

# Specifying the self for cognitive neuroscience

Kalina Christoff<sup>1</sup>, Diego Cosmelli<sup>2</sup>, Dorothee Legrand<sup>3</sup> and Evan Thompson<sup>4</sup>

<sup>1</sup> Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, V6T 1Z4 Canada

<sup>2</sup> Escuela de Psicología, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago, Chile

<sup>3</sup> Centre de Recherche en Epistémologie Appliqué (CREA), ENSTA-32, boulevard Victor, 75015 Paris, cedex 15, France

<sup>4</sup> Department of Philosophy, University of Toronto, 170 St George Street, Toronto, ON, M5R 2M8 Canada

**Cognitive neuroscience investigations of self-experience have mainly focused on the mental attribution of features to the self (self-related processing). In this paper, we highlight another fundamental, yet neglected, aspect of self-experience, that of being an agent. We propose that this aspect of self-experience depends on self-specifying processes, ones that implicitly specify the self by implementing a functional self/non-self distinction in perception, action, cognition and emotion. We describe two paradigmatic cases – sensorimotor integration and homeostatic regulation – and use the principles from these cases to show how cognitive control, including emotion regulation, is also self-specifying. We argue that externally directed, attention-demanding tasks, rather than suppressing self-experience, give rise to the self-experience of being a cognitive-affective agent. We conclude with directions for experimental work based on our framework.**

## Investigating self-experience in cognitive neuroscience

How does the embodied brain give rise to self-experience? This question, long addressed by neurology [1] and neurophysiology [2], now attracts strong interest from cognitive neuroscience and the neuroimaging community [3–6].

Recent neuroimaging studies have investigated self-experience mainly by employing paradigms that contrast self-related with non-self-related stimuli and tasks. Such paradigms aim to reveal the cerebral correlates of ‘self-related processing’ (see Glossary). Recent reviews identify several brain regions that appear most consistently activated in self-related paradigms such as assessing one’s personality, physical appearance or feelings; recognizing one’s face; or detecting one’s first name (see [4,6] for extensive reviews). The medial prefrontal cortex (mPFC) and the precuneus/posterior cingulate cortex (Precuneus/PCC) are the most frequently discussed [4–10], but two additional regions, the temporoparietal junction (TPJ) and temporal pole, are also consistently activated [6].

Although these studies have contributed valuable information about the neural correlates of self-related processing, two issues have recently arisen [3,6]. First, the identified regions, especially the midline regions (mPFC, Precuneus/PCC) often associated with self-related

processing [4,7–10], might not be self-specific, because they are also recruited for a wide range of other cognitive processes – recall of information from memory, inferential reasoning, and representing others’ mental states [3,5,6]. In addition, the PCC appears to be engaged in attentional processes and might be a hub for attention and motivation [11,12], whereas the TPJ is important for attentional reorienting [13]. Hence, describing these regions (singly or collectively) as self-specific could be unwarranted [3,5,6]. Second, studies employing self-related processing approach self-experience through the self-attribution of mental and physical features, and thereby focus on the self as an *object of attribution* and not the self as the *knowing subject* and *agent*. To invoke James’ [14] classic distinction, this paradigm targets the ‘Me’ – the self as known through its physical and mental attributes – and not the ‘I’ – the self as subjective knower and agent. Thus, relying exclusively on this paradigm would limit the cognitive neuroscience of self-experience to self-related processing (the ‘Me’), to the neglect of the self-experience of being a knower and agent (the ‘I’) [6,15].

In this paper, we focus on the ‘I’ – experiencing oneself as the agent of perception, action, cognition and emotion – and

## Glossary

**Cognitive control:** the process by which one focuses and sustains attention on task-relevant information and selects task-relevant behavior.

**Emotion regulation:** the process by which one influences one’s experience and expression of emotion.

**Homeostatic regulation:** the process of keeping vital organismic parameters within a given dynamical range despite external or internal perturbations.

**‘I’ versus ‘Me’:** experiencing oneself as subjective knower and agent versus experiencing oneself as an object of perception or self-attribution.

**Self-related processing:** processing requiring one to evaluate or judge some feature in relation to one’s perceptual image or mental concept of oneself.

**Self-specific:** a component or feature that is exclusive (characterizes oneself and no one else) and noncontingent (changing or losing it entails changing or losing the distinction between self and non-self).

**Self-specifying:** any process that specifies the self as subject and agent by implementing a functional self/non-self distinction.

**Sensorimotor integration:** the mechanisms by which sensory information is processed to guide motor acts, and by which motor acts are guided to facilitate sensory processing.

**Task-negative/default-network brain regions:** regions exhibiting sustained functional activity during rest but showing consistent deactivations during externally directed, attention-demanding tasks. Such regions include the precuneus/posterior cingulate cortex, medial prefrontal cortex and bilateral temporoparietal junction.

**Task-positive brain regions:** regions consistently activated during externally directed, attention-demanding tasks. Such regions include the intraparietal sulcus, frontal eye field, middle temporal area, lateral prefrontal cortex and dorsal anterior cingulate.

we propose a theoretical framework that links this type of self-experience to a wide range of neuroscientific findings at different levels of neural functioning.

According to our proposal, experiencing oneself as an agent depends on the existence of specific types of dynamic interactive processes between the organism and its environment. We call these processes ‘self-specifying’ because they implement a functional self/non-self distinction that implicitly specifies the self as subject and agent [6,16]. To illustrate the basic principles of self-specifying processes, we describe two paradigmatic examples – sensorimotor integration and homeostatic regulation – that underlie the self-experience of being a bodily agent. We then argue that although externally directed attention-demanding tasks can compromise self-related processing [7–10,17–19], such tasks can be expected to enhance another fundamental type of self-experience, namely that of being a cognitive–affective agent [6,15,16]. In support of this point, and to show how cognitive neuroscience can begin to model this type of self-experience, we apply the concept of self-specifying processes to cognitive control, including emotion regulation. We conclude with suggestions for future experimental work based on our framework.

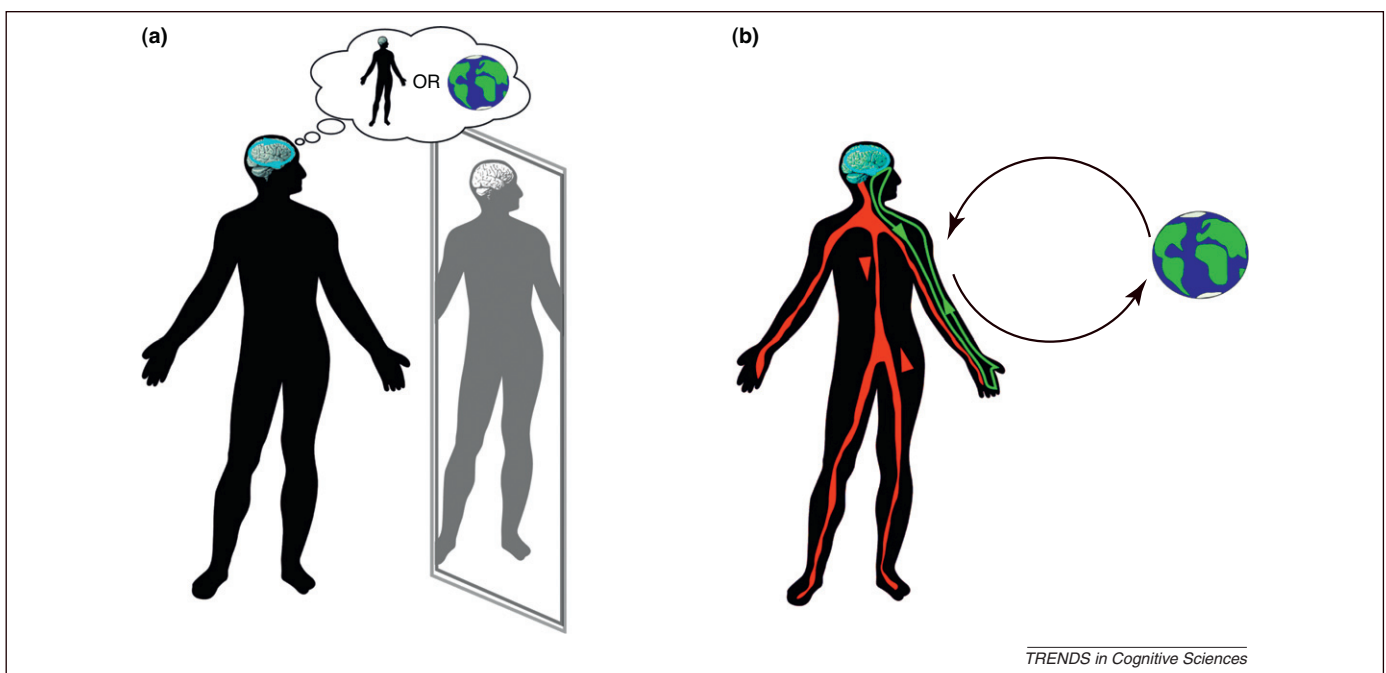
### Self-experience as arising from self-specifying processes

Many neuroimaging studies have focused on the type of self-experience that occurs when a person directs his or her attention away from the external world (e.g. when task demands are low, when performing a self-reflective task or during rest) [7–10,17] (Figure 1a). At the same time, other lines of investigation concerned with embodied experience have examined self-experience during world-directed per-

ception and action [1,20,21] (Figure 1b). These investigations have focused on bodily awareness in sensorimotor integration [20,21] and homeostatic regulation [1,22,23]. Central to this approach is the notion that the organism constantly integrates efferent and afferent signals in a way that distinguishes fundamentally between *reafference* – afferent signals arising as a result of the organism’s own efferent processes (self) – and *exafference* – afferent signals arising as a result of environmental events (non-self). By implementing this functional self/non-self distinction, efferent–afferent integration implicitly specifies the self as a bodily agent [6,16,21].

### Sensorimotor integration

The notion of self-specifying processes is easiest to illustrate through the systematic linkage of sensory and motor processes in the perception–action cycle (Box 1). An organism needs to be able to distinguish between sensory changes arising from its own motor actions (self) and sensory changes arising from the environment (non-self). The central nervous system (CNS) distinguishes the two by systematically relating the efferent signals (motor commands) for the production of an action (e.g. eye, head or hand movements) to the afferent (sensory) signals arising from the execution of that action (e.g. the flow of visual or haptic sensory feedback). According to various models going back to Von Holst [24], the basic mechanism of this integration is a comparator that compares a copy of the motor command (information about the action executed) with the sensory refference (information about the sensory modifications owing to the action) [25]. Through such a mechanism, the organism can register that it has executed a given movement, and it can use this information to



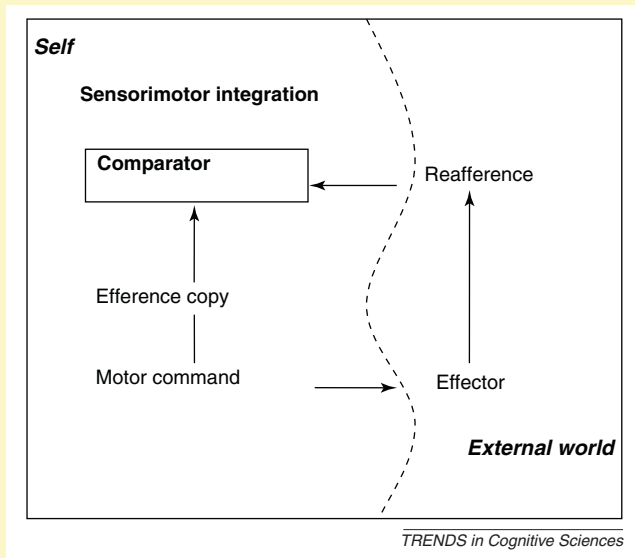
**Figure 1. Two types of self-experience.** (a) The ‘Me’ or self-related processing (here depicted as self-recognition and reflective thinking about oneself). Its neural substrates are thought to be restricted to a subset of midline cortical regions (mPFC and Precuneus/PCC). It is also thought to compete for cognitive resources when some aspect of the world demands attention. (b) The ‘I’ as embodied agent. This type of self-experience arises from the integration of efferent and refferent processes, notably sensorimotor integration (green loop) and homeostatic regulation (red loop), as well as possible higher level efferent–refferent regulatory loops such as the one instantiated by cognitive control processes (blue loop). Such regulatory loops implement a functional self/non-self distinction that implicitly specifies the self as agent. This type of self-experience implicitly occurs during attention-demanding interactions with the environment (black arrows).

### Box 1. Self-experience and sensorimotor integration

The self-experience of being an embodied agent depends on the sensorimotor mechanisms that integrate efference with refference (Figure I). A basic level mechanism allows efferences to be systematically related to their refferent consequences. This anchoring of efference to refference implements a functional self/non-self distinction that implicitly specifies the self as a bodily agent [6,21].

For example, consider the motor act of biting a lemon and the resulting taste. This experience is characterized by (i) a specific content (lemon, not chocolate); (ii) a specific mode of presentation (tasting, not seeing); and (iii) a specific perspective (my experience of tasting). The process of relating an efference (the biting) to a refference (the resulting taste of acidity) is what allows the perception to be characterized not only by a given content (the acidity) but also by a *self-specific perspective* (I am the one experiencing the acidity of the lemon juice) [6,21].

The agent's *perspective* is thus a central concept within this framework. Although the basic sensorimotor integration processes do not involve any representation of the self *per se*, they are nonetheless *self-specifying* [6] because they implement a unique egocentric perspective in perception and action, and thus implicitly specify the self as subject and agent of that perspective. According to this view, self-experience is present whenever a self-specific perspec-



**Figure I.** Sensorimotor integration  
Comparator mechanism for relating efferent signals to refferent sensory feedback.

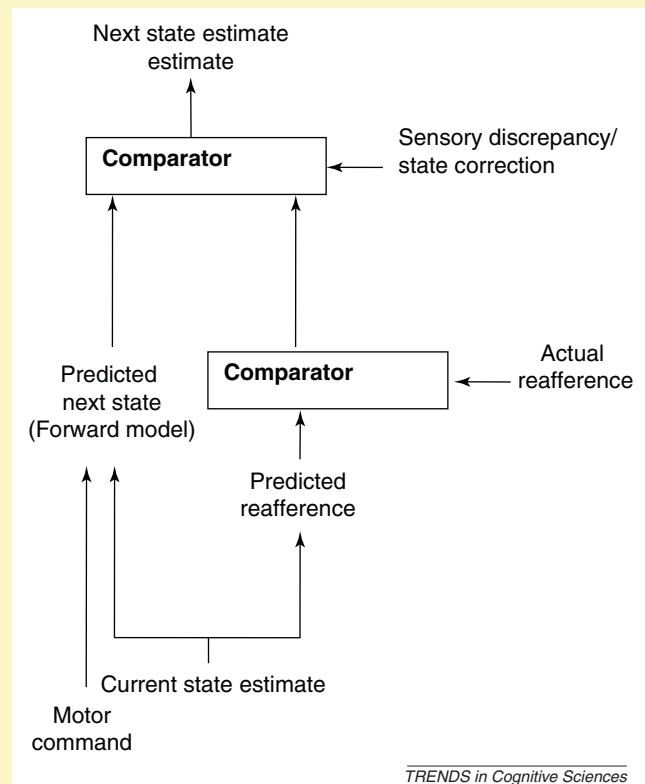
process the resulting sensory refference. The crucial point for our purposes is that refference is *self-specific*, because it is intrinsically related to the agent's own action (there is no such thing as a non-self-specific refference). Thus, by relating efferent signals to their afferent consequences, the CNS marks the difference between self-specific (refferent) and non-self-specific (exafferent) information in the perception-action cycle. In this way, the CNS implements a functional self/non-self distinction that implicitly specifies the self as the perceiving subject and agent.

#### Homeostatic regulation

Self-specifying refferent-efferent processes are key components of homeostatic regulation, which implements the self/non-self distinction at the basic level of life preservation [1,16,22,23]. To ensure the organism's survival through changing internal and external conditions, afferent signals conveying information about the organism's

tive exists, regardless of the properties of the represented content [6,15,16,21].

The original mechanism of sensorimotor integration (Figure I) can be elaborated to include higher level comparators between intended, predicted and actual refference (Figure II). For example, Wolpert and colleagues [25] described a two-process model of action monitoring. The first process (Figure II, left) uses the motor command and the current state estimate to achieve a next state estimate using the forward model (or a prediction) to simulate the arm's dynamics. The second process (Figure II, right) uses the difference between expected and actual sensory feedback to correct the forward model's next state estimate. Through such sophisticated comparators, the model can handle higher level phenomena, such as intentions, predictions, mental simulation and goals [20].



**Figure II.** Two-process model of action monitoring (Ref. [25]).

internal state are continually coupled with corresponding efferent regulatory processes that keep afferent parameters within a tight domain of possible values [1,22,23]. Refferent-efferent loops from spinal nuclei to brainstem nuclei and midbrain structures are involved in somato-autonomic adjustments; these loops are modulated by the hypothalamus as well as mid/posterior insula (sensory) and anterior cingulate (motor) cortices [23]. This vertically integrated, interoceptive homeostatic system specifies the self as a bodily agent by maintaining the body's integrity (self) in relation to the environment (non-self) [22], and by supporting the implicit feeling of the body's internal condition in perception and action [23].

#### Specifying the self as knowing subject and agent

The refferent-efferent processes just described specify the self not as an *object* of perception or attribution (the 'Me') but as the experiential *subject* and *agent* of perception,

action and feeling (the ‘I’). Sensorimotor integration specifies a unique perceptual perspective on the world, whereas homeostatic regulation specifies a unique affective perspective based on the inner feeling of one’s body. The resulting perspective is *self-specific* in the strict sense of being both *exclusive* (it characterizes oneself and no one else) and *noncontingent* (changing or losing it entails changing or losing the distinction between self and non-self) [6]. In the general case, ‘I’ perceive and act from my self-specific perspective while implicitly experiencing myself as perceiver and agent. In some particular cases, what ‘I’ perceive is ‘Me’, such as when I visually recognize myself. Although many non-human animals can implicitly experience themselves as embodied agents through the types of self-specifying sensorimotor and homeostatic processes described above [26], only humans and a few other species seem capable of self-recognition [27], and thus of experientially relating the ‘I’ and the ‘Me’. What we emphasize here is that whereas the ‘Me’ consists in the features one perceives as belonging to oneself, the ‘I’ consists in the self-specific, agentive perspective from which such perceptions occur; hence, to explain the ‘I’ we need to explain how such a perspective is implemented. Our proposal is that the reafferent–efferent processes of sensorimotor integration and homeostatic regulation implement a self-specific, agentive perspective at the bodily level of perception and feeling.

This model predicts that if a brain process involves only afference without a matching efference/reafference, it will not specify the organism as subject or agent, and thus will not constitute a self-specifying process. For example, the ‘feedforward sweep’ in visual processing from early visual areas to extrastriate areas, which Lamme [28] argues is not accompanied by conscious awareness, would not qualify as self-specifying, whereas ‘recurrent processing’ in multiple visual areas, which Lamme argues is associated with ‘phenomenal awareness’ (short-lived awareness that is not necessarily reportable), would qualify as self-specifying only if linked to matching efference/reafference. Our model thus allows that non-self-specifying processes occur in parallel with self-specifying ones, and it leaves open the question whether there exist conscious processes that do not include even minimal self-specification (as Lamme’s proposal suggests) or whether every conscious process is also minimally self-specifying (as others have argued [15]).

Given this model, we next consider the view, prevalent in the recent neuroimaging literature [7–10,17–19], that self-experience is suppressed during externally directed, attention-demanding tasks. We argue that this view needs qualification to take into account the self-experience of being a cognitive–affective agent.

### Is self-experience suppressed during world-directed attention?

One outcome of functional magnetic resonance imaging (fMRI) studies using self-related processing as the main paradigm for understanding self-experience is the view that self-experience occurs mostly when individuals are not preoccupied with externally oriented tasks and that it is suppressed when such tasks do occur [7–10]. This view is

based partly on findings from a growing number of studies examining spontaneous fluctuations in the fMRI signal during task-free, resting-state conditions [29]. These findings have distinguished between (i) task-positive regions (e.g. dorsolateral PFC, inferior parietal cortex and supplementary motor area), whose activity increases during externally oriented attention and (ii) task-negative/default-network regions (e.g. mPFC, Precuneus/PCC and TPJ), whose activity decreases across a wide variety of tasks. These task-positive and task-negative networks also appear to be anticorrelated in their spontaneous activity during the resting state [30], so that increased activity in one network has been noted to correlate with decreased activity in the other [17–19].

A prominent interpretation of these findings is that the brain alternates dynamically between a task-oriented, externally directed state and a task-independent, self-directed state, with self-experience in the form of self-related processing mainly occurring during the task-independent, self-directed state [8–10,18,19]. A wide variety of studies have been taken to support this interpretation; these studies indicate that externally oriented, attention-demanding tasks, which are considered to suppress introspective thoughts, tend to suspend default-network activity, whereas resting conditions, as well as practiced tasks that do not suppress introspective thoughts, correlate with an active default network (see [31] for a comprehensive review). Additional support is thought to come from the finding that tasks requiring individuals to make explicit reference to some aspect of themselves implicate medial prefrontal regions also active as part of the default network [4,5,26,31]. Hence, it has been proposed, on the one hand, that self-experience is largely absent during world-directed attention (because self-related processing is strongly suppressed) [17], and, on the other hand, that during rest conditions, subjects mainly engage in self-referential processing [7–10].

This conclusion, however, rests on the following assumptions: (i) the main way to experience the self is as an *object* of one’s attention (i.e. through self-related processing); (ii) self-reflective, introspective processes are linked to task-negative/default-network regions; and (iii) the brain is organized into a dynamic system of task-positive regions subserving world-directed attention and task-negative/default regions subserving self-directed attention, with these two networks acting in opposition so that recruitment of one suppresses the other.

Each of these assumptions, however, needs qualification in light of the recent theoretical literature and empirical findings.

First, treating self-related processing as the main form of self-experience limits self-experience to the ‘Me’ (self as object of one’s attention) while neglecting the ‘I’ (self as knowing subject and agent). For example, if the agentive ‘I’ is considered at the bodily level of sensorimotor integration, then task-positive regions such as the supplementary motor cortex and inferior parietal cortex could be viewed as crucial to self-experience, for these regions serve to implement sensorimotor integration tasks [25,32,33]. More generally, although world-directed attention can suppress self-related processing, one cannot conclude that it suppresses



every form of self-experience, especially the self-experience of being a cognitive agent (which it can instead enhance).

Second, self-referential and introspective processes have also been linked to recruitment of regions outside the default network. For example, self-related processing activates the temporopolar cortex as consistently as the three main default network regions (mPFC, Precuneus/PCC and TPJ) [34], and is also frequently associated with activations in the insula and lateral PFC [6]. Furthermore, introspective mental processes have been linked to a recruitment of the anterior portion of the lateral PFC, namely the rostrolateral PFC [35–37], which is considered to be part of a cognitive control network separable from the default network [38]. These findings indicate that self-referential processing is not uniquely associated with task-negative/default-network regions. Therefore, reduced or inhibited activity in default network regions does not necessarily indicate that self-directed introspective processes are suppressed, because they can be implemented through regions outside the default network.

Finally, recent studies have begun to qualify the picture of task-positive and task-negative/default networks as invariably acting in opposition to each other. A parallel recruitment of task-positive and task-negative/default-network regions has been observed during several tasks, such as passive sensory stimulation [39], continuous movie viewing [40], narrative speech comprehension [41], autobiographical planning [42] and mind wandering during a sustained attention task [36]. These diverse findings suggest that characterizing brain activity as either task-positive/world-directed or task-negative/self-directed is incomplete. Rather, such neural recruitments and cognitive processes can occur in parallel.

In contrast to the view that attention-demanding tasks suppress self-experience, we propose that such tasks can be expected to enhance the self-experience of being a cognitive-affective agent. An outstanding task for cognitive neuroscience is to integrate this type of self-experience and self-related processing into an overarching explanatory framework that can guide empirical research. In the next section, we propose what we believe is a crucial element of such a framework. By describing how the concept of self-specifying processes can be applied to cognitive control, including emotion regulation, we argue that cognitive-affective processes instantiate the self-experience of being a cognitive-affective agent. In this way, we show how cognitive neuroscience can investigate this type of self-experience by including paradigms involving attention to the external world.

### Self-specifying processes during attention-demanding tasks

Can cognitive control processes in affectively neutral contexts and affectively arousing contexts implicitly specify the self as a cognitive-affective agent?

#### *Cognitive control processes in affectively neutral contexts*

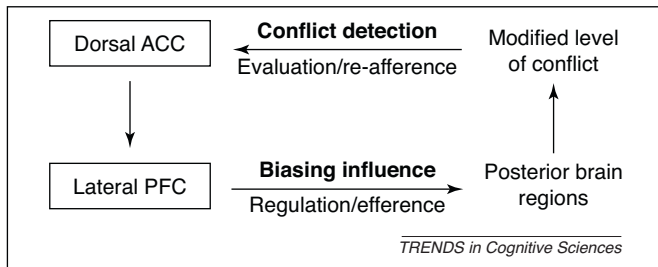
Cognitive control processes serve both to focus attention on task-relevant information versus other competing sources of information and to select task-relevant behavior over

habitual or otherwise prepotent responses. For example, in a Stroop task, the goal is to name the ink color of a printed color name while ignoring the word's meaning. Individuals are slower to respond when the information is incongruent (e.g. the word RED is printed in blue ink) than when it is congruent (e.g. the word RED is printed in red ink), and the slower response time is taken to reflect the need for higher attentional control when a conflict in perceptual information is present.

According to the influential 'conflict-monitoring model' [43], cognitive control is implemented through a regulatory conflict-control loop consisting of two components. An *evaluative* or conflict-monitoring component detects conflicts in the information available for task performance, whereas a *regulative* component exerts a top-down biasing influence on the cognitive and motor processes required for task performance. At the neural level, the dorsal anterior cingulate cortex (dACC) has been proposed to support the evaluative process of conflict monitoring [43,44], whereas lateral PFC regions have been proposed to underlie the regulative process of cognitive control [43,45]. This model predicts that strong ACC activity should be followed by behavior reflecting relatively focused attention, and weak ACC activity by behavior reflecting less focused attention. In keeping with this prediction, Kerns and colleagues [46] found that high dACC activation for incongruent trials in the Stroop task was followed by low interference on the subsequent trial, as well as by strong activation in dorso-lateral PFC. These findings suggest that the dACC could signal the need for control adjustments to lateral PFC and thereby strengthen cognitive control [45].

Our aim in describing the conflict-monitoring model is not to endorse it against other important models of cognitive control [47–49] or ACC functioning [50,51]. In particular, we do not suppose that dACC is involved in cognitive but not emotional functions, whereas ventral ACC does the reverse [52], because recent experimental findings and theoretical considerations argue against both this particular cognitive-affective division [53] as well as emotion-cognition separations more generally in the brain and behavior [53,54]. Instead, we use the model to illustrate how cognitive-control processes can be self-specifying.

For the purposes of the present argument, the key feature of the conflict-monitoring model is the functional distinction between a regulatory function and an evaluative function. The control loop comprising these two functions (Figure 2) strongly resembles the integration of efferent and reafferent information during sensorimotor processing, with the regulative component corresponding to efferent influence and the evaluative component corresponding to a reafferent process. We propose that such a regulative-evaluative loop can implement a functional self/non-self distinction between, on the one hand, reafferent signals about modifications in level of conflict resulting from one's own cognitive-control efforts (self), and, on the other hand, exafferent signals about the level of conflict resulting from environmental sources such as stimulus properties (non-self). By implementing this self-specific, agentive perspective in cognitive control, the regulatory conflict-control loop would implicitly specify the self as a



**Figure 2.** Cognitive control as a self-specifying process. The conflict-monitoring model of cognitive control [43] depicted as implementing a possible efferent/reafferent regulatory loop. This loop can define the functional self/non-self distinction between reafferent signals resulting from one's own cognitive control efforts (self) and exafferent signals about the level of conflict resulting from environmental sources such as stimulus properties (non-self).

cognitive agent. Note that this cognitive form of self-experience would subsume the self-experience of being an embodied agent resulting from sensorimotor integration, because cognitive control operates on sensorimotor processes themselves, and thus occurs at higher levels of integration in the perception–action cycle [55].

As originally conceived, the cognitive control of attention was closely linked to self-regulation [56,57], including the self-experience of being a cognitive agent [57]. Concern with this link, however, seems to have largely disappeared from the recent cognitive neuroscience literature, possibly because of the assumption that self-experience is suppressed during attention-demanding tasks [7–10,17–19], as well as the observation that brain regions associated with cognitive control, such as the lateral PFC and dACC, largely overlap with the task-positive regions outlined earlier. Indeed, meta-analyses show that the lateral PFC and dACC are among the most consistently recruited brain regions across a broad range of attention-demanding tasks, including perception, response selection, executive control, working memory, episodic memory and problem solving [58,59]. Nevertheless, as discussed above, recruitment of these task-positive regions is not mutually exclusive with recruitment of the task-negative/default-network regions. Although intense engagement in sensorimotor tasks can suppress the task-negative/default-network regions that also subserve self-related processing [17–19], one can envision situations (e.g. introspection, envisioning the perspective of others, mind wandering) in which the required mental processes call upon resources from both sets of regions and hence lead to more balanced activations between them, as indicated by recent results [36,39–42]. Furthermore, even in situations where the dACC and lateral PFC are recruited in opposition to task-negative/default-network regions (i.e. with a concomitant deactivation of these regions), self-experience might still be crucially present in the form of the ‘I’ or self-as-cognitive-agent, as a result of cognitive control processes being self-specifying in the way just outlined above.

### Emotion regulation

The cognitive and behavioral control of emotion in affectively arousing or challenging situations [60,61] provides another case where we can expect to find the self-experience of being a cognitive–affective agent. Although emotion

regulation and self-related processing have often been linked by pointing to their common reliance on midline cortical structures [61,62], we propose that another fundamental but less explored link between self-experience and emotion regulation can be found in how emotion regulation processes are also self-specifying.

Recent discussions have proposed a distinction between two main forms of emotion regulation – a deliberate or voluntary form, and an implicit or incidental form [60,61,63–65]. Deliberate emotion regulation relies on the same cognitive control mechanisms required for attention-demanding tasks [61]. Thus, tasks requiring reappraisal – reinterpreting the meaning of a stimulus to change one's emotional response to it [60,61] – recruit dACC and lateral PFC regions [61]. Here these regions are thought to subserve explicit reasoning about how the association between a situation and one's emotional response to it can be changed. For example, if one is viewing a picture of a burn victim in a hospital bed, it might be possible to modify the original emotional response of distress or sadness by focusing on possible positive aspects, such as the victim's successful progress toward a healthier state or that the victim survived. Maintaining such descriptions is thought to bias perceptual and associative-memory systems; these systems in turn send signals to subcortical appraisal systems, such as the amygdala and ventral striatum [61], and thus indirectly modify the original emotional response.

We propose that such a regulatory–evaluative loop can implement a functional self/non-self distinction between the effortful reappraisal process (self) and the target of that process, namely the emotional scene (non-self). In this way, emotion regulation can implicitly specify the self as the cognitive–affective agent engaged in trying to reinterpret and thereby control an emotional response.

Deliberate forms of emotion regulation are associated not only with dACC and lateral PFC – regions crucially involved in cognitive control – but also with recruitment of dorsomedial PFC (dmPFC) [61,64,65], a brain region considered to support reflective awareness of one's feelings, and thus to enable higher level, metarepresentations of one's own experience [63]. By allowing the maintenance of such emotion-specific metarepresentations, and through its dense interconnections with the ventromedial PFC (vmPFC) [66], the dmPFC can exert a biasing influence on emotion processes during deliberate attempts at emotion regulation. Thus, by both influencing and re-representing the emotion processes in more ventral systems, the dmPFC and its interconnected ventral structures can form another regulatory–evaluative loop that implicitly specifies the self as cognitive–affective agent in effortful emotion regulation.

In contrast to deliberate emotion regulation, implicit or incidental emotion regulation has been linked to medial regions such as the rostral ACC (rACC), subgenual ACC and vmPFC [61]. For example, the rACC is associated with regulation of attention to emotional (but not non-emotional) distracters during an emotional version of the Stroop task [67,68]. During this task, subjects are not instructed to regulate their emotions, thus the recruitment of the rACC and its accompanying regulation of emotional attention can

be considered incidental to the main task [65]. Activation in rACC appears to be accompanied by a simultaneous and correlated reduction of amygdala activity; this relation suggests that resolving emotional conflict depends on a rACC–amygdala regulatory loop [67] that also appears to use the general cognitive monitoring mechanism of the dACC to detect the presence of conflict [68]. Thus, a self-specifying evaluative–regulatory loop can be formed between rACC and dACC, analogous to that between lateral PFC and dACC, but dedicated to the resolution of emotional conflict through an rACC biasing influence on amygdala activity.

Furthermore, regions playing a role in deliberate emotion regulation, such as the dACC and dmPFC [63,64], and possibly the right ventrolateral PFC [65], also appear to participate in implicit emotion regulation. For example, the dACC and dmPFC have autonomic regulatory functions mediated by direct neural connections with subcortical visceromotor centers such as the lateral hypothalamus [66]. In addition, neuroimaging studies noting an inverse correlation between medial PFC activity and heart rate variability suggest that medial PFC activity can have a tonic inhibitory effect mediated through the vagus nerve [63]. Based on these findings, researchers have described an evaluative–regulatory feedback mechanism, including an equilibration process between bottom-up and top-down interactions, through which the body state is altered as arousal processes become modulated and differentiated [63]. This mechanism provides another candidate for a self-specifying process at implicit levels of emotion regulation.

Given that these candidate self-specifying processes belong to implicit emotion regulation, the functional self/non-self distinction they implement would be closely related to the one established through homeostatic regulation between the feeling body and the environment. Indeed, implicit emotion regulation processes overlap conceptually and neurally with the higher levels of the homeostatic regulation system described earlier [1,22,23,26]. Thus, the self-experience of being an emotional agent that these processes elicit would occur at the level of affect and action tendencies [26], whereas this bodily level would be subsumed by the self-experience of being a cognitive–affective agent in deliberate emotion regulation, analogous to the way the self-experience of being a cognitive agent also subsumes the self-experience of being an embodied agent in attention-demanding cognitive tasks.

### Concluding remarks and future directions

Using the concept of self-specifying processes, we have outlined a model of how cognitive control processes, including emotion regulation, implicitly specify the self as a cognitive–affective agent. Our model suggests several questions for future investigations (Box 2). We highlight two issues here.

One issue concerns the types of neural mechanisms that integrate the efferent–reafferent and regulatory–evaluative signals in self-specifying processes. On the one hand, the comparison between efferent and reafferent signals can be remapped at higher levels by specific neural structures. For example, the anterior insula can serve to remap the second-order comparison between efferent and reafferent

### Box 2. Questions for further research

- Is the 'I' or self as-subject all or nothing, or graded?
- When multiple self-specifying processes are activated, does a stronger sense of 'I' occur?
- Can self-specifying processes be altered through attentional and emotion regulation training?
- Do self-specifying processes require higher level remapping of efferent–reafferent integration, or can such integration occur through dynamical mechanisms such as phase synchronization?
- Can self-specifying processes be identified in neuroimaging data through functional connectivity measures, and can statistical measures such as Granger causality be used to identify directional influences in such processes?
- Can self-specifying processes be identified as part of the brain's intrinsic functional architecture through intrinsic connectivity measures in resting state neuroimaging data?
- Can transcranial magnetic stimulation interfere selectively in self-specifying loops and thereby alter cognitive–affective self-experience?
- Are self-specifying processes altered in psychiatric disorders, such as schizophrenia or anorexia nervosa, which involve altered self-experience and self-other evaluation?

signals in more posteriorly located motor and sensory regions during homeostatic regulation [23]. Similarly, during cognitive control, anteriorly located lateral PFC regions, such as the rostrolateral PFC, can remap the second-order comparison between the regulative and evaluative outcomes of processes supported by the more posteriorly located dorsolateral PFC and dACC [35]. Such hierarchically organized systems can be present at multiple neural levels and in multiple functional domains. On the other hand, another type of mechanism not requiring explicit remapping by dedicated neural structures, but relying instead on dynamical coupling across multiple areas [69] (e.g. through phase synchronization of neuronal signals [70]), could be responsible for signal integration. Such dynamical mechanisms can also be implemented at multiple neural levels and in various functional domains [69,70]. Whether self-specifying processes depend on either or both of these mechanisms is an important issue for future research.

A second issue concerns the subjective nature of self-experience. Although objective measures from experimentally controlled tasks and uncontrolled rest conditions are certainly useful, we believe a richer understanding of self-experience requires the incorporation of subjective measures such as self-reports into neuroimaging protocols [36,71]. Certain questions seem tractable only with such an approach. For example, is self-experience all-or-nothing or graded in character? When multiple self-specifying processes are activated at various levels of neural functioning, does a stronger sense of self occur than when only a few are recruited? Can mental training of attention and emotion regulation [72,73] alter self-experience and its neural substrates?

As argued here, how cognitive neuroscience specifies the self profoundly shapes our view of self-experience and its neural substrates. By broadening our investigations to include the self-experience of being a cognitive agent, we can deepen our understanding of how the brain and body work together to create our sense of self.



## Acknowledgments

For helpful comments we thank Norm Farb, Alisa Mandrigin, Luiz Pessoa, Rebecca Todd and four anonymous reviewers. K.C. was supported by grants from the Canadian Institutes of Health Research (CIHR MOP 81188), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Michael Smith Foundation for Health Research (MSFHR); D.C. by Fondo Nacional de Desarrollo Científico y Tecnológico Grant 1090612; and E.T. by the Social Sciences and Humanities Research Council of Canada.

## References

- Damasio, A.R. (1999) *The Feeling of What Happens*, Harcourt
- Llinas, R. (2001) *The I of the Vortex*, MIT Press
- Gillihan, S. and Farah, M. (2005) Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychol. Bull.* 131, 76–97
- Northoff, G. et al. (2006) Self-referential processing in our brain – a meta-analysis of imaging studies on the self. *Neuroimage* 31, 440–457
- Uddin, L.Q. et al. (2007) The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends Cogn. Sci.* 11, 153–157
- Legrand, D. and Ruby, P. (2009) What is self-specific? A theoretical investigation and critical review of neuroimaging results. *Psychol. Rev.* 116, 252–282
- Gusnard, D.A. et al. (2001) Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4259–4264
- Gusnard, D.A. (2005) Being a self: considerations from functional imaging. *Conscious. Cogn.* 14, 679–697
- Wicker, B. et al. (2003) A relation between rest and self in the brain? *Brain Res. Rev.* 43, 224–230
- Schneider, F. et al. (2008) The resting brain and our self: self-relatedness modulates resting state neural activity in cortical midline structures. *Neuroscience* 157, 120–131
- Mohanty, A. et al. (2008) The spatial attention network interacts with limbic and monoaminergic systems to modulate motivation-induced attention shifts. *Cereb. Cortex* 18, 2604–2613
- Engelmann, J.B. et al. (2009) Combined effects of attention and motivation on visual task performance: transient and sustained motivational effects. *Front. Hum. Neurosci.* 3, 1–17
- Corbetta, M. et al. (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297
- James, W. (1890/1981) *The Principles of Psychology*, Harvard University Press
- Legrand, D. (2007) Pre-reflective self-as-subject from experiential and empirical perspectives. *Conscious. Cogn.* 16, 583–599
- Thompson, E. (2007) *Mind in Life*, Harvard University Press
- Goldberg, I.I. et al. (2006) When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 50, 329–339
- Fransson, P. (2005) Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis. *Hum. Brain Mapp.* 26, 15–29
- Fransson, P. (2006) How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations. *Neuropsychologia* 44, 2836–2845
- Blakemore, S.-J. and Frith, C. (2003) Self-awareness and action. *Curr. Opin. Neurobiol.* 13, 219–224
- Legrand, D. (2006) The bodily self: the sensori-motor roots of pre-reflective self-consciousness. *Phenom. Cogn. Sci.* 5, 89–118
- Parvizi, J. and Damasio, A.R. (2001) Consciousness and the brainstem. *Cognition* 79, 135–160
- Craig, A.D. (2009) How do you feel – now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70
- Von Holst, E. (1954) Relations between the central nervous system and the peripheral organs. *Br. J. Anim. Behav.* 2, 89–94
- Wolpert, D.M. et al. (1995) An internal model for sensorimotor integration. *Science* 269, 1880–1882
- Northoff, G. and Panksepp, J. (2008) The transpecies concept of self and the subcortical-cortical midline system. *Trends Cogn. Sci.* 12, 259–264
- de Waal, F.B.M. (2008) The thief in the mirror. *PLoS Biol.* 6, e201
- Lamme, V.A.F. (2003) Why visual awareness and attention are different. *Trends Cogn. Sci.* 7, 12–18
- Fox, M.D. and Raichle, M.E. (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711
- Fox, M.D. et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678
- Buckner, R.L. et al. (2008) The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38
- Andersen, R.A. and Buneo, C.A. (2003) Sensorimotor integration in posterior parietal cortex. *Adv. Neurol.* 93, 159–177
- Haggard, P. and Whitford, B. (2004) Supplementary motor area provides an efferent signal for sensory suppression. *Cogn. Brain Res.* 19, 52–58
- Christoff, K. et al. (2004) Neural basis of spontaneous thought processes. *Cortex* 40, 623–630
- Christoff, K. and Gabrieli, J.D.E. (2000) The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology* 28, 168–186
- Christoff, K. et al. (2009) Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8719–8724
- McCaig, R.G. et al. (2010) Improved modulation of rostralateral prefrontal cortex using real-time fMRI and meta-cognitive awareness. *Neuroimage* [Epub ahead of print].
- Vincent, J.L. et al. (2008) Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J. Neurophysiol.* 100, 3328–3342
- Greicius, M.D. and Menon, V. (2004) Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation. *J. Cogn. Neurosci.* 16, 1484–1492
- Golland, Y. et al. (2007) Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cereb. Cortex* 17, 766–777
- Wilson, S.M. et al. (2008) Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb. Cortex* 18, 230–242
- Spreng, R.N. et al. (2010) Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage* 53, 303–317
- Botvinick, M.M. et al. (2001) Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652
- Botvinick, M.M. et al. (2004) Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* 8, 539–546
- Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
- Kerns, J.G. et al. (2004) Anterior cingulate conflict monitoring and adjustments in control. *Science* 303, 1023–1026
- Enger, T. (2008) Multiple conflict-driven control mechanisms in the human brain. *Trends Cogn. Sci.* 12, 374–380
- Vergut, T. and Notebaert, M. (2009) Adaptation by binding: a learning account of cognitive control. *Trends Cogn. Sci.* 13, 252–257
- Mayr, U. and Ach, E. (2009) The elusive link between conflict and conflict adaptation. *Psychol. Res.* 73, 794–802
- Rushworth, M.F.S. et al. (2007) Contrasting roles for anterior cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn. Sci.* 11, 169–176
- Etkin, A. et al. (2010) Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* DOI: 10.1016/j.tics.2010.11.004
- Bush, G. et al. (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* 4, 215–222
- Pessoa, L. (2008) On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 9, 148–158
- Pessoa, L. (2010) Emotion and cognition and the amygdala: from 'what is it?' to 'what's to be done?'. *Neuropsychologia* 48, 3416–3429
- Botvinick, M.M. (2007) Multilevel structure in behaviour and in the brain: a model of Fuster's hierarchy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1615–1626
- Norman, D.A. and Shallice, T. (1986) Attention to action: willed and automatic control of behavior. In *Consciousness and Self-regulation. Advances in Research and Theory* (Vol. 4) (Davidson, R.J. et al., eds), In pp. 1–18, Plenum Press
- Posner, M.I. and Rothbart, M.K. (1998) Attention, self-regulation and consciousness. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1915–1927



- 58 Duncan, J. and Owen, A.M. (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483
- 59 Corbetta, M. and Shulman, G. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215
- 60 Gross, J.J. and Thomspon, R.A. (2007) Emotion regulation: conceptual foundations. In *Handbook of Emotion Regulation* (Gross, J.J., ed.), pp. 3–25, Guilford
- 61 Ochsner, K.N. and Gross, J.J. (2005) The cognitive control of emotion. *Trends Cogn. Sci.* 9, 242–249
- 62 Northoff, G. (2005) Is emotion regulation self-regulation? *Trends Cogn. Sci.* 9, 408–409
- 63 Lane, R.D. (2008) Neural substrates of implicit and explicit emotional processes: a unifying framework for psychosomatic medicine. *Psychosom. Med.* 70, 214–231
- 64 Phillips, M.L. *et al.* (2008) A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Mol. Psychiatr.* 13, 833–857
- 65 Berkman, E.T. and Lieberman, M.D. (2009) Using neuroscience to broaden emotion regulation: theoretical and methodological considerations. *Soc. Pers. Psychol. Comp.* 3/4, 475–493
- 66 Price, J.L. *et al.* (1996) Networks related to the orbital and medial prefrontal cortex; a substrate for emotional behavior? *Prog. Brain Res.* 107, 523–536
- 67 Etkin, A. *et al.* (2006) Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* 51, 871–882
- 68 Egner, T. *et al.* (2008) Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cereb. Cortex* 18, 1475–1484
- 69 Bressler, S.L. and Menon, V. (2010) Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290
- 70 Varela, F.J. *et al.* (2001) The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239
- 71 Jack, A. and Roepstorff, A. (2002) Introspection and cognitive brain mapping: from stimulus-response to script-report. *Trends Cogn. Sci.* 6, 333–339
- 72 Lutz, A. *et al.* (2008) Attention regulation and monitoring in meditation. *Trends Cogn. Sci.* 12, 163–169
- 73 Farb, N.A.S. *et al.* (2007) Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Soc. Cogn. Affect. Neurosci.* 2, 313–322