91 Traxler, M.J. and Pickering, M.J. (1996) Plausibility and the processing of unbounded dependencies: an eyetracking study. *J. Mem. Lang.* 35, 454–475

92 Murray, W.S. and Liversedge, S.P. (1994) Referential context effects on syntactic processing. In *Perspectives on Sentence Processing* (Clifton, J. *et al*., eds), pp. 359–388, Erlbaum

93 Altmann, G.T.M. *et al.* (1992) Avoiding the garden path: eye movements in context. *J. Mem. Lang.* 31, 685–712

94 Britt, M.A. (1994) The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *J. Mem. Lang.* 33, 251–283

95 Liversedge, S.P. *et al.* (1998) Processing arguments and adjuncts in isolation and context: the case of by-phrase ambiguities in passives. *J. Exp. Psychol.* 24, 461–475

96 Spivey-Knowlton, M.J. *et al.* (1993) Context effects in syntactic ambiguity resolution. *Can. J. Psychol.* 47, 276–309

97 Spivey, M.J. and Tanenhaus, M.K. (1998) Syntactic ambiguity resolution in discourse:modeling the effects of referential context and lexical frequency. *J. Exp. Psychol.* 24, 1521–1543

98 Trueswell, J.C. *et al.* (1994) Semantic influences on parsing: use of thematic role information in syntactic disambiguation. *J. Mem. Lang.* 33, 285–318

99 MacDonald, M. (1994) Probabilistic constraints and syntactic ambiguity resolution. *Lang. Cognit. Processes* 9, 157–201

100 Spivey-Knowlton, M.K. and Sedivy, J. (1995) Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227–267

101 Traxler, M.J. *et al.* (1998) Adjunct attachment is not a form of lexical ambiguity resolution. *J. Mem. Lang.* 39, 558–592

102 Van Gompel, R.P.G. *et al.* Unrestricted race: a new model of syntactic ambiguity resolution. In *Reading as a Perceptual Process* (Kennedy, A. *et al*., eds), Elsevier (in press)

103 Garrod, S. *et al.* (1994) The role of different types of anaphor in the on-line resolution of sentences in a discourse. *J. Mem. Lang.* 33, 39–68

104 Duffy, S.A. and Rayner, K. (1990) Eye movements and anaphor resolution: effects of antecedent typicality and distance. *Lang. Speech* 33, 103–119

Philosophical conceptions of the self: implications for cognitive science

Shaun Gallagher

Several recently developed philosophical approaches to the self promise to enhance the exchange of ideas between the philosophy of the mind and the other cognitive sciences. This review examines two important concepts of self: the 'minimal self', a self devoid of temporal extension, and the 'narrative self', which involves personal identity and continuity across time. The notion of a minimal self is first clarified by drawing a distinction between the sense of self-agency and the sense of self-ownership for actions. This distinction is then explored within the neurological domain with specific reference to schizophrenia, in which the sense of self-agency may be disrupted. The convergence between the philosophical debate and empirical study is extended in a discussion of more primitive aspects of self and how these relate to neonatal experience and robotics. The second concept of self, the narrative self, is discussed in the light of Gazzaniga's left-hemisphere 'interpreter' and episodic memory. Extensions of the idea of a narrative self that are consistent with neurological models are then considered. The review illustrates how the philosophical approach can inform cognitive science and suggests that a two-way collaboration may lead to a more fully developed account of the self.

S. Gallagher is at the Department of Philosophy and Cognitive Science, Canisius College, Buffalo, NY 14208, USA.

tel: +1 716 888 2329
fax: +1 716 888 3122
e-mail: gallaghr@canisius.edu

Ever since William James[1] categorized different senses of the self at the end of the 19th century, philosophers and psychologists have refined and expanded the possible variations of this concept. James' inventory of physical self, mental self, spiritual self, and the ego has been variously supplemented. Neisser, for example, suggested important distinctions between ecological, interpersonal, extended, private and conceptual aspects of self[2]. More recently, when reviewing a contentious collection of essays from various disciplines, Strawson found an overabundance of delineations between cognitive, embodied, fictional and narrative selves, among others[3]. It would be impossible to review all of these diverse notions of self in this short review. Instead, I have focused on several recently developed approaches that promise the best exchange of ideas between philosophy of mind and the other cognitive sciences and that convey the breadth of philosophical analysis on this topic. These approaches can be divided into two groups that are focused, respectively, on two important aspects of self – the **'minimal' self** and the **'narrative' self** (see Glossary).

1364-6613/00/$ – see front matter PII: S1364-6613(99)01417-5

The first approach involves various attempts to account for a 'minimal' sense of self. Even if all of the unessential features of self are stripped away, we still have an intuition that there is a basic, immediate, or primitive 'something' that we are willing to call a self. This approach leaves aside questions about the degree to which the self is extended beyond the short-term or 'specious present' to include past thoughts and actions. Although continuity of identity over time is a major issue in the philosophical definition of personal identity, the concept of the minimal self is limited to that which is accessible to immediate self-consciousness. Certain aspects of the minimal self are relevant to current models in robotics. Furthermore, aspects of the minimal self that involve senses of ownership and agency in the context of both motor action and cognition can be clarified by neurocognitive models of schizophrenia that suggest the involvement of specific brain systems (including prefrontal cortex, supplementary motor area, and cerebellum) in the manifestation of neurological symptoms in this disorder.

A second approach to the concept of self involves conceiving of the self in terms of narrative. This notion was imported into the cognitive sciences by Dennett[4], but it might have a more complex significance than is indicated in Dennett's account. The narrative self is extended in time to include memories of the past and intentions toward the future. It is what Neisser refers to as the extended self, and what Dennett calls a 'nonminimal selfy' self. Neuropsychological descriptions of episodic memory and its loss can help to circumscribe the neural substrates of the narrative self.

Glossary

Immunity principle: When a speaker uses the first-person pronoun ('I') to refer to him or herself, she cannot make a mistake about the person to whom she is referring. Philosophers call this 'immunity to error through misidentification relative to the first-person pronoun'[5].

Minimal self: Phenomenologically, that is, in terms of how one experiences it, a consciousness of oneself as an immediate subject of experience, unextended in time. The minimal self almost certainly depends on brain processes and an ecologically embedded body, but one does not have to know or be aware of this to have an experience that still counts as a self-experience.

Narrative self: A more or less coherent self (or self-image) that is constituted with a past and a future in the various stories that we and others tell about ourselves.

Non-conceptual first-person content: The content of a primitive self-consciousness that is not informed by conceptual thought. For example, the ecological content of perception that specifies one's own embodied position in the environment.

Sense of agency: The sense that I am the one who is causing or generating an action. For example, the sense that I am the one who is causing something to move, or that I am the one who is generating a certain thought in my stream of consciousness.

Sense of ownership: The sense that I am the one who is undergoing an experience. For example, the sense that my body is moving regardless of whether the movement is voluntary or involuntary.

Self-reference and misidentification

There are a number of ways to understand the notion of a minimal sense of self. In this section, I approach the problem by discussing how we use the first-person pronoun in a self-referring way that should never permit a mistake. This kind of self-reference has a feature that some philosophers call 'immunity to error through misidentification relative to the first-person pronoun'[5]. I will refer to this as the **immunity principle** (see Glossary). Once this principle is clarified we can ask whether, in actuality, it can ever fail and if so, what this might reveal. In the next section, I will explore this possibility in relation to a neurocognitive model of schizophrenia that requires us to make a distinction between two aspects of the minimal sense of self: the sense of self-ownership and the sense of self-agency.

Wittgenstein distinguished between two uses of the first-person pronoun in self-reference: 'as subject' and 'as object'[6]. Use of the first-person pronoun as subject might best be discerned by understanding what a speaker could be wrong about, and the kinds of questions that one could sensibly ask them. For example, if someone says 'I think it is raining outside', she could be wrong about the rain. It might not be raining. But it seems that she could not be wrong about the 'I'. That is, she could not misidentify herself when she states that it is she who is thinking. So, according to Wittgenstein, the following question would be nonsensical: 'Are you sure that *you* are the one who thinks it is raining?' Such use of the first-person pronoun is immune to error through misidentification. By contrast, when we use the first-person pronoun 'as object' it is possible to misidentify ourselves. For example, in some experimental situations a subject's arm may be deafferented (that is, the subject is deprived of normal proprioceptive feedback about the position of their limb and therefore cannot keep track of it without vision). Their visual perception of arm movements are then manipulated through mirrors or videotape[7,8]. In such cases, the subject might be led to say, 'I am moving my arm to the left', when in fact the basis for his judgment is a videotape of someone else moving their own arm to the left. In that case, the subject makes a mistake about *who* is moving their arm to the left. To say 'I' in such a case involves an objective misidentification of oneself.

Shoemaker[5] suggests that the immunity principle applies only to the use of 'I' as subject because when we use the first-person pronoun as subject we are not actually attempting to identify ourselves. In other words, when I self-refer in this way I do not go through a cognitive process in which I try to match up first-person experience with some known criterion in order to judge the experience to be my own. My access to myself (my *self*) in first-person experience is immediate and non-observational; that is, it doesn't involve a perceptual or reflective act of consciousness. In this sense, the immediate self that is referred to here is the pre-reflective point of origin for action, experience and thought.

Are there any exceptions to the immunity principle? Is there any instance of someone using a first-person pronoun as subject, and being wrong in their reference? Following suggestions made by Feinberg[9] and Frith[10] about certain schizophrenic experiences (including auditory hallucination, thought insertion, and delusions of control in which subjects report that their body is under the control of other people or things), Campbell has proposed that such experiences might be counterexamples to the immunity principle[11]. A schizophrenic patient who suffers thought insertion, for example, might claim that she is not the one who is thinking a particular thought, when in fact she is the one who is thinking the thought. The following example of a schizophrenic's account of her own thought processes illustrates this: 'Thoughts are

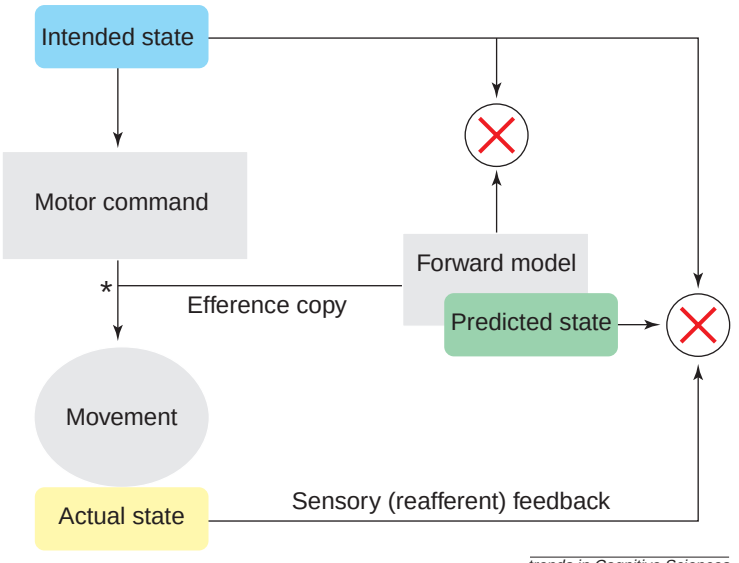

Fig. 1. The forward and feedback comparators. This model represents processes that generate two aspects of the 'minimal self' in normal experience (see text). Match at the forward comparator provides a sense of agency for movement; match at the feedback comparator provides a sense of ownership for movement.

put into my mind like "Kill God". It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts' (Ref. 10, p. 66). In such cases the schizophrenic patient misidentifies the source of the thought and seemingly violates the immunity principle.

It may be argued whether or not Campbell is correct in his claim that this is a counterexample to the immunity principle[12], but the implications of his analysis are quite productive. His argument implies that a scientific explanation of schizophrenic phenomena such as thought insertion might also count as a scientific explanation of how the immunity principle works. Frith's neurocognitive model of the breakdown of self-monitoring in schizophrenia turns out to be a good candidate for explaining immunity to error through misidentification. If we can identify which mechanisms fail at the cognitive or neurological level when the schizophrenic patient suffers from thought insertion, then we also have a good indication of the mechanisms responsible for (or at least involved in) the normal immunity to error found in self-reference, and the immediate sense of self. This insight moves us from the conceptual and often abstract arguments of philosophy to the more empirical inquiries of neuropsychology and neurophysiology.

A neurocognitive model of immediate self-awareness

A brief consideration of motor action will help to clarify two closely related aspects of minimal self-awareness: self-ownership – the sense that it is my body that is moving; and self-agency – the sense that I am the initiator or source of the action. In the normal experience of voluntary or willed action, the **sense of agency** and the **sense of ownership** coincide and are indistinguishable. When I reach for a cup, I know this to be my action. This coincidence may be what leads us to think of ownership of action in terms of agency: that the owner of an action is the person who is, in a specific way, causally involved in the production of that action, and is thus the author of the action. In the case of involuntary action, however, it is quite possible to distinguish between sense of agency and sense of ownership. I may acknowledge ownership of a movement – that is, I have a sense that I am the one who is moving or is being moved – and I can self-ascribe it as *my* movement, but I may not have a sense of causing or controlling the movement. That is to say, I have no sense of agency. The agent of the movement is the person who pushed me from behind, for example, or the physician who is manipulating my arm in a medical examination. Thus, my claim of ownership (my self-ascription that I am the one who is undergoing an experience) can be consistent with my lack of a sense of agency. Phenomena such as delusions of control, auditory hallucinations, and thought insertion appear to involve problems with the sense of agency rather than the sense of ownership[13].

There is good evidence to suggest that the sense of ownership for motor action can be explained in terms of ecological self-awareness built into movement and perception[2,14]. By contrast, experimental research on normal subjects suggests that the sense of agency for action is based on that which precedes action and translates intention into action[15,16]. In addition, research that correlates initial awareness of action with scalp recordings of the lateralized readiness potential in motor cortex, and with transcranial magnetic stimulation of the supplementary motor area, strongly indicates that one's initial awareness of a spontaneous voluntary action is tied to the anticipatory or pre-movement motor commands relating to relevant effectors[17,18].

It turns out that some schizophrenic patients who suffer from thought insertion also make mistakes about the agency of various bodily movements. To explain this, Frith[10] appeals to the notions of efference copy and comparator mechanisms that were originally used to explain motor control[19,20]. According to the most recent version of this model, a comparator mechanism operates as part of a non-conscious premotor or 'forward model' that compares efference copy of motor commands with motor intentions and allows for rapid, automatic error corrections[21,22]. Such a mechanism is consistent with the findings cited above. This comparator process anticipates the sensory feedback from movement and underpins an online sense of self-agency that complements the ecological sense of self-ownership based on actual sensory feedback[12] (Fig. 1). If the forward model fails, or efference copy is not properly generated, sensory feedback may still produce a sense of ownership ('I am moving') but the sense of agency will be compromised ('I am not causing the movement'), even if the actual movement matches the intended movement[23].

Schizophrenic patients who suffer from thought insertion and delusions of control also have problems with this forward, pre-action monitoring of movement, but not with motor control based on a comparison of intended movement and sensory feedback[24,25]. The control based on sensory feedback is thought to involve the cerebellum[21]. By contrast, problems with forward monitoring are consistent with studies of schizophrenia that show abnormal pre-movement brain potentials associated with elements of a neural network involving supplementary motor, premotor and prefrontal cortexes[26]. Problems with these mechanisms might therefore result in the

lack of a sense of agency that is characteristic of these kinds of schizophrenic experience.

Following a suggestion made by Feinberg[9], Frith postulates a similar model for cognition – specifically, for thought and inner speech[10]. Phenomena such as thought insertion, hearing voices, or perceiving one's own acts as alien, suggest that something is wrong with the self-monitoring mechanism. Frith's model assumes not only that thinking, insofar as it is intended and self-generated, is a kind of action, but also that thinking has to match the subject's intention for it to feel self-generated, as in the case of a motor action. This suggests that, although such intentions are not always consciously accessible, comparator processes that match intentions to the generation of thought and to the stream of thought bestow, respectively, a sense of agency and a sense of ownership for thought, in a similar way to motor action. If the mechanism that constitutes the forward aspect of this monitoring process fails, a thought occurs in the subject's own stream of consciousness but does not seem to the subject to be self-generated or to be under the subject's control. Rather, it appears to be an alien or inserted thought (Fig. 2).

Whether or not this kind of model is adequate to account for the phenomenon of inserted thought, I would suggest that the approaches taken by Frith and Campbell promise a way to explain in specific neurological terms the immediacy involved in the senses of self-ownership and self-agency, and in the immunity principle. Such aspects of the minimal self may find a neurological explanation in the proper workings of the mechanisms described above and are threatened by their failure.

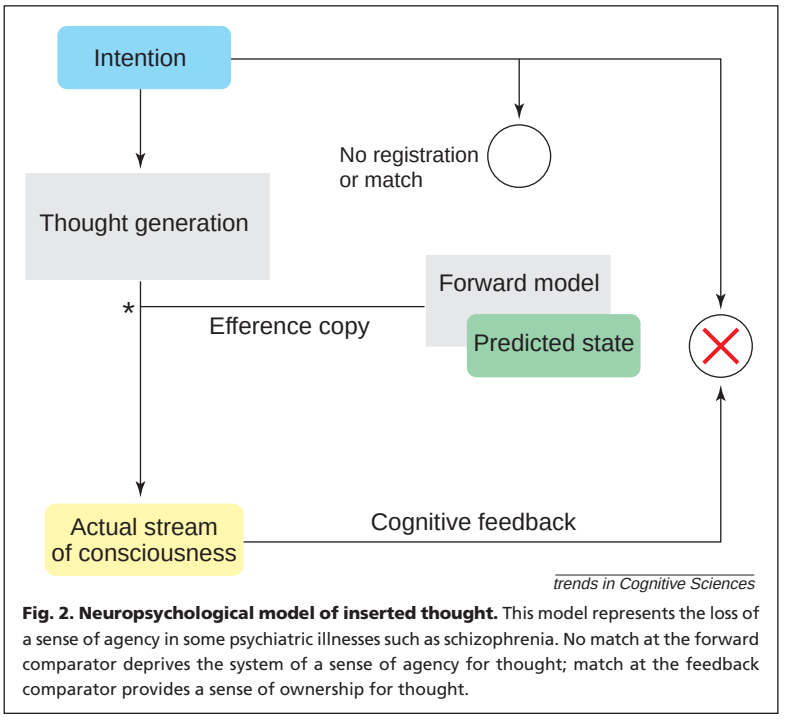

Fig. 2. Neuropsychological model of inserted thought. This model represents the loss of a sense of agency in some psychiatric illnesses such as schizophrenia. No match at the forward comparator deprives the system of a sense of agency for thought; match at the feedback comparator provides a sense of ownership for thought.

The minimal self: embodied or disembodied?

Taking the immunity principle as a point of departure, there are two other directions that one could follow. The first explores the idea that there is an even more primitive and embodied sense of self than that involved in the use of the first-person pronoun. This approach pursues the implications of what developmental psychologists have recently discovered about the experience of neonates. The second involves a more abstract self-reflective access to first-person experience, which, among other things, leads to issues that concern AI applications in robotics.

Taking the first approach, are there any aspects of the minimal self that are more primitive than those identified in the immunity principle? In speaking about self-reference, we are already speaking of a self that is capable of linguistic communication – at the very least, the person is capable of using the first-person pronoun. If one considers that language and conceptual capacity develop in parallel, this might mean that the person's immediate and pre-reflective access to the self already involves the mediation of a conceptual framework. Is it possible to speak of a non-conceptual access to the self – a more primitive self-consciousness that does not depend on the use of a first-person pronoun?

Bermúdez explores the many implications of this question[27]. Following on from Gibson's ecological psychology, part of what Bermúdez calls **'non-conceptual first-person content'** consists of the self-specifying information attained in perceptual experience. When I perceive objects or movement in the external environment, I also gain information about myself – information that is pre-linguistic and non-conceptual. This is what Neisser calls the ecological self[2]. The fact that non-conceptual, ecological self-awareness exists from the very beginning of life can be demonstrated by the important role it plays in neonatal imitation. Neonates less than an hour old are capable of imitating the facial gestures of others in a way that rules out reflex or release mechanisms, and that involves a capacity to learn to match the presented gesture[28,29]. For this to be possible the infant must be able to do three things: (1) distinguish between self and non-self; (2) locate and use certain parts of its own body proprioceptively, without vision; and (3) recognize that the face it sees is of the same kind as its own face (the infant will not imitate non-human objects[30]). One possible interpretation of this finding is that these three capacities present in neonates constitute a primitive self-consciousness, and that the human infant is already equipped with a minimal self that is embodied, enactive and ecologically tuned[31–33].

One can, however, move in a second direction and ask whether it is possible to capture and make explicit the pre-reflective minimal self in a reflective, and conceptually informed, introspection. In this case, one may still talk about the most abstract aspect of what we experience to be ourselves, even if it is mediated through reflection. Galen Strawson's recent essays on the self make it clear that he is seeking the most basic and stripped-down version of a self that can still be called self[3,34,35]. He begins with a reflective description of his *experience* of the self. This phenomenologically reflective approach then naturally leads to a characterization of the self as a subject of experience. Thus, Strawson is led to define the self as a subject of experience that is a single (hiatus-free) mental thing. This is a momentary self without long-term continuity, and thus, without a history – 'a bare locus of consciousness, void of personality' (Ref. 3, p. 492).

On this view, a human being consists of a series of such transient selves, each one lasting only as long as a unique period of experience lasts, coming into existence and going out of

Box 1. Robotics and the minimal self

Tani explores the possibility of establishing an artificial version of Strawson's minimal self in a machine (Ref. a). He takes up the challenge of developing an objective definition of this concept in the context of robot design, although he is still forced to use terms like 'subjective mind' and 'self-consciousness' in his objective account (Ref. a, pp. 150, 173). However, in contrast to Strawson, Tani makes it clear that the robotic self he is designing is the result of physical interaction between the robotic body and its environment. Specifically, its short-term existence is generated only in cases where the interaction fails to go smoothly. What Tani retains from Strawson's model is the idea that the self is only a momentary phenomenon, called into and going out of existence from one moment to the next, as the situation requires.

Following theoretical suggestions made by Varela *et al.* (Ref. b), Tani constructed a robot (Fig. I) that operates by integrating information derived from bottom-up (sensory–motor) processes that represent environmental conditions, with top-down, abstract, predictive modeling of the world. In general, robotic processes run smoothly as long as there is a good relative match between the top-down model and the bottom-up input. Problems occur when there are inconsistencies between these two pathways. These problems are difficult to resolve in hybrid systems where top-down mechanisms are designed to follow the logic of symbol manipulation, and bottom-up processes are based on analogue pattern-matching. By contrast, Tani's design bases both pathways on dynamic systems, so that their interface takes place in a shared metric space. During unsteady or conflicting phases of the system dynamics, an arbitration process takes place in that shared space, and the robot is required to take its own current state into account. Specifically it needs to take into account ('become aware of') the conflict in its own system, and its own degree of familiarity with the surrounding environment. This, in Tani's view, is the robotic equivalent of self-consciousness. A self comes into existence when the relationship between top-down 'mental' processes and bottom-up sensory-motor processes becomes incoherent; that is, in the event of a failure to cope with environmental demands. This self is not an entity that has continuity over the long-term, but is a short-term phenomenon (consistent with Strawson's view), and in this case one that emerges only on occasions that motivate self-reference.

References

a Tani, J. (1998) An interpretation of the 'self' from the dynamical systems perspective: a constructivist approach. *J. Conscious. Stud.* 5, 516–542

b Varela, F., Thompson, E. and Rosch, E. (1991) *The Embodied Mind*, MIT Press

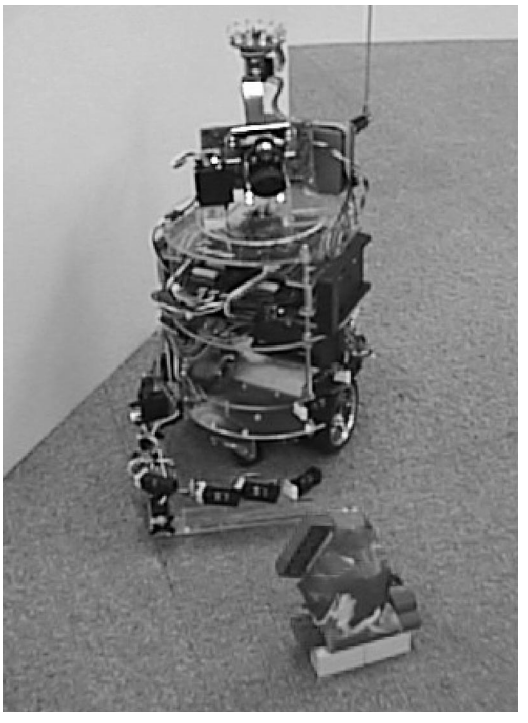

Fig. I. Tani's vision-based robot. (Reproduced, with permission, from Ref. a.)

existence, without continuity. Despite the 'local' character of Strawson's approach, that is, an approach that focuses on his own experience, the self that he seeks to define is not restricted to the human case. It would be entirely possible for the immediate self he describes to be instantiated in a non-human animal that has the right cognitive equipment. It has been argued that it might even be possible to create the minimal self in a machine, or more precisely, in a robot (see Box 1), but this would entail dropping the part of Strawson's proposal that defines the self as a *conscious* subject of experience.

Furthermore, on Strawson's view, it is not *essential* that this minimal self be embodied or enactive within an environment. The self-consciousness that captures this self is not ecologically embedded, but is one that operates on a conceptual level, already in possession of the concept 'self'. Strawson is nonetheless a materialist, and considers the self as 'mental thing' to be a physical entity which, in the human case, is likely to be manifest in terms of brain processes. What is metaphysically the case, however, is not always revealed in the phenomenological record: I can be conscious of myself as a minimal subject of experience that is a single mental thing, without being aware of the embodiment or brain functions that may (or may not) generate the self. This is entirely consistent with self-reference that is immune from error through misidentification, as it specifies an access to the self that does not depend on applying empirical (in this case physical) criteria of identity. Even if it is the case that the information that constitutes the minimal self is generated in ecologically embodied experience, and even if, in practice, a human being is capable of knowing that this is the case, one does not gain the self-consciousness that goes along with the minimal self by knowing this to be true or by being able to employ empirical criteria to verify it.

The self extended and mediated by narrative

So far we have considered only a minimal self, which is a concept of self that seems quite at odds with our commonsense conception of who we are. Surely we think and speak of ourselves as entities extended in time? Indeed, it seems undeniable that we have memories and that we make plans, and that there is continuity between our past and our future. And do we not, as selves with individual identity, encompass that continuous experience? What is the nature of this sense of a continuous self? Is it carried by a succession of momentary minimal selves that are tied together by real connections? Or

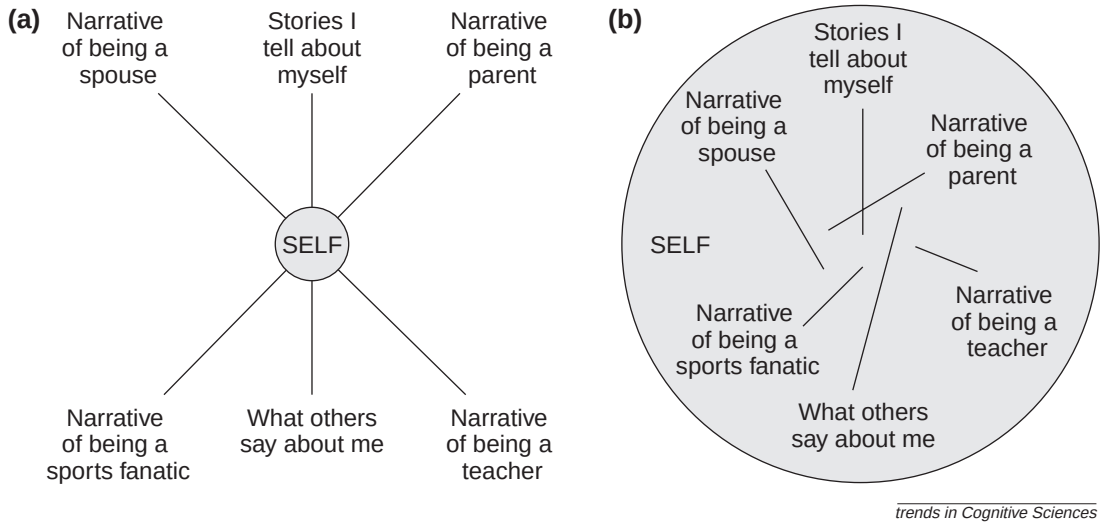

Fig. 3. Two models of the narrative self. (a) Center of narrative gravity. In this model (formulated by Dennett[4]), the self is defined as an abstract 'center of narrative gravity', where the various stories told about the person, by himself and others, meet. **(b)** An extended and more distributed model of the narrative self, less unified than the one proposed by Ricoeur, but importantly distinct from Dennett's model in that the self is not an abstract 'center', but rather, an extended self which is decentered and distributed.

are momentary minimal selves simply abstractions from a more substantial continuity that is the more genuine self? The philosophical traditions are replete with a variety of answers to these questions.

One famous answer given by Hume suggests that the self consists of a bundle of momentary impressions that are strung together by the imagination[36]. In effect, an extended self is simply a fiction, albeit a useful one because it lends a practical sense of continuity to life, but a fiction nonetheless. The narrative theory of self is a contemporary reading of this view. Dennett offers one version of this theory which he sees as consistent with recent developments in our understanding of how the brain functions[4,37]. The consensus from contemporary neuroscience is that neurological processing is for the most part distributed across various brain regions, and it cannot be said that there is a real, neurological center of experience. Thus, there is no real simplicity of experience at one time nor real identity across time that we could label the self. At best, we might refer to a minimal biological self as something real. But the latter is nothing more than a principle of organization involving the distinction between self and non-self. Furthermore, this principle is found throughout living nature, and is not something sufficient for the purpose of a coherent continuity or identity over time, such as is found at the level of human experience. Humans, however, do have something more than this – we have language. And with language we begin to make our experience relatively coherent over extended time periods. We use words to tell stories, and in these stories we create what we call our selves. We extend our biological boundaries to encompass a life of meaningful experience.

Two things are to be noted from Dennett's account. First, we cannot prevent ourselves from 'inventing' our selves. We are hardwired to become language users, and once we are caught up in the web of language and begin spinning our own stories, we are not totally in control of the product. As Dennett puts it, 'for the most part we don't spin them [the stories]; they spin us' (Ref. 4, p. 418). Second, an important product of this spinning is the narrative self. The narrative self, however, is nothing substantially real. Rather, it is an empty abstraction. Specifically, Dennett defines a self as an abstract 'center of narrative gravity,' and likens it to the theoretical fiction of the center of gravity of any physical object. In the case of narrative gravity, however, an individual self consists of the abstract and movable point where the various stories (of fiction or biography) that the individual tells about himself, or are told about him, meet up (Fig. 3a).

The notion of a narrative self-constitution finds confirmation in psychology and neuroscience. In the former, Neisser's concepts of the extended and the conceptual self, initially explained in terms of memory, have been enhanced by considerations of the role that language and narrative play in developing our own self-concept[38]. In the realm of neuroscience, Gazzaniga has suggested that one function of the left hemisphere of the brain is to generate narratives, using what he terms an 'interpreter'. Gazzaniga proposed this interpreting function based on his studies of split-brain patients. In these patients, the left hemisphere has no internal access to right-hemisphere experience because the corpus callosum has been severed. Nonetheless, in properly designed experimental circumstances, the left hemisphere devises interpretations for meanings, actions and emotions produced by the right hemisphere. Such interpretations show consistency with the experiential context belonging to the left hemisphere rather than with the original right-hemisphere context. The left hemisphere, for example, might remain ignorant of the content or cause of an emotion generated in the right hemisphere, but the left-hemisphere experience of that emotion motivates an interpretation of the event in terms relevant to the content available to the left hemisphere. In Gazzaniga's model, the interpreter weaves together autobiographical fact and inventive fiction to produce a personal narrative that enables the sense of a continuous self[39,40]. Gazzaniga, however, contends that the self, in this regard, is not a fiction because the normal functioning of the interpreter tries to make sense of what actually happens to the person. At most, in the non-pathological case, it may be only 'a bit fictional' (Ref. 41,

Outstanding questions

- What relationship exists between the minimal self and the narrative self? Is one generated from the other? Do they operate independently of each other?
- Shoemaker[5] has maintained that immunity to error does not necessarily extend to episodic memory. What status do truth-claims concerning episodic memory have?
- Because the sense of self-agency is absent both in the case of involuntary thoughts and in cases of schizophrenic experiences of inserted thoughts, the lack of a sense of self-agency by itself cannot fully explain the schizophrenic patient's sense that certain of his thoughts belong to someone else. What else is required to explain this?
- If some aspect of the minimal self depends on a forward model of control, this complicates Tani's design for a self-referring robot. To what extent can feedback mechanisms by themselves provide the requisite minimal reference to self?
- What are the precise neurological mechanisms involved in the left-hemisphere interpreter?
- In narrative theory, self-constitution is meant to imply a situation in which the self is both the narrator and the narrated. If there is more than one narrator (if we are also constituted by stories about ourselves told by others), what mechanisms at the psychological level integrate or adjudicate these diverse constitutions?

p. 713). Perhaps we cannot help but enhance our personal narratives with elements that smooth over discontinuities and discrepancies in our self-constitution.

A necessary condition for the non-fictional aspects of a narrative self is the proper working of episodic memory. Pribram suggests that this depends on a fronto-limbic system that includes the anterior poles of the frontal and temporal lobes, and elements of the limbic formation[42]. Specifically, this system is involved in providing a sense of time. The importance of the proper functioning of episodic memory and time-sense on the formation of the narrative self is indicated by the case of a young boy diagnosed with congenital damage to the right hemisphere and frontal cortex. He suffers from a profound episodic amnesia and because he lacks the ability to quantify the passage of time or to appreciate the meaning of temporal units[42,43], he is unable to formulate certain essential structures of narrative, namely, sequential structure and the demarcations of beginning and end.

Further extensions of the narrative self

In the current context of contentious disagreements on a large range of issues surrounding the self[3], a general consensus among a diverse group of cognitive scientists concerning the constitution of the narrative self might seem surprising. However, it is perhaps even more surprising that there is some consensus within philosophy on this point, even across the great divide between continental and analytical philosophers. What Dennett, Neisser, Gazzaniga and Pribram have to say about the narrative self echoes in some respects an earlier discussion among continental philosophers. Perhaps Ricoeur is the best representative of this earlier discussion concerning the nature of narrative and the making of the narrative self[44,45]. Ricoeur carefully explored these issues in depth, and reached conclusions that are not inconsistent with the view outlined in the cognitive science discussion above. In contrast to Dennett, however, Ricoeur conceives of the narrative self, not as an abstract point at the intersection of various narratives, but as something richer, more substantial and concrete. Ricoeur insists that, importantly, one's own self narrative is always entangled in the narratives of others.

We may extend Ricoeur's model beyond what he takes to be a unified life narrative and suggest that the self is the sum total of its narratives, and includes within itself all of the equivocations, contradictions, struggles and hidden messages that find expression in personal life. In contrast to Dennett's center of narrative gravity, this extended self is decentered, distributed and multiplex (Fig. 3b). At a psychological level, this view allows for conflict, moral indecision and self-deception, in a way that would be difficult to express in terms of an abstract point of intersection. Furthermore, with respect to neurological models, this extended model is even more consistent than Dennett's abstract center with the concept of distributed processing, and with what Gazzaniga describes as the mixing of fact and fiction by the left-hemisphere 'interpreter'. By extending the ideas of a narrative self, we are perhaps coming closer to a concept of the self that can account for the findings of the cognitive sciences and neurosciences, as well as our own experience of what it is to be a continuous, phenomenological self.

Concluding remarks

In a recent book, Damasio has insightfully captured the difficulty involved in expressing the interrelations between the minimal ('core') self and the narrative ('autobiographical') self[46]. The difficulty is due to complexities that are apparent on both the personal and the sub-personal, neurological levels. Episodic memory, which is necessary for the construction of the narrative self, is subject to constant remodeling under the influence of factors that include innate and acquired dispositions as well as social and cultural environments. The registration of episodic memory as 'my' memory of 'myself' clearly depends on a minimal but consistently reiterated sense of self that I recognize, without error, as myself. In some respects, as Damasio insists, this depends on narrowly defined, embodied capabilities and feelings. In other regards, however, the core features of the self are constantly being reinterpreted by the narrative process. In the neurological terms that Damasio uses this means that there are extremely complex demands made on the processes that link early sensory cortexes that hold information on the minimal or core self, and convergence or dispositional zones that contribute to the generation of the narrative self. In this regard, he makes it clear that at present the neuroscientist, like the philosopher, can offer, at best, informed speculation on these processes.

In this review, I have tried to show that philosophical ideas about the self can be aligned with, and can inform, current ideas in cognitive science. I also believe that philosophers can learn about the nature of the self from psychologists, neuroscientists and other cognitive scientists. Thus, collaborative efforts between philosophers and scientists promise to open up more subtle and sophisticated avenues of research, which will define more fully the concept of the self.

Acknowledgements

I thank S-J. Blakemore, U. Neisser, G. Strawson, and anonymous reviewers for helpful comments on an earlier version of this paper. Parts of my research were supported by a fellowship at the NEH Summer Institute on Mind, Self, and Psychopathology, directed by J. Whiting and L. Sass at Cornell University in 1998, and by a sabbatical leave from Canisius College in 1999.

References

1 James, W. (1890) *The Principles of Psychology*, Dover Publ. (reprinted 1950)
2 Neisser, U. (1988) Five kinds of self-knowledge. *Philos. Psychol.* 1, 35–59
3 Strawson, G. (1999) The self and the SESMET. In *Models of the Self* (Gallagher, S. and Shear, J., eds), pp. 483–518, Imprint Academic
4 Dennett, D. (1991) *Consciousness Explained*, Little Brown & Co.
5 Shoemaker, S. (1984). *Identity, Cause, and Mind*, Cambridge University Press
6 Wittgenstein, L. (1958) *The Blue and Brown Books,* Blackwell
7 Jeannerod, M. (1994) The representing brain: neural correlates of motor intention and imagery. *Behav. Brain Sci.* 17, 187–245.
8 Daprati, E. *et al*. (1997) Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition* 65, 71–86
9 Feinberg, I. (1978) Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophr. Bull.* 4, 636–640
10 Frith, C. (1992) *The Cognitive Neuropsychology of Schizophrenia*, Erlbaum
11 Campbell, J. Schizophrenia, the space of reasons and thinking as a motor process. *The Monist* (in press)
12 Gallagher, S. Self-reference and schizophrenia: a cognitive model of immunity to error through misidentification. In *Problems of the Self* (Zahavi, D. and Parnas, J. eds), John Benjamins (in press)
13 Stephens, G.L. and Graham, G. (1994) Self-consciousness, mental agency, and the clinical psychopathology of thought insertion. *Philos. Psychiatry Psychol.* 1, 1–12
14 Gallagher, S. and Marcel, A.J. (1999) The self in contextualized action. *J. Conscious. Stud.* 6, 4–30
15 Marcel, A.J. The sense of agency: awareness and ownership of actions and intentions. In *Agency and Self-Awareness* (Roessler, J. and Eilan, N. eds), Oxford University Press, (in press)
16 Fourneret, P. and Jeannerod, M. (1998) Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* 36, 1133–1140
17 Haggard, P. and Eimer, M. (1999) On the relation between brain potentials and the awareness of voluntary movements. *Exp. Brain Res.* 126, 128–133
18 Haggard, P. and Magno, E. (1999) Localising awareness of action with transcranial magnetic stimulation. *Exp. Brain Res.* 127, 102–107
19 Sperry, R.W. (1950) Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* 43, 482–489
20 Holst, E. von and Mittelstaedt, H. (1950). Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* 37, 464–476
21 Frith, C.D. *et al.* Abnormalities in the perception and control of action. *Proc. R. Soc. Lond. Ser. B* (in press)
22 Georgieff, N. and Jeannerod, M. (1998) Beyond consciousness of external events: a 'who' system for consciousness of action and self-consciousness. *Conscious. Cognit.* 7, 465–477
23 Spence, S.A. *et al*. (1997) A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain* 120, 1997–2011
24 Frith, C.D. and Done, D.J. (1988) Towards a neuropsychology of schizophrenia. *Br. J. Psychiatry* 153, 437–443
25 Malenka, R.C. *et al*. (1982) Impaired central error correcting behaviour in schizophrenia. *Arch. Gen. Psychiatry* 39, 101–107
26 Singh, J.R. *et al*. (1992) Abnormal premovement brain potentials in schizophrenia. *Schizophr. Res.* 8, 31–41
27 Bermúdez, J. (1998) *The Paradox of Self-Consciousness*, MIT Press
28 Meltzoff, A. and Moore, M.K. (1977) Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78
29 Meltzoff, A. and Moore M.K. (1983) Newborn infants imitate adult facial gestures. *Child Dev.* 54, 702–709
30 Legerstee, M. (1991) The role of person and object in eliciting early imitation. *J. Exp. Child Psychol.* 51, 423–433
31 Bermúdez, J. (1996) The moral significance of birth. *Ethics* 106, 378–403
32 Gallagher, S. (1996) The moral significance of primitive self-consciousness. *Ethics* 107, 129–140
33 Rochat, P., ed. (1995) *The Self in Infancy: Theory and Research*, Elsevier
34 Strawson, G. (1997) The self. *J. Conscious. Stud.* 4, 405–428
35 Strawson, G. (1999) Self, body, and experience. *Proceedings of the Aristotelian Society (Supplement)* 73, 307–332
36 Hume, D. (1739) *A Treatise of Human Nature*, Clarendon Press (reprinted 1975)
37 Dennett, D. (1988) Why everyone is a novelist. *Times Literary Supplement* Sept.16–22, pp. 1016, 1028–1029
38 Neisser, U. and Fivush, R. (1994) *The Remembering Self : Construction and Accuracy in the Self-Narrative*, Cambridge University Press
39 Gazzaniga, M. (1998) *The Mind's Past*, Basic Books
40 Gazzaniga, M. (1995) Consciousness and the cerebral hemispheres. In *The Cognitive Neurosciences* (Gazzaniga, M., ed), pp. 1391–1400, MIT Press
41 Gazzaniga, M. and Gallagher, S. (1998) The neuronal Platonist. *J. Conscious. Stud.* 5, 706–717
42 Pribram, K.H. (1999) Brain and the composition of conscious experience. *J. Conscious. Stud.* 6 , 19–42
43 Ahern, C.A. *et al*. (1998) Preserved semantic memory in an amnesic child. In *Brain and Values: Is a Biological Science of Values Possible?* (Pribram, K.H., ed.), pp. 277–297, Erlbaum
44 Ricoeur, P. (1984) *Time and Narrative* (3 Vols), University of Chicago Press
45 Ricoeur, P. (1992) *Oneself As Another*, University of Chicago Press (transl. from *Soi-même comme un autre*, Editions du Seuil, 1990)
46 Damasio, A. (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Harcout Brace