## RESEARCH PAPERS

# Neuroimaging studies of self-reflection[*]

ZHU Ying[**]

(Department of Psychology, Peking University, Beijing 100871, China)

**Abstract**    This paper reviews some basic findings and methodological issues in neuroimaging studies of self-referential processing. As a general rule, making judgments about one's self, inclusive of personality trait adjectives or current mental states (person's preferences, norms, aesthetic values and feeling) uniformly generates medial prefrontal activations, regardless of stimulus materials (words or pictures) and modality (visual or auditory). Cingulate activations are also observed in association with most self-referential processing. Methodological issues include treating self-referential processing as either representing one's own personality traits or representing one's own current mental states. Finally, self-referential processing could be considered as implement of "I think therefore I am" approach to neuroimaging the self.

**Keywords:   self, self-referential processing, medial prefrontal cortex, cingulate, brain image.**

Self-referential processing is an important aspect of processing the information about the self. There are two types of self-referential processing in terms of neuroimaging studies of self-referential processing.

First, self-referential processing refers to the processing of personality trait adjectives with reference to self or making judgments about one's own personality traits. A number of behavioral studies have shown that words processed with reference to the self are better remembered than words processed in other ways, including semantic processing[1,2]. For example, the word "friendly" is better remembered after answering the self-referential question "Does the word 'friendly' describe you?" than after answering the semantic processing question "Does 'friendly' mean the same as 'kindly'?" Why does this self-reference effect emerge? It has been long debated by cognitive and social psychologists. However, data from purely behavioral studies cannot be used to solve the problem[3,4]. There has been now neuroimaging studies aiming to find if there are special neural correlates of self-referential processing[4~7].

Second, self-referential processing refers to making judgments about one's own present mental states after reading many different questions about knowledge of one's self. For example, Zysset et al.[8] pre-sented subjects with a series of statements such as "I like Leipzig", "I enjoy going to New Year's parties", and "Gerhard Schroder is a good chancellor" and asked them to make a yes or no answer to the statements one by one. This type of self-referential processing involves various aspects of the self, such as person's value system (e.g. preferences, norms and aesthetic values, see [8]), pleasant and unpleasant feeling[9], and abilities, traits and attitudes[10]. Researchers also use pictures in this type of study of self-referential processing[9]. Therefore, the second type of self-referential processing is more feasible than the first one in neuroimaging the self.

All known (nine) studies are listed in Tables 1 and 2. Table 1 lists the numbers of participants, materials used in experiment, and scanning conditions of each study. Table 2 summarizes the type of self-referential processing, contrast relating to the self, and brain areas yielded by self-referential processing in the contrast. The major findings and their significance will be discussed in the next section.

## 1  Medial prefrontal cortex and cingulate

It is surprising that previous studies found that self-referential processing uniformly yield medial prefrontal activations (see Table 2). These medial pre-

frontal cortices are localized in Brodmman areas 9 and 10 (BA9 and BA10, see Fig. 1). We should note that this general pattern appears no matter what materials (words, sentences or pictures) are used, what modalities are used (visual or auditory), who (American, Canadian, German, Japanese or Chinese) are the participants, and what design paradigms (block-design or event-related design) are used.

Table 1.   Summary of the nine studies

| Study | $n$ | Materials | Modality | Scanning conditions |
|---|---|---|---|---|
| Craik et al., 1999; PET[5] | 8 | Trait adjectives | Visual | 4 conditions: 1, self-referential encoding ("How well does the adjective describe you?" —— a 4-point scale); 2, other-referential encoding; 3, semantic encoding; 4, syllable encoding ("How many syllables does the adjective contain?"). |
| Kelley et al., 2002; fMRI[4] | 24 | Trait adjectives | Visual | 3 conditions: 1, self-referential encoding ("Does this adjective describe you?"); 2, other-referential encoding; 3, case encoding ("Is this adjective printed in uppercase letter?"). |
| Li Zhang et al., 2003; fMRI[6] | 14 | Trait adjectives (Chinese) | Visual | 5 conditions: 1, self-referential encoding ("Does this adjective describe you?"); 2, other-referential encoding; 3, mother-referential encoding ("Does the trait adjective describe your mother?"); 4 semantic encoding; 5, phonetic encoding ("How many Chinese first tone does the adjective contain?"). |
| Horinchi et al., 2002; e-fMRI[7] | 6 | Trait adjectives (Japanese) | Visual | 3 conditions: 1, self-referential encoding ("Does this word describe you?"); 2, semantic encoding; 3, physical encoding ("Does this word have closed parts?"). |
| Kircher et al., 2000; fMRI[11] | 6 | Trait adjectives (German) | Visual | 2 conditions: 1, self describing condition ("Does this adjective describe how you typically feel and think about yourself?"); 2, non self describing condition.<br>The words describing personal traits and physical features were rated by the subjects 6 weeks prior to imaging data acquisition and they were categorized into self describing and non self describing. During scanning subjects had to indicate whether a word was self-descriptive or not by means of a button pressing. |
| Kircher et al., 2002; fMRI[12] | 6 | Trait adjectives (German) | Visual | **Experiment 1:**<br>2 conditions: 1, intentional self describing condition ("Does this adjective describe how you typically feel and think about yourself?"); 2, non self describing condition.<br>The words describing personal traits and physical features were rated by the subjects 6 weeks prior to imaging data acquisition and they were categorized into self describing and non self describing. During scanning subjects had to indicate whether a word was self-descriptive or not by means of a button pressing.<br>**Experiment 2:**<br>2 conditions: 1, incidental self description condition; 2, incidental non self description condition.<br>The words describing personal traits and physical features were rated by the subjects 6 weeks prior to imaging data acquisition and they were categorized into self describing and non self describing. During scanning subjects were requested to indicate by button pressing whether the word described a physical or psychological feature, while unaware that the words had been arranged in blocks according to self-descriptiveness. |
| Johnson et al., 2002, fMRI[10] | 11 | Sentences | Auditory | 2 conditions: 1, experimental condition (e. g., "I forget important things."); 2, semantic condition (e. g., "You need water to live"). |
| Zysset et al., 2002; fMRI[8] | 13 | Sentences (German) | Visual | 4 conditions: 1, evaluative judgment (e. g., "I like Leipzig"); 2, semantic memory retrieval; 3, episodic memory retrieval; 4, baseline condition (press the left or right button). |
| Gusnard et al., 2001; fMRI[9] | 24 | Color pictures | Visual | 2 conditions: 1, the internally cued condition (emotional pictures which induce pleasant or unpleasant feeling); 2, the externally cued condition (the pictures depicted an indoor or outdoor scene). |

Table 2.　Brain areas active during self-referential processing

| Study | Type of self-referential processing | Contrast | Prefrontal cortex L | Prefrontal cortex Medial | Prefrontal cortex R | Cingulate | Other |
|---|---|---|---|---|---|---|---|
| Craik et al. (1999), PET[5] | Type 1: Making judgment about one's own personal traits | Self-referential processing minus other-referential processing, semantic processing and syllable encoding | | BA10( -6,56,8) BA9(6,40,28) | BA10(30,6020) BA45(52,26,4) | | |
| Kelley et al. (2002), e-fMRI[4] | Type 1 | Self-referential processing minus other referential processing | | BA10(10,52,2) | | BA7(12, -48, 50) | |
| Li Zhang et al. fMRI[6] | Type 1 | Self-referential processing minus semantic processing | | BA9(4, 50, 26) BA10(2,52,12) | | | |
| | | Self-referential processing minus other processing | | BA9( -2,46,18) BA9(0,44,26) | | BA32(6,42,12) | |
| Horinchi et al. (2002), e-fMRI[7] | Type 1 | Self-referential processing minus semantic processing | | Medial frontal lobe | | | Thalamus and nucleus |
| Kircher et al. 2000; fMRI[11] | Type 1 | Self descriptors minus non-self descriptors | | | | BA23( -12, -22,31) BA31( -23, -67,9) | L: precuneus (BA7), superior parietal lobe(BA7), postcentral gyrus (BA40), fusiform gyrus (BA19) |
| Kircher et al. 2002; fMRI[12] | Type 1 | Intentional self minus non-self descriptors | BA44 ( -38,0,20) | | | BA23( -12, -22,31) BA31( -23, -67,9) | L: precuneus (BA7), postcentral gyrus (BA40) |
| | | Incidental self minus non-self descriptors | BA44 ( -46,3,31) | | | | R: middle temporal gyrus (BA21), inferior parietal lobe (BA40) L: superior temporal gyrus ( BA22 ), lingual gyrus (BA18) |
| Johnson et al. (2002), fMRI[10] | Type 2: Making judgment about one's own present mental states | Self-referential processing ( making judgment about abilities, traits and attitudes) minus semantic processing | | BA9/10(0,54,8) | | BA23,30 and 31 ( -2, -62, 32) | Thalamus, bilateral poster orbital cortex, bilateral inferior temporal gyrus, bilateral cerebellum |
| Zysset et al. (2002), fMRI[8] | Type 2 | Self-referential processing ( making judgment about person's preferences, norms, aesthetic values) minus semantic memory retrieval | BA45/47 ( -42,16, -4) | BA10/9 ( -6,55,13) | | BA23/31 ( -10, -51,36) | |
| Gusnard et al. (2001), fMRI[9] | Type 2 | Self-referential processing ( making judgment about pleasant vs. unpleasant pictures) minus semantic processing (making judgment about picture indoors vs. outdoors) | | BA8/9( -9,39,42) BA10( -3,53,24) BA6/8( -11,23,52) BA8( -11,30,44) BA9(7,45,25) | | BA32( -3,41,8) | |

However, Kircher et al.[11,12] did not observe medial prefrontal activations in association with the processing of the self. The basic difference between Kircher et al.'s studies and other seven studies shown in Table 2 is in the control condition used for comparison with self-referential processing. While non-self describing condition was used as the control condition in Kircher et al.'s work, other researchers employed other-referential processing or semantic processing as the control condition. Kircher et al. had subjects indicate whether a presented trait adjective was self descriptive or not by means of a button pressing. They
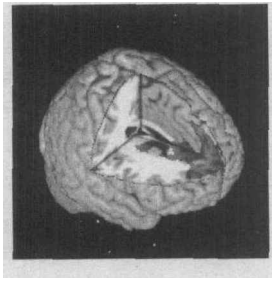
Fig. 1.   3D rendering of activations of medial frontal cortex (self minus semantic condition from Zhang Li, et al. 2003). MRIcro settings: positive range is from 2.00 to 13.65, minimum cluster size is 2.

used the result of self descriptors minus non-self descriptors to indicate self-referential processing. However, according to the levels-of-processing theory[13,14], making judgment about a word as non-self descriptive still fell on self-referential processing category. Therefore, we think self descriptors minus non-self descriptors could not lead to activations of the brain regions related to self-referential processing. In other words, the baseline was not proper in Kircher et al.'s work.

We want to emphasize here that medial prefrontal cortex may modulate self-referential processing. Until now there has been no neuroimaging evidence that shows other types of information processing about the self yield medial prefrontal activations except self-referential processing. First, recognition of one's own face does not yield activations of the medial prefrontal cortex[11,15,16]. Second, autobiographical memory retrieval does not either[17~19]. Finally, episodic memory retrieval which necessarily involves the concept of the self does not yield activations of the medial prefrontal cortex either[20]. Another interesting and important point related to medial prefrontal cortex is that, since medial prefrontal cortex engages self-referential processing including representing one's own personal traits and representing one's own current mental states, the two types of representations can be grouped together under the same catalogue from the cognitive neuroscience point of view. This point of view is different from the argument that representations of one's own personal traits and representations of one's own mental states are two subsystems of the unitary self composed of six subsystems based on social psychological and neuropsychological studies[21]. We may consider that representation of one's own personal traits is a very important part of one's own mental states for normal people.

As to cingulated cortex, it is also worth mentioning that three posterior and two anterior cingulate activations are associated with self-referential processing in seven studies excluding two studies of Kircher et al.[11,12] since their baseline is not proper. Although the medial prefrontal cortex may only modulate self-referential processing, cingulated activations are in common with recognition of one's own face (two anterior cingulated activations appear in Kircher et al.'s studies[11,15]), autobiographical event memories (two anterior and two posterior cingulated activations appear in 11 studies summarized in Maguire's paper[19]), and episodic memory retrieval (anterior cingulated activations appear in Lepage et al.'s work[20], which involves a total of 53 subjects), and self-referential processing besides. In other words, anterior or posterior cingulated may be more common brain areas than medial prefrontal cortex for the multifaceted self[22,23].

It is interesting in finding that self-referential processing activated medial prefrontal cortex and cingulated cortex are also activated by other mind reading. In answering the question "Do we activate the same brain regions to read our own and other minds?" Ross pointed out that "Neuroimaging studies show that theory of mind activity occurs in medial frontal cortex and paracingulate cortex for both kinds of mind reading"[24]. Since recognition of one's own face, autobiographical memory retrieval and episodic memory retrieval do not cause activations of the medial prefrontal cortex as mentioned above, self-referential processing may be more important than these processings of information about the self in relation to theory of mind.

## 2   A flexible paradigm in neuroimaging the self

Why does self-referential processing uniformly evoke medial prefrontal activations? It is possible that the activations result from the uniform experimental procedure used in the studies, i.e. participants are asked to make judgments about themselves or to think of themselves. In the first type of self-referential processing, participants make judgment about themselves as responses to personal trait adjectives presented; in the second type of self-referential processing, they make judgments about themselves as responses to statements about themselves presented.

Although making judgments about one's self is

the same for two types of self-referential processings, the second type of self-referential processing is more flexible than the first one in neuroimaging the self. For example, Johnson et al.[10] used only two conditions in their neuroimaging the self. In the experimental condition, participants responded to a variety of statements about their own abilities, traits and attitudes presented via headphones, such as "my future is bright", "I'm a good friend", and "I have a quick temper". In the control condition (used to control for auditory processing, attention, language comprehension, decision making response and retrieval), participants responded to statements about semantic knowledge, such as "Ten seconds is more than a minute" and "You need water to live". Self-referential processing minus semantic processing in fMRI data analysis showed that all 11 participants individually activated the anterior medial prefrontal cortex and posterior cingulate. Of course, the activations in the group analysis were highly consistent with the individual analysis. As this study shows, you can use sentences as materials (not just individual trait adjectives), and the sentences in the experimental condition can be different from the ones in the control condition (not the same trait adjectives required in the first type of self-referential processing). These characters are very important to extend our studying on the self with neuroimaging. Although personal trait is very important for one's self, it is not the whole of the self. If using sentences as materials and the sentences in the experimental condition can be different from the ones in the control condition, then it is possible for us to explore the self in its various aspects without limitation, not to mention using pictures as the material in neuroimaging the self. For example, you can describe a person by a series of numbers in modern society: number of licence, phone, many different passwords, bankbook, number plate, birthday, and so on. Using these numbers in sentences as stimulus you can get a numerical self in neuroimaging the self.

## 3  Self-referential processing as implement of "I think therefore I am" approach[1]

There are a number of approaches to processing information about the self. Self-referential processing is one of them and needs to be considered in the context of the multifaceted self.

According to Kihlstrom et al.[3] there are three

representations of self-knowledge: (1) the self as concept, (2) the self as image, and (3) the self as memory. Though this categorization is useful in theory, the first two types of representations do not lend themselves easily to empirical neuroimaging study. Based on the representations of self-knowledge[3] and neuroimaging studies of the self so far, we think that we can learn self-knowledge from how we get its information by neuroimaging. We propose that there are three approaches to neuroimaging the self:

(1) "I recognize therefore I am" approach, which treats the self as a percept and tests one's ability to recognize oneself from percept[25-27]. For example, people recognize him/herself from a photograph, a mirror image and a voice[11,15,28].

(2) "I remember therefore I am" approach, which treats the self as the memories of one's own life and tests the existence of one's self in the past by autobiographical memory and episodic memory[3,17,19~21,29~31].

(3) "I think therefore I am" approach, which treats the self as the metarepresentations and tests one's ability to reflect one's own present mental activities and states[21,29]. The nine studies of the self using neuroimaging summarized here employed this approach.

These three approaches cover a number of topics of the multifaceted self from percept, memory to thinking and emotion, which is consistent with the idea that self relevant information processing includes sensory input, memory and emotion[11,32].

Why do we need such categorizations in neuroimaging the self besides the representations of self-knowledge? The key point is that cognitive neuroscience mainly based on neuroimaging studies "can provide a new way of dividing cognition into topics of study. That is, ······ cognitive neuroscience may parse cognition differently than does strictly behavioral work or computational analysis"[33]. For example, we have shown above that representation of one's own personal traits and representation of one's own mental states can be grouped under the same category from neuroimaging data, or brain/mind point of view, but they are two representations from social psychological and neuropsychological point of

---

1) In one of the most famous statements in Western idealism, Descartes (1596~1650) concluded "I think therefore I am." Here I borrow his saying to express an approach in neuroimaging the self.

view[21]. Another example is that based on neuroimaging data we think that type 1 of self-referential processing (processing information about personality trait adjective with reference to self) belongs to "I think therefore I am" approach, but it was traditionally categorized as self-reference effect in memory field by cognitive psychology and social psychology[1].

Why does self-referential processing belong to "I think therefore I am" approach? In other words, why is self-referential processing an implement of the approach? We consider that the nature of self-referential processing makes it. Type 1 of self-referential processing—making judgment about one's own personal traits, is a self-reflective encoding, and encoding the meaningful attributes of a trait adjective can be seen as the deepest semantic processing according to the levels-of-processing theory[13,14]. A deep semantic processing must be an elaborate thinking processing. Type 2 of self-referential processing—making judgment about one's own present mental states, was called self-reflective thought (when subjects respond to a variety of statements requiring knowledge of and reflection on their own abilities, traits and attitudes) by Johnson et al.[10]. Similarly, we can think of an experimental condition[9]—an internally cued condition induced self-referential mental activity and emotional processing—as thinking processing. In Zysset et al.'s study[8], subjects made evaluative judgment in that the internal scale was related to the person's value system. Evaluating a person's value system can be referred to as a metacontrol process or thinking processing. In short, two types of self-referential processings are thinking processes and can be an implement of "I think therefore I am" approach.

## 4   Conclusion

We have reviewed here previous functional neuroimaging studies of self-referential processing. The evidence shows that the medial prefrontal cortex and cingulated cortex are the neural correlates of self-referential processing. We propose that self-referential processing is an implement of "I think therefore I am" approach in neuroimaging the self. Type 2 of self-referential processing is more feasible in neuroimaging the self and can provide more space to explore the multifaceted self in its various aspects. Finally, we argue that the self is a neurocognitive (mind/brain) system[34] and three approaches to neuroimaging the self fit well with its three central components: the self as percept, the self as memory and the self as thinking and emotion.

## References

1   Symons, C. S. et al. The self-reference effect in memory: A meta-analysis. Psychological Bulletin, 1997, 121: 371.

2   Zhu, Y. et al. An experimental study on the self-reference effect. Science in China (Series C), 2002, 45(2): 120.

3   Kihlstrom, J. F. et al. Self and identity as memory. In: Handbook of Self and Identity ( ed. Leary, M. R ), New York: Guilford Press, 2002, 68.

4   Kelley, W. M. et al. Finding the self? An event-related fMRI study. Journal of Cognitive Neuroscience, 2002, 14(5): 785.

5   Craik, F. I. M. et al. In search of the self: A positron emission tomography study. Psychological Science, 1999, 10(1): 26.

6   Zhang, L. et al. In search of the Chinese self: An fMRI study. 10th Annual Meeting of Cognitive Neuroscience, New York, 2003.

7   Horiuchi, T. et al. Differences between self-referent encoding and other encodings: An event related fMRI study. The 8th International Conference on Functional Mapping of Human Brain, 2002, 308.

8   Zysset, S. et al. The anterior frontomedian cortex and evaluative judgment: An fMRI study. NeuroImage, 2002, 15: 983.

9   Gusnard, D. A. et al. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. Proc. Natl. Acad. Sci. USA, 2001, 98: 4259.

10   Johnson, S. C. et al. Neural correlates of self-reflection. Brain, 2002, 125: 1808.

11   Kircher, T. J. et al. Towards a functional neuroanatomy of self processing: effects of faces and words. Cognitive Brain Research, 2000, 10: 133.

12   Kircher, T. T. J. et al. The neural correlates of intentional and incidental self processing. Neuropsychologia, 2002, 40: 683.

13   Craik, F. I. M. et al. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11: 671.

14   Craik, F. I. M. et al. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104: 268.

15   Kircher, T. T. J. et al. Recognizing one's own face. Cognition, 2001, 78: B1.

16   Sugiura, M. et al. Passive and active recognition of one's own face. Neuroimage, 2000, 11: 36.

17   Conway, M. A. et al. A positron emission tomography (PET) study of autobiographical memory retrieval. Memory, 1999, 7: 679.

18   Fink, G. R. et al. Cerebral representation of one's own past: neural networks involved in autobiographical memory. Journal of Neuroscience, 1996, 16(13): 4275.

19   Maguire, E. A. Neuroimaging studies of autobiographical event memory. Phil. Trans. R. Soc. Lond. B, 2001, 356: 1441.

20   Lepage, M. et al. Prefrontal cortex and episodic memory retrieval mode. Proc. Natl. Acad. Sci. USA, 2000, 97: 506.

21   Klein, S. B. et al. A social-cognitive neuroscience analysis of the self. Social Cognition, 2002, 20(2): 105.

22   Devinsky, O. et al. Contributions of anterior cingulated cortex to behaviour. Brain, 1995, 118(12): 279.

23  Stuss, D. et al. Adult clinical neuropsychology: lessons from stud-
ies of the frontal labes. Annual Review Psychology, 2002, 53:
401.

24  Ross, J. A. The self: From soul to brain. Journal of Consciousness
Studies, 2003, 10(2): 67.

25  Keenan, J. P. et al. Self-recognition and the right prefrontal cor-
tex. Trends in Cognitive Science, 2000, 4: 338.

26  Keenan, J. P. et al. Self-recognition and the right hemisphere.
Nature, 2001, 409: 305.

27  Turk, D. J. et al. Mike or me? Self recognition in a spilt-brain pa-
tient. Nature Neuroscience, 2002, 5(9): 841.

28  Nakamura, K. et al. Neural substrates for recognition of familiar
voices: a PET study. Neuropsychologia. 2001, 39: 1047.

29  Klein, S. B. A self to remember. In: Individual Self, Relational
Self, Collective Self (ed. Sedikides, C.). London: Psychology
Press, 2001, 25.

30  Cabeza, R. et al. Lateralization of prefrontal activity during
episodic memory retrieval: Evidence for the production-monitoring
hypothesis. Journal of Cognitive Neuroscience, 2003, 15 ( 2 ):
249.

31  Tulving, E. Episodic memory: From mind to brain. Annu. Rev.
Psychol., 2002, 53: 1.

32  Dolan, R. Feeling the neurobiological self. Nature, 1999, 401:
847.

33  Smith, E. E. Infusing cognitive neuroscience into cognitive psy-
chology. In: Mind and Brain Science in the 21st Century (ed. Sol-
so, R. L. ), Massachusetts: The MIT Press, 1999, 71.

34  Churchland, P. Self-representation in nervous systems. Science,
2002, 296: 308.