55 years) who were randomly assigned to OT or placebo treatment ($n = 26$ for each group). Experiments 1–3 recruited independent subject groups. Exclusion criteria were any self-reported history of medical or psychiatric disorder and of medication/drug/alcohol abuse. All were right-handed and had normal or corrected-to-normal vision. All were paid for their participation. Informed consent was obtained prior to participation. This study was approved by the ethics committee of the School of Psychological and Cognitive Sciences, Peking University.

OT and placebo administration

A double-blind placebo-controlled between-subjects design was used in all the three experiments. The self-referential task required participants to perform a 'surprising' memory test and this does not allow us to employ a within-subjects design in the current work. Twenty-four IU OT or placebo (containing all of the active ingredients except for the neuropeptide) was administered with a nasal spray 45 min before the behavioral test or fMRI scanning. The spray was administered to participants three times and each administration consisted of one inhalation of the spray with 4 IU into each nostril. This procedure is similar to that in the previous work (Petrovic *et al.*, 2008; De Dreu *et al.*, 2010; Mikolajczak *et al.*, 2010; Sheng *et al.*, 2013; Ma *et al.*, 2016b).

Stimuli and procedure

In Experiment 1, 192 trait adjectives (each with 2 Chinese characters) were selected from an established personality trait adjective pool (Liu, 1990) for trait and font judgment tasks. These adjectives were randomly assigned to 2 groups of 96 words (half positive and half negative) for two sessions. There were 8 blocks of 12 trials (half positive and half negative adjectives) in each session of the encoding phase. Each trial consisted of a cue word (i.e. self, a friend's name, a celebrity's name or Font) above a trait adjective presented at the center of a screen for 2250 ms followed by a fixation cross of 750 ms (Figure 1). The cue and trait adjectives subtended a visual angle of $1.2^\circ \times 0.6^\circ$ at a viewing distance of 95 cm. In each session of the encoding phase, participants had to judge whether a trait adjective can describe the self, a friend or Liu Xiang (a well-known Chinese athlete) in two blocks of trials and had to judge the font of an adjective (bold- *vs* light-faced character) in two blocks by pressing one of two buttons. Friend- and celebrity-judgments were used to control for familiarity and general person/semantic processing. Font-judgments were used to control general perceptual processing. There was a break of 8 s between two

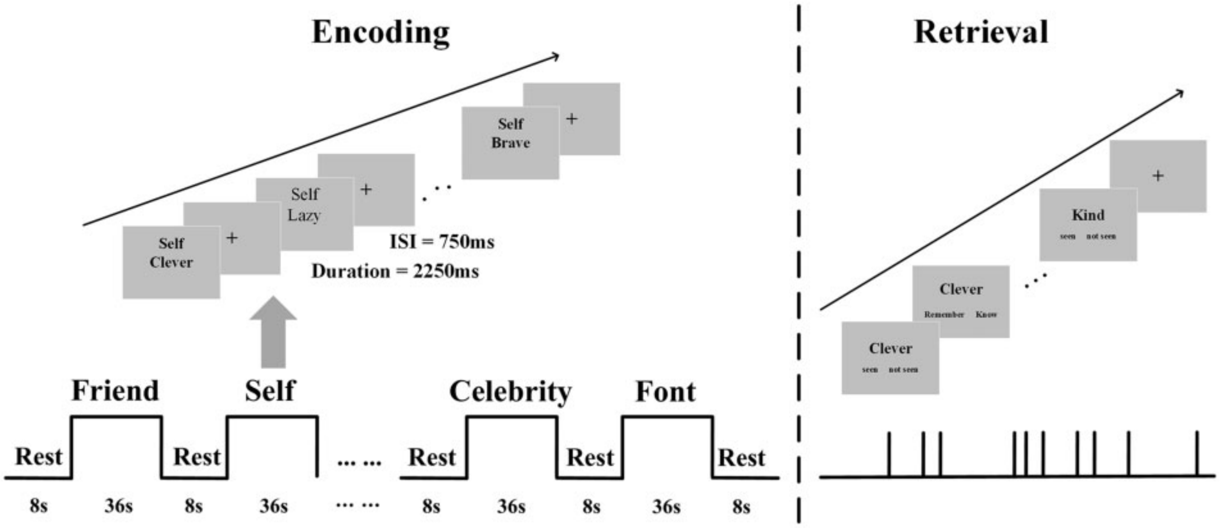

Fig. 1. Illustration of the stimuli and procedure of the current work. The left panel illustrates the stimuli and procedure during the encoding phase of the self-referential task. The right panel illustrates the stimuli and procedure during the retrieval phase of the self-referential task.

consecutive blocks of trials while only a fixation cross was presented on the screen. After the judgment tasks participants were asked to perform a surprise memory test—the retrieval phase. During the memory test, the trait adjectives used in the judgment tasks and 192 new trait adjectives were presented on a screen in a random order. Participants were asked to identify old or new items by a button press without a time limit. After yes responses participants were further asked to make an *R/K* judgment (Tulving, 1985, 1999) for old words by indicating whether 'remembering' (consciously recollect specific details of the item that appeared in the earlier list) or simply 'knowing' (not accompanied by recollective experience but has a feeling of knowing or familiarity to the subjects) the item. The total recognition score, the score of remembering (i.e. *R* score), and the score of knowing (i.e. *K* score) were calculated to estimate participants' memory retrieval.

The stimuli and procedure in Experiment 2 were the same as those in Experiment 1 except that participants were scanned during the judgment tasks and there was no memory test after the trait and font judgment tasks. The stimuli and procedure in Experiment 3 were the same as those in Experiment 1 except the following. Participants performed only trait judgments on the self, a friend, and a celebrity. There were 256 trait adjectives (each with 2 Chinese characters) for trait judgment tasks and recognition task. A total of 192 adjectives were randomly assigned to 2 groups of 96 words (half positive and half negative) for two judgment sessions. There were 6 blocks of 16 trials (half positive and half negative adjectives) in each session for trait judgment tasks outside scanner. Participants were then administered with OT or placebo treatment, and thereafter, were scanned during a surprise memory test. The 192 trait adjectives used in the judgment tasks and 64 new trait adjectives were organized into 4 groups of 64 words (half positive and half negative) for 4 functional scans. In each scan there were 16 words for self-/friend-/celebrity-judgments and 16 new words which were presented in a random order. Each word was presented at the center of a screen for 3 s followed by a fixation with a duration varying among 1, 3, 5 and 7 s. Participants were asked to identify old or new items by a button press. No *R/K* judgment was required after a yes response during scanning.

All participants completed the Self-Esteem Scale (Rosenberg, 1995), the Self-Construal Scale (Singelis, 1994), and the Inclusion of other in the self-scale (Aron *et al.*, 1992) before the OT or placebo treatment.

fMRI data acquisition and analysis

Brain images were acquired in Experiments 2 and 3 using a 3.0 T GE Signa MR750 scanner (GE Healthcare; Waukesha, WI, USA) with a standard head coil. Functional images were acquired by using T2-weighted, gradient-echo, echo-planar imaging sequences sensitive to BOLD contrast (64 × 64 × 32 matrix with 3.75 × 3.75 × 5 mm^3 spatial resolution, repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°, field of view = 24 × 24 cm). A high-resolution T1-weighted structural image (512 × 512 × 180 matrix with a spatial resolution of 0.47 × 0.47 × 1.0 mm^3, repetition time = 8.204 ms, echo time = 3.22 ms, flip angle = 12°) was acquired before the functional scans.

Functional images were preprocessed using SPM8 (the Wellcome Trust Centre for Neuroimaging, London, UK). Head movements were corrected within each run and six movement parameters (translation; *x*, *y*, *z* and rotation; pitch, roll, yaw) were extracted for further analysis in the statistical model. The anatomical image was coregistered with the mean realigned functional image and then was normalized to the standard Montreal Neurological Institute (MNI) template. The functional images were resampled to 3 × 3 × 3 mm^3 voxels, normalized to the MNI space using the parameters of anatomical normalization and then spatially smoothed using an isotropic of 8 mm full-width half-maximum (FWHM) Gaussian kernel.

Fixed effect analyses of the fMRI data in Experiment 2 were first conducted by applying a general linear model (GLM) to the fMRI data. All five conditions (Self, Friend, Celebrity, Font and rest) were included in the model. The design matrix also included the realignment parameters to account for any residual movement-related effect. A box-car function was used to convolve with the canonical hemodynamic response in each condition. To examine possible differential OT effects on brain activities underlying encoding of self-related and other-related information, we conducted a whole brain ANOVA with Target person (self, friend and celebrity) as a within-subjects variable and Treatment (OT *vs* placebo) as a between-subjects variable. Whole-brain random effect analyses were then conducted on the contrast images of self- *vs* celebrity-judgments, friend- *vs* celebrity-judgments and celebrity- *vs* font-judgments to reveal the brain regions involved in self-related, friend-related and celebrity-related processing in the OT and placebo groups, respectively. These contrast images were further subject to two-sample (OT *vs* placebo groups) *t*-tests to identify the OT effect on the neural activities involved in self-, friend- and celebrity-judgments, respectively. The 'rest' between two blocks of trials was used as a baseline when calculating the contrast image of font-judgments.

Because retrieval effort and retrieval success were two different processes involved in memory retrieval (Buckner *et al.*, 1998), fixed effect analyses of the fMRI data in Experiment 3 were conducted by applying a GLM to fMRI data in two ways. A box-car function was used to convolve with the canonical hemodynamic response in each condition. First, we classified trait adjectives into four categories, i.e. adjectives used for self-/friend-/celebrity-judgments and new words, which were included in model estimation. The contrasts of old *vs* new worlds were then calculated to identify the neural activities related to recollection of or familiarity with stored information (Rugg and Curran, 2007). Second, we classified trait adjectives into five categories, i.e. remembered trait adjectives used for self-/friend-/celebrity-judgments, missed old words, and correctly rejected new words, which were included in model estimation. The contrasts of corrected recognized old *vs* missed old words were then calculated to identify the neural activities involved in successful memory retrieval. Since the number of trials was different in the five conditions, we randomly selected the same number of trials in each condition for data analyses based on the minimal number of trials across the five conditions for each participant. The mean numbers of trials per condition used for fMRI data analyses was 32 and did not differ between OT and placebo groups [$t(50) = -0.939$, $P = 0.352$, Cohen's $d = 0.26$). Besides the words in the five conditions, new words identified as old and the unselected words for contrast analyses trials were also included in the model. In both models, the design matrix also included the realignment parameters to account for any residual movement-related effect. These contrast images were also subject to two-sample *t*-tests to identify the OT effect on the brain activity involved in memory retrieval. Contrast values in activated brain regions were extracted from each condition using MarsBaR (http://marsbar.sourceforge.net) for the region-of-interest analyses. Brain activations in the whole brain analyses in Experiments 2 and 3 were defined using a threshold of $P < 0.05$ (corrected by a combined voxel-intensity and cluster-size threshold of single voxel $P < 0.001$ and cluster extent > 21 voxels, based on Monte-Carlo simulation (1000 iterations, Slotnick *et al.*, 2003).

Results

Experiment 1: OT effects on behavioral performance

Experiment 1 recorded behavioral performance during both encoding and retrieval phases of the self-referential task from the OT and placebo groups, respectively. The OT and placebo groups were matched in questionnaire measures of self-esteem, self-construals and closeness between the self and a friend (Ps > 0.05; see Table 1). We conducted a repeated measure analysis (ANOVA) of RTs during trait and font judgments with Target (Self, Friend, Celebrity, Font) as a within-subjects variable and Treatment (OT *vs* placebo) as a between-subjects variable to estimate the OT effect on memory encoding. This revealed a significant main effect of Target [$F(3, 147) = 76.401$, $P < 0.001$, $\eta_p^2 = 0.609$] because participants responded faster to font-judgments than trait-judgments of self/friend/celebrity (Ps < 0.001; Figure 2A), whereas RTs to traits judgments of self/friend/celebrity did not differ significantly (Ps > 0.05). The main effect of Treatment was marginally significant [$F(1, 49) = 3.597$, $P = 0.064$, $\eta_p^2 = 0.068$] as OT *vs* placebo treatments tended to speed participants' responses. Most important, the ANOVA of RTs revealed a significant interaction of Target × Treatment [$F(3, 147) = 3.286$; $P = 0.023$, $\eta_p^2 = 0.063$] due to distinct OT effects on RTs to different targets. Further simple effect analyses confirmed that OT *vs* placebo treatments led to shorter RTs to self-judgments (mean difference = −1678 ms, 95% CI = (−2869, −487), $P = 0.007$] but did not influence RTs to other targets (Ps >0.05) (Figure 2A), suggesting that OT *vs* placebo treatments made participants spend less time on judgments of the self but not of others. We also conducted ANOVAs of the percentage of yes response to different targets with Target (Self, Friend and Celebrity) as a within-subjects variable and Treatment (OT *vs* placebo) as a between-subjects variable. This revealed only a significant main effect of Target [$F(2, 98) = 5.831$, $P = 0.004$, $\eta_p^2 = 0.106$] because participants tended to made more yes responses during self- than friend-/celebrity-judgments [self *vs* friend: mean difference = 0.030, 95% CI = (−0.003, 0.062), $P = 0.072$; self *vs* celebrity: mean difference = 0.053, 95% CI = (0.019, 0.087), $P = 0.003$; friend *vs* celebrity: mean difference = 0.024, 95% CI = (−0.004, 0.051), $P = 0.09$]. The mean accuracy of font-judgments was high (93.75%) and did not differ significantly between the OT and placebo groups ($P > 0.05$).

To estimate participants' performance during memory retrieval, we calculated the recognition score (i.e. hit rates minus false alarm rates), *R* score (the percentage of remembering answer after yes responses to old words minus that to new words), and *K* score (percentage of knowing answer after yes responses to old words minus that to new words), respectively. A 2 × 4 ANOVA of the recognition score with Target (Self, Friend, Celebrity and Font) as a within-subjects variable and Treatment (OT *vs* placebo) as a between-subjects variable showed only a significant main effect of Target [$F(3, 147) = 69.338$; $P < 0.001$, $\eta_p^2 = 0.586$] due to better memory performance on trait adjectives used for self-judgments than for other judgment tasks (Ps <0.001), replicating the self-reference effect (Rogers *et al.*, 1977) (Figure 2B). The ANOVA of the *R* score revealed a significant main effect of Target [$F(3, 147) = 57.689$; $P < 0.001$, $\eta_p^2 = 0.541$] and a marginally significant effect of Treatment [$F(1, 49) = 3.872$; $P = 0.055$]. Participants tended to show a greater *R* score of trait adjectives used for self-judgments than for other judgment tasks and OT *vs* placebo treatments tended to increase the *R* score. Although the Target × Treatment interaction on the *R* score was not significant [$F(1, 49) = 1.600$; $P = 0.192$, $\eta_p^2 = 0.023$], separate analyses revealed a significant OT effect on the *R* score of trait adjectives used for self-judgments (mean difference = 0.122, 95% CI = [0.011, 0.234], $P = 0.032$) but not for other judgment tasks (Ps > 0.05; Figure 2C). The ANOVA of the *K* score showed a significant interaction between Target and Treatment [$F(3, 147) = 3.047$; $P = 0.031$, $\eta_p^2 = 0.059$] because OT *vs* placebo treatments tended to decrease the *K* score of trait adjectives used for self-judgments (mean difference = −0.090, 95% CI = [−0.187, 0.006], $P = 0.065$) but did not affect the *K* scores of trait adjectives used for other judgment tasks (Ps > 0.05; Figure 2D). The results suggest that OT (*vs* placebo) treatments tended to facilitate memory retrieval of self-related information by

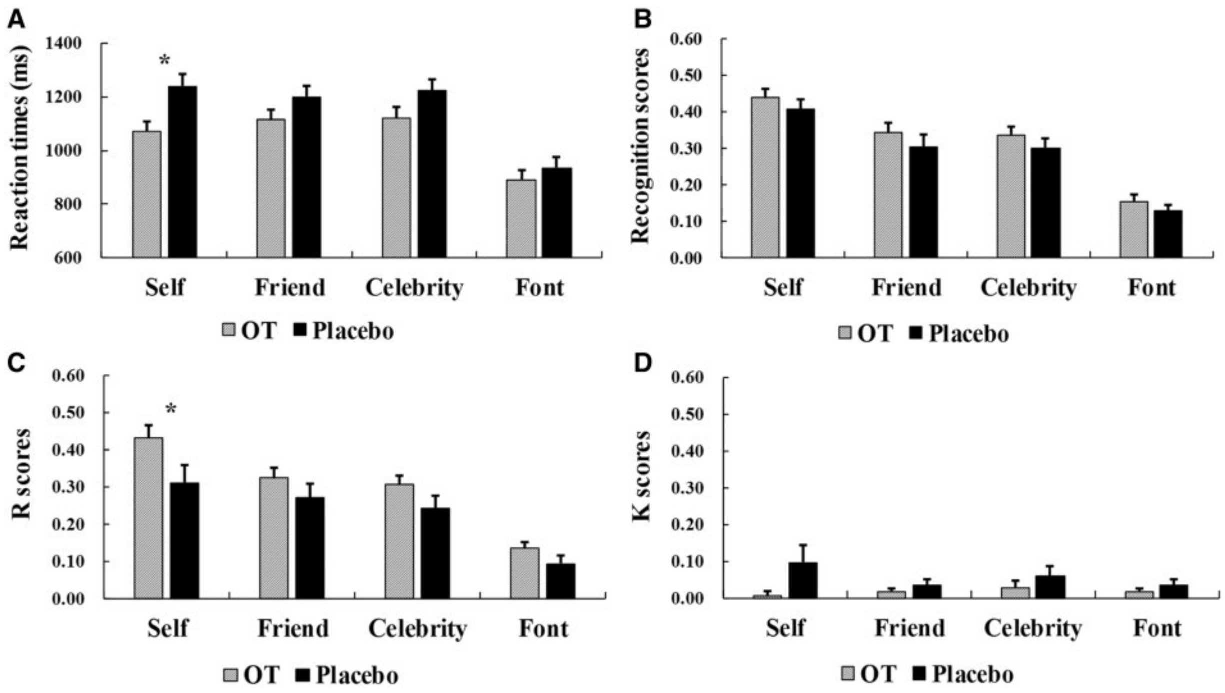

Fig. 2. Behavioral results of Experiment 1. In the encoding phase, results were shown for (A) RTs. In memory retrieval phase, results were shown for (B) recognition scores calculated by using hit rates minus false alarm rates, (C) R scores (the percentage of 'remembering answer after yes responses to old words minus that to new words) and (D) K scores (percentage of knowing answer after yes responses to old words minus that to new words). OT: oxytocin. $^{*}P < 0.05$.

Table 1. The results of self-esteem, self-construal and closeness measures

	Experiment 1	Experiment 2	Experiment 3
Mean (s.d.)	OT/placebo group	OT/placebo group	OT/placebo group
Self-esteem	27.77/26.20 (3.13/3.48)	26.32/27.12 (2.59/2.30)	22.12/21.69 (2.53/2.59)
Self-construal			
Interdependent	63.58/63.24 (6.30/8.44)	62.60/63.28 (6.15/8.69)	60.46/61.50 (7.37/6.27)
Independent	58.64/56.96 (7.50/8.22)	57.84/57.92 (6.80/10.25)	58.00/57.33 (5.89/7.81)
Closeness			
Self-friend	4.27/4.44 (1.08/1.12)	4.56/4.64 (1.00/0.99)	4.58/4.38 (0.95/1.47)
Self-celebrity	1.46/1.36 (0.76/0.64)	1.68/1.44 (0.85/0.65)	1.46/1.58 (0.95/0.86)

OT, oxytocin.
Note: The comparisons of self-esteem, self-construals and closeness between OT and placebo groups failed to reach significance in all experiments ($Ps > 0.05$).

increasing recollect specific details of self-related items but deceasing the feeling of knowing or familiarity of the items used for self-judgments during the encoding phase.

Experiment 2: OT effects on brain activity during encoding

Experiment 2 recorded brain activities during the encoding phase from the OT and placebo groups, respectively. The OT and placebo groups were matched in questionnaire measures of self-esteem, self-construals and closeness between the self and a friend ($Ps > 0.05$; see Table 1). The mean accuracy of font-judgments was high (95.50%) and did not differ between the OT and placebo groups [$t(48) = 1.214$; $P = 0.231$, Cohen's $d = 0.34$]. The results of RTs showed a pattern similar to that observed in Experiment 1 (Supplementary Figure S1), though the OT effect on RTs to self-judgments did not reach significance. As can be seen in Figure 1 and Supplementary Figure S1, the mean RTs appeared to be longer in Experiment 2 than in Experiment 1, reflecting possible influences of scanner noise, body gestures, and state of arousal, which in turn might weakened OT effects on RTs behavioral performance.

To examine differential OT effects on brain activities underlying encoding of self-related and other-related information, we conducted a whole brain ANOVA with Target (self, friend and celebrity) as a within-subjects variable and Treatment (OT *vs* placebo) as a between-subjects variable. This analysis revealed significant interactions of Target × Treatment on the activities in the vMPFC, dMPFC, bilateral orbital frontal cortices (OFC), and the bilateral frontal and parietal cortices (Figure 3A; Table 2). To illustrate the pattern of OT effects on the brain activities, we extracted the contrast values of self-/friend-/celebrity-judgments *vs* rest from the spheres with 10 mm diameter centered at the peak voxes in the activated regions shown in the interaction analysis from the OT and placebo groups, respectively. As shown in Figure 3B and Supplementary Figure S2, OT (*vs*

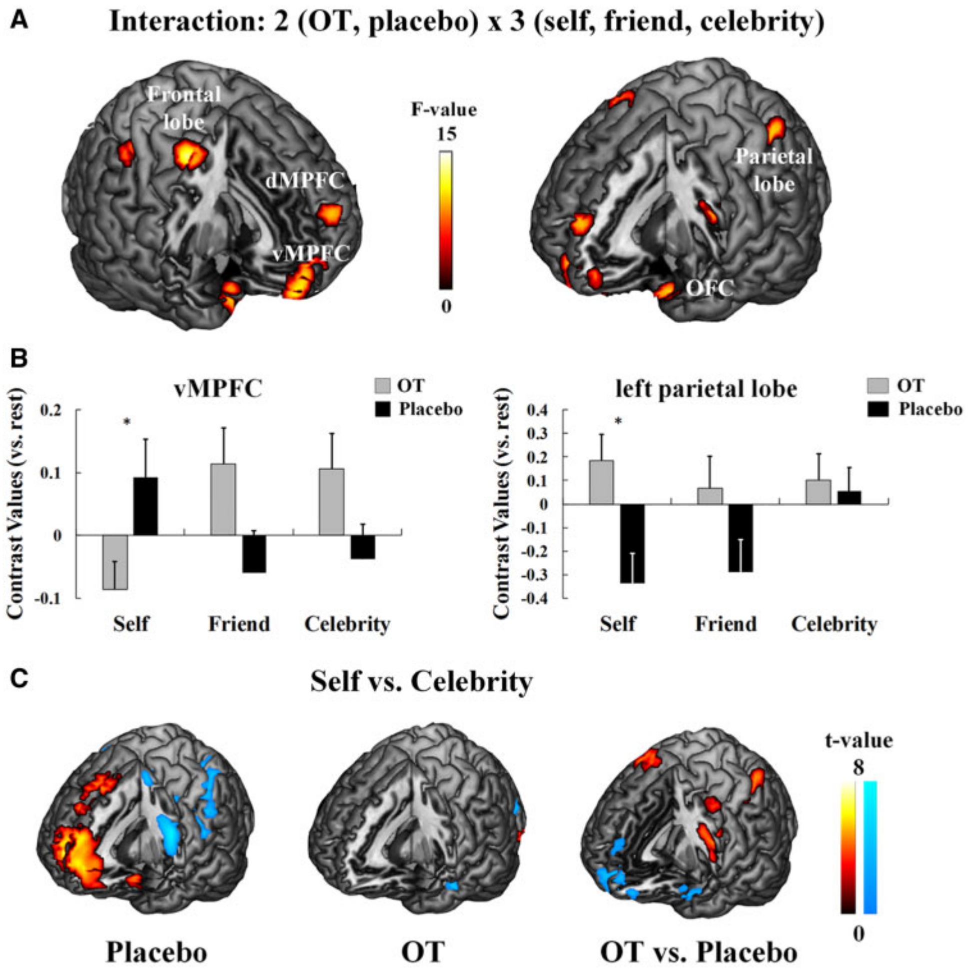

Fig. 3. fMRI results in Experiment 2. (A) Illustration of brain regions showed significant 2 × 3 interaction of Treatment (OT, placebo) and Target (self, friend and celebrity) and (B) the contrast values of self-, friend- and celebrity-judgments *vs* rest in vMPFC and left parietal lobe as an example. (C) Brain activity during self- *vs* celebrity-judgments in placebo and OT groups and the difference between two groups. Red regions were positive activation and blue regions were negative activation. dMPFC, dorsal medial prefrontal cortex; vMPFC, ventral medial prefrontal cortex; OFC, orbital frontal cortex; OT, oxytocin; *$P < 0.05$.

placebo) treatments tended to decrease the MPFC and OFC activities but increase the parietal activity during self-judgments. The OT effects on brain activities underlying self-referential processing were further examined by conducting a whole brain two-sample *t*-test of the contrast of self- *vs* celebrity-judgments, which showed that, relative to the placebo group, the OT group showed decreased activations in the vMPFC/dMPFC and bilateral OFC but increased activations in the bilateral frontal and parietal cortices (Figure 3C).

We also assessed whether OT *vs* placebo treatments modulated the association between participants' traits and brain activities involved in self-referential processing. We conducted whole-brain analyses and revealed significantly greater activations in the MPFC during self- *vs* celebrity-judgments ($x/y/z = -3/59/13$, $z = 5.41$; $k = 1395$) when collapsing the data from the placebo and OT groups. We then examined the association between the MPFC activity shown in the contrast of self- *vs* celebrity-judgments and participants' interdependence (defined by the differential score of interdependent *vs* independent self-construal items) and found a significant negative correlation for the placebo group [$r(25) = -0.477$; $P = 0.016$; Figure 4A]. This replicated the previous findings (Ma *et al.*, 2014) and suggests a coupling between participants' cultural traits and brain activity underlying self-referential processing. However, the OT group did not show significant correlation between interdependence and the MPFC activity [$r(25) = 0.159$; $P = 0.447$; Figure 4B]. We further conducted Fisher-z transformation and confirmed the significant group difference in the correlation between interdependence and the MPFC activity ($z = 2.25$, $P = 0.024$).

Experiment 3: OT effects on brain activity during memory retrieval

Experiment 3 recorded brain activities during memory retrieval from the OT and placebo groups, respectively. OT or placebo treatments were administered after the encoding phase outside the scanner but before the retrieval phase during scanning. The OT and placebo groups were matched in questionnaire measures of self-esteem, self-construals and closeness between the self and a friend ($Ps > 0.05$; see Table 1). RTs and recognition scores during the retrieval phase were subject to ANOVAs with Target (Self, Friend, Celebrity and New) as a within-subjects variable and Treatment (OT *vs* placebo) as a between-subjects variable. These analyses revealed only significant main effects of Target [RTs: $F(3, 150) = 25.480$; $P < 0.001$, $\eta_p^2 = 0.338$; recognition score: $F(2, 100) = 51.658$; $P < 0.001$, $\eta_p^2 = 0.508$], suggesting a descending sequence of RTs in response to new words, old

Table 2. Brain activations shown in different contrasts in Experiments 2 and 3

Region	MNI Coordinates			Cluster	PeakZ
	x	y	z		
Experiment 2					
2 ×3 ANOVA interactions					
Dorsal MPFC	−9	65	25	32	3.82
Ventral MPFC	9	53	−23	186	4.16
Left lateral OFC	−36	23	−20	46	4.02
Right lateral OFC	27	26	−29	46	3.99
Left frontal lobe	−48	−1	28	26	3.99
Right frontal lobe	36	−4	58	73	4.48
Left parietal lobe	−33	−55	52	73	3.87
Right parietal lobe	45	−40	49	34	3.41
Experiment 3					
Old>New					
Precuneus	0	−46	49	2437	5.60
Left parietal lobe	−39	−79	34	[a]	5.64
Right parietal lobe	42	−67	31	29	3.80
Left inferior frontal gyrus	−27	11	61	395	4.65
Hit>Miss					
precuneus/PCC	−6	−79	34	2777	6.25
MPFC	3	68	7	664	5.59
Left superior frontal gyrus	−27	26	55	344	4.78
Left parahippocampal gyrus	−18	−4	−29	75	4.46
Right parahippocampal gyrus	21	−4	−26	36	4.02
Right hippocampus	30	−19	−17	30	4.10
OT>placebo					
Self-celebrity					
RSFG	27	17	55	38	3.70
Hit-miss					
Right hippocampus	30	−19	−14	25	3.81

[a] Left parietal lobe was the same cluster with precuneus.

Note: MPFC, medial prefrontal cortex; OFC, orbital frontal cortex; PCC, posterior cingulate cortex; RSFG, right superior frontal gyrus; OT, oxytocin.

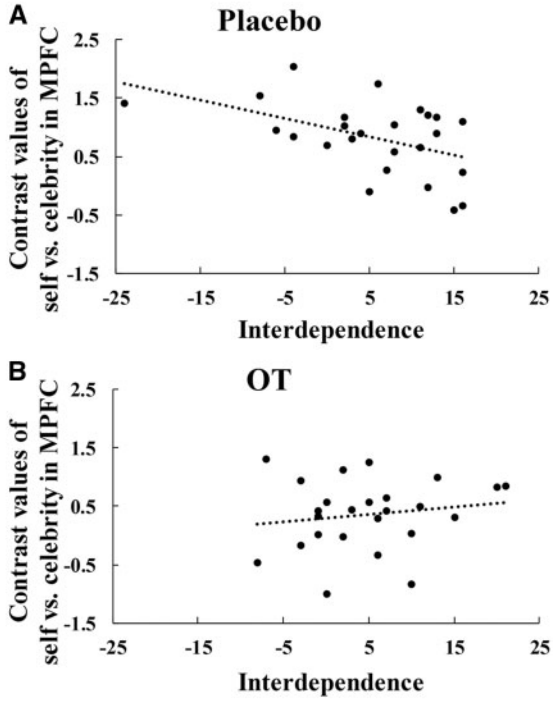

Fig. 4. Correlations of interdependence self-construals and brain activity of MPFC during self- *vs* celebrity in (A) placebo group and (B) OT group. MPFC, medial prefrontal cortex; OT, oxytocin. The correlation between MPFC activity and interdependence in the placebo group was significant even removing one outlier [$r(24) = -0.470$; $P = 0.020$].

words related to a celebrity, a friend, and the self and an increasing order of recognition scores of the old words related to the self, a friend, and a celebrity (Supplementary Figure S3).

The whole brain fMRI data analysis first identified brain regions involved in recognition of the old words used during trait judgments by contrasting neural responses to old *vs* new words collapsed for the placebo and OT groups. The contrast of old (collapsing the words used for self-/friend/celebrity-judgments) *vs* new words showed activations in the bilateral parietal cortex, precuneus, and left superior frontal cortex (Figure 5A; Table 2). Brain activities related to successful retrieval were further estimated by conducting whole-brain analyses of the contrast of old words that were successfully recognized *vs* those that were missed during the retrieval task. This contrast, collapsed for the placebo and OT groups, showed activations in the precuneus/PCC, MPFC, left superior frontal gyrus, left parahippocampal gyrus, right parahippocampal gyrus and right hippocampus (Figure 5B; Table 2).

To examine OT effects on brain activities during memory retrieval of trait adjectives related to different targets, we conducted a whole brain ANOVA with Target (self, friend and celebrity) as a within-subjects variable and Treatment (OT *vs* placebo) as a between-subjects variable. This analysis, however, did not find significant activations. To further explore the OT effects on brain activities during memory retrieval of trait adjectives, we conducted whole brain two sample *t*-tests of the contrast of self-(or friend- or celebrity-)related old words *vs* new words between the OT and placebo groups. The analyses revealed increased activation in the right superior frontal gyrus (RSFG) (x/y/z = 27/14/58, z = 3.70; *k* = 32) only for the contrast of self-related old *vs* new words (Figure 5C). We also conducted whole brain two sample *t*-tests of the contrast of self-related *vs* celebrity-related words between the OT and placebo groups and also found a significant activation in the RSFG (x/y/z = 27/17/55, z = 3.70; *k* = 38; Table 2). Similarly, we conducted whole brain two-sample *t*-tests of the contrast of the old words that were successfully recognized *vs* those that were missed during the retrieval task between the OT and placebo groups. The analyses revealed that OT *vs* placebo treatments significantly increased the activity in the right hippocampus (x/y/z = 30/−19/−14, z = 3.81; k = 25) when collapsing the data for self-/friend-/celebrity-judgments (Figure 5D; Table 2).

To estimate the possible routine that the OT treatment facilitates memory retrieval of self-related information, we extracted the old *vs* new contrast values in the brain regions that showed significant activations in old *vs* new contrast in Table 2 and examined the correlation between the contrast values and participants' memory performance. This revealed that the left parietal activity positively predicted participants' recognition scores (hit minus false alarm) [$r(52) = 0.530$, $P < 0.001$; Figure 6A]. Interestingly, the left parietal activity was also positively correlated with the activity in the RSFG identified in the whole brain two sample *t*-tests of the contrast of self-related *vs* celebrity-related words between the OT and placebo groups [$r(52) = 0.280$, $P = 0.044$; Figure 6B], suggesting a possible mechanism through which OT (*vs* placebo) treatments influenced participants' memory performance during the retrieval phase.

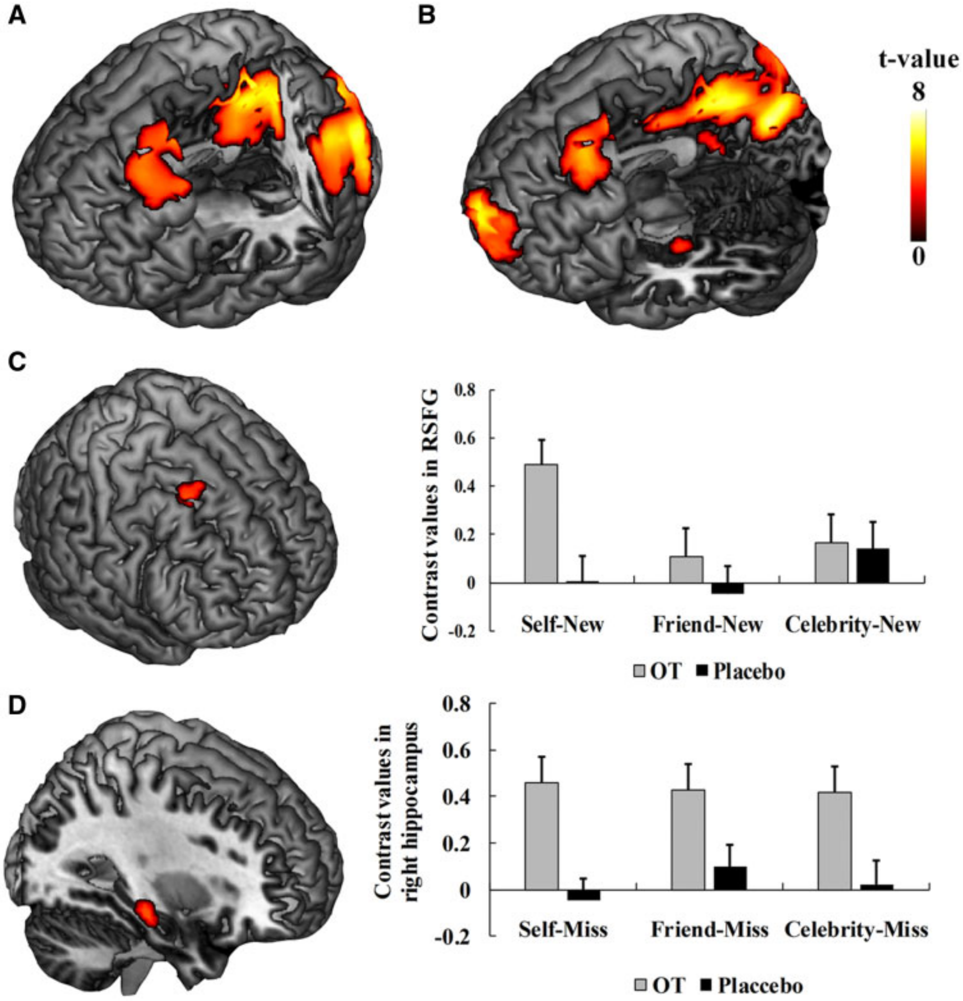

Fig. 5. fMRI results in Experiment 3. (A) Brain regions showed significant activation using old words minus new words. (B) Brain regions showed significant activation using corrected recognized old *vs* missed old words. (C) Brain regions showed significant OT effect (OT > placebo) on self-related memory retrieval (*vs* new words) and the contrast values of self-/friend-/celebrity-related old words *vs* new words in OT and placebo groups. (D) Brain regions showed significant OT effect (OT > placebo) for corrected recognized old *vs* missed old words and the contrast values of self-/friend-/celebrity-related hit words *vs* miss words in OT and placebo groups. RSFG, right superior frontal gyrus; OT, oxytocin.

Discussion

The current work investigated the functional roles of OT in self-referential processing by integrating OT/placebo administration and the self-referential task (Rogers *et al.*, 1977). Both behavioral and fMRI results suggest that OT (*vs* placebo) treatments produced opposite effects on the encoding and retrieval processes of self-related information. Experiment 1 revealed that OT (*vs* placebo) treatments decreased participants' response times during trait judgments of the self but not during trait judgments of a friend and a celebrity. The OT effects during the encoding phase of self-referential processing, however, did not result in deteriorated memory performance on self-related information during the surprising memory test. In contrast, OT (*vs* placebo) treatments enhanced memory retrieval of trait adjectives related to the self, which was manifested in the fact that OT *vs* placebo treatments tended to increase the *R* score but decrease the *K* score of trait adjectives used for self-judgments. The measure of remember/know responses was initially designed to distinguish between two conscious states of awareness associated with memory retrieval (Tulving, 1985, 1999). It is believed that a remember judgment is made when participants are able to consciously recollect information associated with the item's original presentation, whereas a known response is made when participants feel familiar with a word but cannot consciously recollect contextual details regarding the item's original presentation. Later studies, however, suggest that remember and know judgments do not reflect two qualitatively different memory processes (Donaldson, 1996) and both recollection and familiarity may be continuous signal-detection processes (Wixted and Mickes, 2010). Although these models have different viewpoints regarding whether or not recollection and familiarity are two different processes, it is commonly agreed that a higher criteria of memory strength is required for remembering judgments than for knowing judgments. Thus applying this proposition to our behavioral results, if a larger *R* score reflects better remember of specific details of the items encoded whereas a greater *K* score reflects stronger feeling of knowing or familiarity but less recollective experience with the items encoded, our results suggest that OT (*vs* placebo) treatments seemed to enhance memory retrieval of self-related information by increasing recollection of specific details of the encoded items and decreasing the feeling of knowing or familiarity associated with the encoded items. Interestingly, the OT effects on

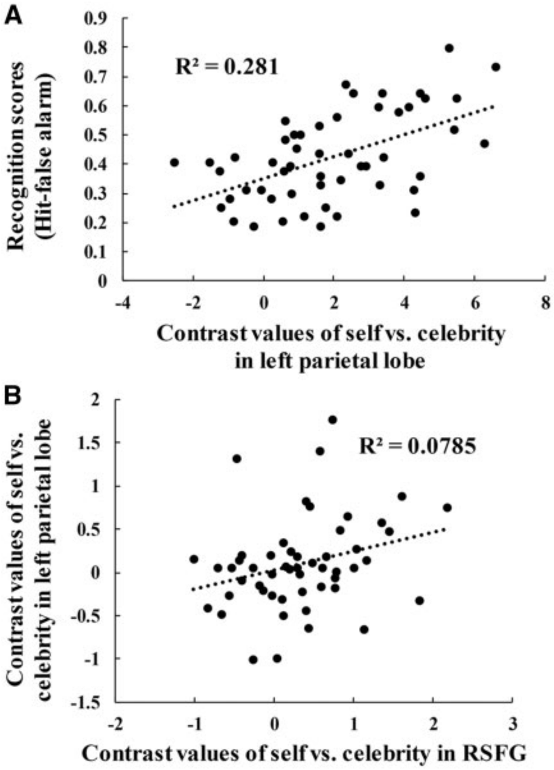

Fig. 6. Correlations in Experiment 3. (A) positive correlation of recognition scores and contrast values of self *vs* celebrity in left parietal lobe and (B) positive correlation of contrast values of self *vs* celebrity in left parietal lobe and contrast values of self *vs* celebrity in RSFG. RSFG, right superior frontal gyrus.

behavioral performance during both the encoding and retrieval phases were specific to the processing of self-related trait adjectives during the self-referential task. The behavioral results allowed expectation of distinct patterns of OT modulations of brain activities involved in encoding and retrieval of self-related items in the self-referential task.

Indeed, Experiment 2 showed that OT (*vs* placebo) administration modulated brain activity involved in self-referential processing in two aspects. First, while trait judgments of the self *vs* a celebrity activated the MPFC in the placebo group, the vMPFC/dMPFC activities underlying trait judgments of the self were significantly reduced in the OT *vs* placebo group. In addition, the OT effect was evident for self-judgments but not for trait judgments of a friend and a celebrity, suggesting self-specific OT effect on the neural correlates of self-referential processing. The OT-induced decreased MPFC activity engaged in trait judgments of the self is consistent with the OT effect on response times observed in Experiment 1. Both the OT effects on self-relevant behavioral performance and brain activity suggest that OT administration allowed less involvement of neural and mental resources during self-reflection on personality traits. The fMRI results in Experiment 2 are consistent with the results of our previous ERP research (Liu *et al.*, 2013) and indicate that the OT effect on self-referential processing did not arise from general OT effects on person or semantic processing involved in trait judgments. The MPFC has been shown to be engaged in the processing of self-related information in various paradigms. Besides the finding of increased MPFC activities during self-reflection of one's own personality traits (Kelley *et al.*, 2002; Macrae *et al.*, 2004; Zhu *et al.*, 2007; Ma and Han, 2011; Ma *et al.*, 2014), the MPFC was also activated when perceiving a morphed face perceived as one's own (*vs* others') face (Ma and Han, 2012), responding to agreement (*vs* disagreement) during self-relevant decision-making (Dong *et al.*, 2016), reflecting on one's own future-oriented core values (*vs* everyday activities) (Cascio *et al.*, 2016), perceiving a shape as being self-associated (*vs* other-associated) (Sui *et al.*, 2013), and being engaged in inward (*vs* outward) attention (Boehme *et al.*, 2015). The MPFC activity was decreased during cognitively demanding tasks that require attention to the external world in healthy adults (Greicius *et al.*, 2003) but not in individuals with autism who are characterized by excessive self-focus (Kennedy *et al.*, 2006). There is also evidence for the association between the resting state MPFC activity and the processing of the self (Qin and Northoff, 2011; Huang *et al.*, 2016). These findings indicate a key role of the MPFC in self-oriented processes or internally oriented attention. Our findings of decreased MPFC activity underlying self-referential processing by OT administration are consistent with the OT effects on response times to self-judgments and provide a potential neural mechanism of how intranasal OT administration modulates encoding of self-relevant items. We also found that OT (*vs* placebo) administration increased the activities in the bilateral frontal and parietal cortices, a neural network that is engaged in attention toward the external world and cognitively demanding tasks (Hopfinger *et al.*, 2000; Corbetta and Shulman, 2002). These results together suggest a functional role of OT in shifting neurocognitive resources between self-oriented and other-oriented cognitive processes during the self-referential task.

The second interesting finding of Experiment 2 was that, while the MPFC activity underlying self-referential processing was negatively correlated with participants' cultural trait (i.e. interdependence) in the placebo group, the coupling between the MPFC activity and cultural orientation was weakened by intranasal OT administration. The association between the MPFC activity underlying self-referential processing and interdependence was reported in the previous research (Ma *et al.*, 2014). Together with the finding of group differences in the MPFC activity and interdependence between East Asians and Westerners (e.g. Ma *et al.*, 2014), it has been argued that cultural experiences and environments shape the brain activity involved in self-related processing (Han *et al.*, 2013; Han and Ma, 2015a) and the coupling between brain activity and cultural traits may facilitate individuals' social interactions in their own cultural environment. However, a rigid coupling between brain activity and cultural traits does not allow an individual to quickly fit into a new cultural environment and to adapt to various social interactions. Thus it is likely that OT may decouple the association between brain activity and cultural traits so that the brain is able to work in a more flexible way during social interaction. This proposition is consistent with the proposition that OT facilitates social adaption that requires the brain to work flexibly in response to cognitive tasks and social environments (Ma *et al.*, 2016a).

Experiment 3 revealed OT effects on the brain activity supporting memory retrieval of self- and other-related information. The contrast of old *vs* new words during memory retrieval revealed activations in the bilateral parietal cortex, precuneus, and left superior frontal cortex. These are consistent with the results of a meta-analysis of the old/new effects during memory retrieval (Kim, 2013, 2016). The contrast of old words that were recognized *vs* missed during the retrieval phase uncovered activations in the precuneus/PCC, MPFC, left superior frontal gyrus, bilateral parahippocampal gyrus, and right hippocampus. The previous study also reported that recognized *vs* missed old words during the retrieval phase were associated with greater

MPFC activity during encoding of self-related trait adjectives (Macrae *et al.*, 2004), which also predicted better memory of self-related information during the retrieval phase (Ma and Han, 2011). Thus the MPFC activities at both the encoding and retrieval phases seem to differentiate between recognized *vs* missed worlds. The hippocampus and parahippocampus activations observed here are also consistent with the widely acknowledged role of these brain regions in memory retrieval (e.g. Maguire *et al.*, 2001; Yonelinas *et al.*, 2001; Zeidman and Maguire, 2016). Interestingly, although the whole brain ANOVA did not show evidence for distinct OT effects on brain activities underlying memory retrieval of trait adjectives related to the self and others, the results of two sample analyses suggest that intranasal OT *vs* placebo tended to increase the RSFG activity specifically in response to self-related old *vs* new words during the retrieval phase. How did the OT administration facilitate memory retrieval? The correlation analyses suggest a possible pathway from the OT modulated RSFG activity to the left parietal activity because the RSFG activity was positively correlated with the left parietal activity which further positively predicted participants' recognition scores. However, the correlation results cannot demonstrate a causal relationship between the increased RSFG activity and enhanced memory performance and the neural pathway through which OT administration facilitates memory retrieval of self-related information should be verified by future research.

Taken together, our behavioral and neuroimaging results provide evidence for modulations of neurocognitive processes of self-related information during both the encoding and retrieval phases of the self-referential task. Our findings complement the previous behavioral and neuroimaging studies of OT effects on other-oriented or externally oriented mental and neural processes, and reinforce our understanding of the functional roles of OT in social communication and interaction. Our findings suggest two new mechanisms through which OT may facilitate successive social interactions, i.e. to modulate encoding and retrieval of self-related information. It appears that OT modulates the neurocognitive strategies at different stages of the processing of self-related information so that individuals can adjust behaviors for successive social interactions. The previous fMRI study reported results that tended to be consistent with our findings but did not identify self-specific significant OT effects (Zhao *et al.*, 2016). This might owe to several differences between the previous and the current studies. For instance, relative to the current work, the previous work tested a smaller sample ($n \leq 20$ in the OT or place group) during trait judgments of more target persons (i.e. self, mother, a class-mate, a friend and a stranger). Moreover, the brain activities underlying self-referential processing were recorded during the encoding phase but not during the retrieval phase. Future research should clarify how the differences in the experimental design and the testing sample size influence the observation of reliable OT effects on the neurocognitive processes of self-related information.

The OT effects on both behavioral performance and brain activity related to self-referential processing are consistent with the social salience hypothesis of OT function (Bartz *et al.*, 2011; Shamay-Tsoory and Abu-Akel, 2016). If OT facilitates attention orienting to external contextual social cues by salience assignment, OT then should produce opposite effects on responses to internal self-related cues and brain activities underlying inward attention. Our findings support this analysis by showing that OT (*vs* placebo) led to shorter RTs during self-judgments, decreased mPFC activity that is associated with inward attention (Boehme *et al.*, 2015) but increased frontal/parietal activities that mediate attention toward the external world and cognitively demanding tasks (Hopfinger *et al.*, 2000; Corbetta and Shulman, 2002). It appears that OT produce opposite effects on shifting outward attention to the external social cues of contexts and other people and inward attention to information related to oneself.

The current work had a few limitations that should be acknowledged. First, the current work tested OT effects on neurocognitive processes of self-related information only in male participants. Recent studies have revealed opposite OT effects on other-oriented processing in male and female participants. For example, OT (*vs* placebo) administration decreased amygdala responses to emotional faces (Kirsch *et al.*, 2005; Domes *et al.*, 2007) but enhanced amygdala activity in response to fearful faces and threatening pictures in women (Domes *et al.*, 2010; Wittfoth-Schardt *et al.*, 2012). Our findings left an open question of whether OT modulates memory encoding and retrieval of self-related information in a similar vein in male and female adults. Second, self-referential processing may recruit distinct neural substrates depending on the domains of personal attributes (e.g. Ma *et al.*, 2014). The processing of one's own images (e.g. face) recruits the frontal lobe such as the MPFC as well as the occipito-temporal cortices (e.g. the fusiform gyrus) (Ma and Han, 2012; Hu *et al.*, 2016). Thus it is interesting to examine OT effects on neural correlates of the processing of self-related information in other domains. Third, there has been ample evidence that the neural correlates of self-referential processing are sensitive to individuals' cultural experiences (Han *et al.*, 2013; Han and Ma, 2015b). Moreover, the OT effects on both brain (Liu *et al.*, 2013) and behavioral (Pfundmair *et al.*, 2014) responses varied across individuals with their cultural orientations such as interdependence. Our current study tested only Chinese participants who are dominated by interdependent self-construals (e.g. Ma *et al.*, 2014). Thus it is essential to clarify whether OT effects on neurocognitive processes of self-related information are independent of individuals' cultural orientations in future research. Finally, although the OT effects on memory performances in Experiment 1 and brain activities in Experiments 2 and 3 appeared to be consistent, the statistical power is always an important issue for studies of intranasal manipulation of OT due to small sample sizes in behavioral and brain imaging research (Walum *et al.*, 2016). Increasing sample size and replication should be considered in future research of OT effects on neurocognitive processes involved in social cognition.

In conclusion, the current work showed both behavioral and neuroimaging evidence that intranasal administration of OT *vs* place modulates encoding and retrieval of self-related information in the self-referential task. Our findings expand the previous OT studies by showing that OT plays an important role in both sides of social cognition, i.e. the processing of other-related and self-related information. The OT effects on both sides of social cognition promote social interactions and facilitate social adaptation.

Supplementary data

Supplementary data are available at *SCAN* online.

Funding

This work was supported by the National Natural Science Foundation of China (Projects 91332125, 31421003, 31470986). We thank B. Li for preparing OT and placebo.

Conflict of interest. None declared.

References

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*, **7**(4), 268–77.

Aron, A., Aron, E.N., Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *J Person Soc Psychol*, **63**(4), 596–612.

Arueti, M., Perach-Barzilay, N., Tsoory, M.M., Berger, B., Getter, N., Shamay-Tsoory, S.G. (2013). When two become one: the role of oxytocin in interpersonal coordination and cooperation. *J Cogn Neurosci*, **25**(9), 1418–27.

Banaji, M.R., Prentice, D.A. (1994). The self in social contexts. *Annu Rev Psychol*, **45**(1), 297–332.

Bartz, J.A., Zaki, J., Bolger, N., Ochsner, K.N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends Cogn Sci*, **15**(7), 301–9.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron*, **58**(4), 639–50.

Bedny, M., Pascual-Leone, A., Saxe, R.R. (2009). Growing up blind does not change the neural bases of theory of mind. *Proc Natl Acad Sci USA*, **106**(27), 11312–7.

Boehme, S., Miltner, W.H., Straube, T. (2015). Neural correlates of self-focused attention in social anxiety. *Soc Cogn Affect Neurosci*, **10**(6), 856–62.

Buckner, R.L., Koutstaal, W., Schacter, D.L., Wagner, A.D., Rosen, B.R. (1998). Functional-anatomic study of episodic retrieval using fMRI. I. Retrieval effort versus retrieval success. *NeuroImage*, **7**(3), 151–62.

Cascio, C.N., O'Donnell, M.B., Tinney, F.J., *et al.* (2016). Self-affirmation activates brain systems associated with self-related processing and reward and is reinforced by future orientation. *Soc Cogn Affect Neurosci*, **11**(4), 621–9.

Corbetta, M., Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*, **3**(3), 201–15.

Cross, S.E., Vick, N.V. (2001). The interdependent self-construal and social support: the case of persistence in engineering. *Person Soc Psychol Bull*, **27**(7), 820–32.

Declerck, C.H., Boone, C., Kiyonari, T. (2014). The effect of oxytocin on cooperation in a prisoner's dilemma depends on the social context and a person's social value orientation. *Soc Cogn Affect Neurosci*, **9**(6), 802–9.

De Dreu, C.K. (2012). Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda. *Horm Behav*, **61**(3), 419–28.

De Dreu, C.K.W., Greer, L.L., Handgraaf, M.J.J., *et al.* (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, **328**(5984), 1408–11.

Ditzen, B., Schaer, M., Gabriel, B., Bodenmann, G., Ehlert, U., Heinrichs, M. (2009). Intranasal oxytocin increases positive communication and reduces cortisol levels during couple conflict. *Biol Psychiatry*, **65**(9), 728–31.

Domes, G., Heinrichs, M., Gläscher, J., Büchel, C., Braus, D.F., Herpertz, S.C. (2007). Oxytocin attenuates amygdala responses to emotional faces regardless of valence. *Biol Psychiatry*, **62**(10), 1187–90.

Domes, G., Heinrichs, M., Michel, A., Berger, C., Herpertz, S.C. (2007). Oxytocin improves "mind-reading" in humans. *Biol Psychiatry*, **61**(6), 731–3.

Domes, G., Lischke, A., Berger, C., *et al.* (2010). Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology*, **35**(1), 83–93.

Donaldson, Z.R., Young, L.J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science*, **322**(5903), 900–4.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Mem Cogn*, **24**(4), 523–33.

Dong, S.Y., Kim, B.K., Lee, S.Y. (2016). Implicit agreeing/disagreeing intention while reading self-relevant sentences: a human fMRI study. *Soc Neurosci*, **11**(3), 221–32.

Frith, C.D., Frith, U. (2006). The neural basis of mentalizing. *Neuron*, **50**(4), 531–4.

Gamer, M., Zurowski, B., Büchel, C. (2010). Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc Natl Acad Sci USA*, **107**(20), 9400–5.

Gardner, W.L., Gabriel, S., Hochschild, L. (2002). When you and I are "we", you are not threatening: the role of self-expansion in social comparison. *J Person Soc Psychol*, **82**(2), 239–51.

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc Natl Acad Sci USA*, **100**(1), 253–8.

Grimm, S., Pestke, K., Feeser, M., *et al.* (2014). Early life stress modulates oxytocin effects on limbic system during acute psychosocial stress. *Soc Cogn Affect Neurosci*, **9**(11), 1828–35.

Gordon, I., Vander Wyk, B.C., Bennett, R.H., *et al.* (2013). Oxytocin enhances brain function in children with autism. *Proc Natl Acad Sci USA*, **110**(52), 20953–8.

Groppe, S.E., Gossen, A., Rademacher, L., *et al.* (2013). Oxytocin influences processing of socially relevant cues in the ventral tegmental area of the human brain. *Biol Psychiatry*, **74**(3), 172–9.

Gutchess, A.H., Kensinger, E.A., Schacter, D.L. (2010). Functional neuroimaging of self-referential encoding with age. *Neuropsychologia*, **48**(1), 211–9.

Han, S., Ma, Y. (2015a). A culture-behavior-brain loop model of human development. *Trends Cogn Sci*, **9**(11), 666–76.

Han, S., Ma, Y. (2015b). Cultural neuroscience studies of self-reflection. In: Chiao, J., Li, S.-C., Seligman, R., Turner, R. editors. *The Oxford Handbook of Cultural Neuroscience*. New York: Oxford University Press, pp. 197–208.

Han, S., Mao, L., Gu, X., Zhu, Y., Ge, J., Ma, Y. (2008). Neural consequences of religious belief on self-referential processing. *Soc Neurosci*, **3**(1), 1–15.

Han, S., Northoff, G., Vogeley, K., Wexler, B.E., Kitayama, S., Varnum, M.E.W. (2013). A cultural neuroscience approach to the biosocial nature of the human brain. *Annu Rev Psychol*, **64**(1), 335–59.

Hopfinger, J.B., Buonocore, M.H., Mangun, G.R. (2000). The neural mechanisms of top-down attentional control. *Nat Neurosci*, **3**(3), 284–91.

Hu, C., Di, X., Eickhoff, S.B., *et al.* (2016). Distinct and common aspects of physical and psychological self-representation in the brain: a meta-analysis of self-bias in facial and self-referential judgments. *Neurosci Biobehav Rev*, **61**(2), 197–207.

Hu, J., Qi, S., Becker, B., *et al.* (2015). Oxytocin selectively facilitates learning with social feedback and increases activity and functional connectivity in emotional memory and reward processing regions. *Hum Brain Mapp*, **36**(6), 2132–46.

Huang, Z., Obara, N., Davis, H.H., Pokorny, J., Northoff, G. (2016). The temporal structure of resting-state brain activity in the medial prefrontal cortex predicts self-consciousness. *Neuropsychologia*, **82**(2), 161–70.

Insel, T.R. (1992). Oxytocin—a neuropeptide for affiliation: evidence from behavioral, receptor autoradiographic, and comparative studies. *Psychoneuroendocrinology*, **17**(1), 3–35.

Kanat, M., Heinrichs, M., Schwarzwald, R., Domes, G. (2015). Oxytocin attenuates neural reactivity to masked threat cues from the eyes. *Neuropsychopharmacology*, **40**(2), 287–95.

Kelley, W.M., Macrae, C.N., Wyland, C.L., Caglar, S., Inati, S., Heatherton, T.F. (2002). Finding the self? An event-related fMRI study. *J Cogn Neurosci*, **14**(5), 785–94.

Kennedy, D.P., Redcay, E., Courchesne, E. (2006). Failing to deactivate: resting functional abnormalities in autism. *Proc Natl Acad Sci USA*, **103**(21), 8275–80.

Keverne, E.B., Kendrick, K.M. (1992). Oxytocin facilitation of maternal behavior in sheep. *Ann N Y Acad Sci*, **652**(1), 83–101.

Kim, H. (2013). Differential neural activity in the recognition of old versus new events: an activation likelihood estimation meta-analysis. *Hum Brain Mapp*, **34**(4), 814–36.

Kim, H. (2016). Default network activation during episodic and semantic memory retrieval: a selective meta-analytic comparison. *Neuropsychologia*, **80**(1), 35–46.

Kirsch, P., Esslinger, C., Chen, Q., *et al.* (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *J Neurosci*, **25**(49), 11489–93.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, **435**(7042), 673–6.

Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. *Ann Rev Psychol*, **58**, 259–89.

Liu, Y. (1990). *Modern Lexicon of Chinese Frequently-Used Word Frequency*. Beijing: Space Navigation Press.

Liu, Y., Sheng, F., Woodcock, K.A., Han, S. (2013). Oxytocin effects on neural correlates of self-referential processing. *Biol Psychol*, **94**(2), 380–7.

Ma, Y., Han, S. (2011). Neural representation of self-concept in sighted and congenitally blind adults. *Brain*, **134**(Pt 1), 235–46.

Ma, Y., Han, S. (2012). Functional dissociation of the left and right fusiform gyrus in self-face recognition. *Hum Brain Mapp*, **33**(10), 2255–67.

Ma, Y., Li, B., Wang, C., *et al.* (2014). 5-HTTLPR polymorphism modulates neural mechanisms of negative self-reflection. *Cereb Cortex*, **24**(9), 2421.

Ma, Y., Li, S., Wang, C., *et al.* (2016b). Distinct oxytocin effects on belief updating in response to desirable and undesirable feedback. *Proc Natl Acad Sci USA*, **113** (33), 9256–61.

Ma, Y., Liu, Y., Rand, D.G., Heatherton, T.F., Han, S. (2015). Opposing oxytocin effects on intergroup cooperative behavior in intuitive and reflective minds. *Neuropsychopharmacology*, **40**(10), 2379–87.

Ma, Y., Shamay-Tsoory, S., Han, S., Zink, C.F. (2016a). Oxytocin and social adaptation: insights from neuroimaging studies of healthy and clinical populations. *Trends Cogn Sci*, **20**(2), 133–45.

Macrae, C.N., Moran, J.M., Heatherton, T.F., Banfield, J.F., Kelley, W.M. (2004). Medial prefrontal activity predicts memory for self. *Cereb Cortex*, **14**(6), 647–54.

Maguire, E.A., Vargha-Khadem, F., Mishkin, M. (2001). The effects of bilateral hippocampal damage on fMRI regional activations and interactions during memory retrieval. *Brain*, **124**(6), 1156–70.

Markus, H.R., Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychol Rev*, **98**(2), 224–53.

Meyer-Lindenberg, A., Domes, G., Kirsch, P., Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat Rev Neurosci*, **12**(9), 524–38.

Mikolajczak, M., Pinon, N., Lane, A., de Timary, P., Luminet, O. (2010). Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance. *Biol Psychol*, **85**(1), 182–4.

Mu, Y., Guo, C., Han, S. (2016). Oxytocin enhances inter-brain synchrony during social coordination in male adults. *Soc Cogn Affect Neurosci*, **11**(12), 1882–93.

Mu, Y., Han, S. (2010). Neural oscillations involved in self-referential processing. *NeuroImage*, **53**(2), 757–68.

Nave, G., Camerer, C., McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspect Psychol Sci*, **10**(6), 772–89.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., Panksepp, J. (2006). Self-referential processing in our brain–a meta-analysis of imaging studies on the self. *NeuroImage*, **31**(1), 440–57.

Pedersen, C.A., Ascher, J.A., Monroe, Y.L., Prange, A.J. (1982). Oxytocin induces maternal behavior in virgin female rats. *Science*, **216**(4546), 648–50.

Petrovic, P., Kalisch, R., Singer, T., Dolan, R.J. (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J Neurosci*, **28**(26), 6607–15.

Pfundmair, M., Aydin, N., Frey, D., Echterhoff, G. (2014). The interplay of oxytocin and collectivistic orientation shields against negative effects of ostracism. *J Exp Soc Psychol*, **55**(11), 246–51.

Qin, P., Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, **57**(3), 1221–33.

Rilling, J.K., DeMarco, A.C., Hackett, P.D., *et al.* (2012). Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology*, **37**(4), 447–61.

Rogers, T.B., Kuiper, N.A., Kirker, W.S. (1977). Self-reference and the encoding of personal information. *J Person Soc Psychol*, **35**(9), 677–88.

Rosenberg, M. (1995). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University.

Rugg, M.D., Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn Sci*, **11**(6), 251–7.

Saxe, R., Moran, J.M., Scholz, J., Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Soc Cogn Affect Neurosci*, **1**(3), 229–34.

Scheele, D., Kendrick, K.M., Khouri, C., *et al.* (2014). An oxytocin-induced facilitation of neural and emotional responses to social touch correlates inversely with autism traits. *Neuropsychopharmacology*, **39**(9),2078–85.

Shamay-Tsoory, S.G., Fischer, M., Dvash, J., Harari, H., Perach-Bloom, N., Levkovitz, Y. (2009). Intranasal administration of oxytocin increases envy and schadenfreude (gloating). *Biol Psychiatry*, **66**(9), 864–70.

Shamay-Tsoory, S.G., Abu-Akel, A. (2016). The social salience hypothesis of oxytocin. *Biol Psychiatry*, **79**(3), 194–202.

Shalvi, S., De Dreu, C.K. (2014). Oxytocin promotes group-serving dishonesty. *Proc Natl Acad Sci USA*, **111**(15), 5503–7.

Sheng, F., Liu, Y., Zhou, B., Zhou, W., Han, S. (2013). Oxytocin modulates the racial bias in neural responses to others' suffering. *Biol Psychol*, **92**(2), 380–6.

Singelis, T.M. (1994). The measurement of independent and interdependent self-construals. *Person Soc Psychol Bull*, **20**(5), 580–91.

Slotnick, S.D., Moo, L.R., Segal, J.B., Hart Jr, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cogn Brain Res*, **17**(1), 75–82.

Stavropoulos, K.K., Carver, L.J. (2013). Research review: social motivation and oxytocin in autism-implications for joint attention development and intervention. *J Child Psychol Psychiatry*, **54**(6), 603–18.

Symons, C.S., Johnson, B.T. (1997). The self-reference effect in memory: a meta-analysis. *Psychol Bull*, **121**(3), 371–94.

Sui, J., Rotshtein, P., Humphreys, G.W. (2013). Coupling social attention to the self forms a network for personal significance. *Proc Natl Acad Sci USA*, **110**, 7607–7612.

Triandis, H.C. (1989). The self and social behavior in differing cultural contexts. *Psychol Rev*, **96**(3), 506–20.

Tulving, E. (1985). Memory and consciousness. *Can Psychol*, **26**(1), 1–12.

Tulving, E. (1999). On the uniqueness of episodic memory. In: Nilsson, L.G., Markowitsch, H.J., editors. *Cognitive Neuroscience of Memory*. Göttingen: Hogrefe and Huber Publishers, pp. 11–42.

Van IJzendoorn, M.H., Bakermans-Kranenburg, M.J. (2012). A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology*, **37**(3), 438–43.

Wittfoth-Schardt, D., Gründing, J., Wittfoth, M., *et al.* (2012). Oxytocin modulates neural reactivity to children's faces as a function of social salience. *Neuropsychopharmacology*, **37**(8), 1799–807.

Wixted, J.T., Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychol Rev*, **117**(4), 1025–54.

Walum, H., Waldman, I.D., Young, L.J. (2016). Statistical and methodological considerations for the interpretation of intranasal oxytocin studies. *Biol Psychiatry*, **79**(3), 251–7.

Yonelinas, A.P., Hopfinger, J.B., Buonocore, M.H., Kroll, N.E.A., Baynes, K. (2001). Hippocampal, parahippocampal and occipital-temporal contributions to associative and item recognition memory: an fMRI study. *Neuroreport*, **12**(2), 359–63.

Young, L.J., Wang, Z. (2004). The neurobiology of pair bonding. *Nat Neurosci*, **7**(10), 1048–54.

Zak, P.J., Stanton, A.A., Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS One*, **2**(11), e1128.

Zeidman, P., Maguire, E.A. (2016). Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nat Rev Neurosci*, **17**(3), 173–82.

Zhao, W., Yao, S., Li, Q., *et al.* (2016). Oxytocin blurs the self-other distinction during trait judgments and reduces medial prefrontal cortex responses. *Hum Brain Mapp*, **37**(7), 2512–27.

Zhu, Y., Zhang, L., Fan, J., Han, S. (2007). Neural basis of cultural influence on self-representation. *NeuroImage*, **34**(3), 1310–6.