

In recent philosophy of mind, informed by ongoing research in the cognitive neurosciences, there has been a tendency to offer deflationary or reductive explanations of self and self-identity. The background to such accounts includes a complex history of the problem of personal identity from Hume to Parfit. Paul Ricoeur has provided an insightful perspective on this history based on his distinction between *ipse* identity and *idem* identity.¹ My intention is not to rehearse that history, or even to update it, but rather, assuming that history as background, to engage in a dialogue with the more recent proponents of deflationary, reductive, and internalist accounts of the self, and especially those accounts that put the notion of a narrative self into question.²

Philosophically there is some question as to whether an account of the narrative self is a sufficient account of the self more generally, that is, whether the whole story about the self is just that it is *nothing more than* a story about the self (this is one version of a deflationary account of the self). There is, nevertheless, a growing consensus that the concept of the narrative self captures something essential about human existence. In this regard, specifically, there are three issues that I want to address. First, is it possible to defend an account of the narrative self which is consistent with discoveries in cognitive neuroscience, but that remains nonetheless non-reductive? Second, how does such an account relate to an embodied-enactive approach to questions of self-identity? And third, to what extent does such an account involve dimensions of intersubjectivity?

The appeal to neuroscience may seem misplaced to some readers. Ricoeur is clear in his insistence that questions of the mind, addressed by phenomenology and hermeneutics, belong to a discourse different from and irreducible to neuroscience.³ To speak of the body from a phenomenological perspective, or from the perspective of lived action, is quite different from speaking of the body from the perspective of natural science. Yet Ricoeur is careful not to suggest an ontological dualism; rather we have here a semantic dualism, two different languages attempting to give an account of human action. He asks a basic question: “does any knowledge that I may have of the brain add to the knowledge that I have of myself simply through direct acquaintance with my body, without knowing anything about my brain?”⁴ To this question I want to give a qualified affirmative answer. To the extent that I am not just a mind, and not just a lived body, and to the extent that my existence depends on a biological constitution, knowledge about the brain can enrich my knowledge of myself and of the capacities that make me what I am. The idea is not to replace one discourse with another, but to supplement one with the other. Furthermore, the idea is not simply to look for correlations, but to show a certain consistency between these two ap-

¹ Paul Ricoeur, *Oneself as Another*, trans. Kathleen Blamey (Chicago: University of Chicago, 1992).

² For purposes of this paper, a *deflationary* account of the self is any attempt to explain the self as simply a product of language, for example, our use of the first-person pronoun. For a good discussion of such an account, see Jose L. Bermudez, *The Paradox of Self-Consciousness* (Cambridge, Mass.: MIT Press, 1998), 9ff. An *internalist* account of the self is any attempt to explain the self as a product of narrowly defined mental or physical processes, where such an account would exclude or ignore embodied action in social contexts. A *reductionist* account is any attempt to explain the self entirely in terms of brain processes. These distinctions and definitions are oversimplified, but nonetheless sufficient for purposes of my general argument, which treats all such accounts as one large target.

³ Jean-Pierre Changeux and Paul Ricoeur, *What Makes Us Think?* trans. Malcolm B. DeBevoise (Princeton, N.J.: Princeton University Press, 2000), 14; also Ricoeur, *Oneself as Another*, 132.

⁴ *Ibid.*, 20.

proaches. If neurobiology supports the conclusions reached by a hermeneutically informed narrative theory, and if narrative theory makes sense out of certain discoveries in neuroscience, I see no reason not to strive for a dialogical enlightenment. Likewise, I see no reason to ignore those areas where neurobiology and narrative theory conflict, for they present the possibility of improving accounts.

Two Models of the Narrative Self

We normally think and speak of ourselves as entities extended in time. We have memories and we make plans. There seems to be a continuity between our past and our future which, as individually consistent selves, we encompass in a more or less coherent way. This notion of a continuous self, however, has been under attack since Hume's famous analysis of personal identity.⁵

Hume suggested that the self consists of a bundle of momentary impressions that are strung together by the imagination. 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. Recently, cognitive scientists have joined forces with philosophers of mind to show how neuroscientific explanations also undermine the notion of an extended, unified self.⁶

Neuroscientists find in the brain something analogous to what Hume had found in the mind, a collection of distributed processes with no central theater, no real neurological center of experience. Thus, as Hume put it, there is no real simplicity of experience at one time nor real identity across time that we could label the self. At the material level one finds distributed processes of experience, but nothing resembling an extended unified self. A good example of how this neuroscientific view has recently been translated into a philosophical conception of the self can be found in Galen Strawson's account of the self.⁷

He develops a model of what he calls a "minimal" self, for which he presents both introspective and neurological evidence. The minimal self lasts for a short period of time, approximating to 3 seconds, very much in line with processing times established in neuropsychological testing.⁸

Yet, if what there is in any real (material) sense is a neurological matrix that generates only a minimal self, then it is incumbent on these authors, as it was for Hume, to explain how it is that we nonetheless have a sense of a continuous self over time. Is the sense of self-continuity in some way carried by a succession of synchronic or momentary selves that are tied together by neural connections? Or is the continuous self a mere fiction, and if so, what sort of fiction is it? Or should we go in a different direction and argue that momentary minimal selves are 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.

⁵ David Hume, *A Treatise of Human Nature* (Oxford: Clarendon Press, 1975).

⁶ See, for example, Antonio Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (New York: Harcourt Brace and Co., 1999); Michael Gazzaniga, *The Mind's Past* (Berkeley, Calif.: University of California Press, 1998); Vilayanur S. Ramachandran and William Hirstein, "Three Laws of Qualia: What Neurology Tells Us about the Biological Functions of Consciousness, Qualia and the Self," in *Models of the Self*, ed. Shaun Gallagher and Jonathan Shear (Exeter: Imprint Academic, 1999), 83-111.

⁷ Galen Strawson, "The Self," in *Models of the Self*, 1-24.

⁸ See, e.g., Ernst Pöppel, "Time Perception," in *Handbook of Sensory Physiology*, vol. 8, ed. Richard Held, Herschel W. Leibovitz, and Hans-Lukas Teuber (New York: Springer, 1978).

The narrative theory of self is one contemporary answer to this problem. Daniel Dennett offers a version of this theory that is consistent with recent neuroscience.⁹

Humans not only have hard-wired brains; they also have language, as well as learning capabilities and memory systems. Language, according to Dennett, facilitates virtual connections among perception, memory, and motor systems in the brain. 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 virtual selves. We extend our biological boundaries to encompass a life of meaningful experience.

Narrative practices generate the narrative self. For Dennett, however, the narrative self is not substantially real; it is an empty abstraction. Specifically, Dennett defines the 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 an individual self consists of the abstract and movable point where the various stories (of fiction or biography) told by the individual, coalesce (*Fig. 1*). Importantly, the self is an inevitable invention. We cannot prevent ourselves from ‘inventing’ our selves. We are hard-wired to become language users, and once we are caught up in the web of language and begin to spin 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.”¹⁰

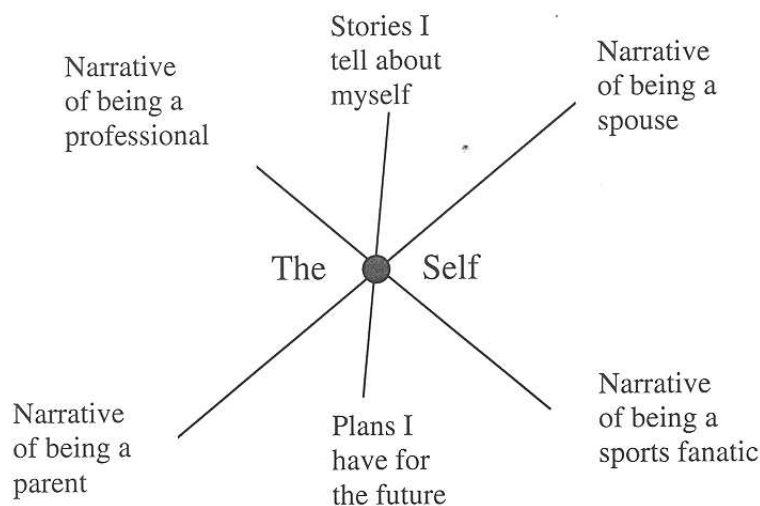


Figure 1: Dennett’s Center of Narrative Gravity

An alternative view of the narrative self is provided by Ricoeur.¹¹ Ricoeur carefully explores the nature of narrative and the constitution of the narrative self, and reaches conclusions that are not inconsistent with the picture painted by cognitive neuroscience. 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 sedimented and concrete. This enrichment is carried through his analysis of the relation between narrative identity, action, and moral relations. In addition, Ricoeur importantly insists that one’s own self-narrative is always entangled in the narratives of others.

⁹ Daniel Dennett, “Why Everyone is a Novelist,” *Times Literary Supplement* 4459 (Sept.16-22, 1988): 1016, 1028-1029.

¹⁰ Daniel Dennett, *Consciousness Explained* (New York: Little Brown and Company, 1991), 418.

¹¹ Ricoeur, *Oneself as Another*; and Paul Ricoeur, *Time and Narrative* (Chicago: University of Chicago Press, 1984).

Ricoeur's model allows for a productive tension between a unified life structure and certain instabilities found within the variety of perspectives that contribute to the narrative. In contrast to Dennett's notion of a movable abstract point, Ricoeur's account suggests that the self is the sum total of its narratives, and includes within itself all of the equivocations, contradictions, struggles and hidden attitudes that find expression in personal life. There is at least some place in this account to allow the extended self both a degree of concordance and the possibility of being decentered, distributed and multiplex (*Fig. 2*).

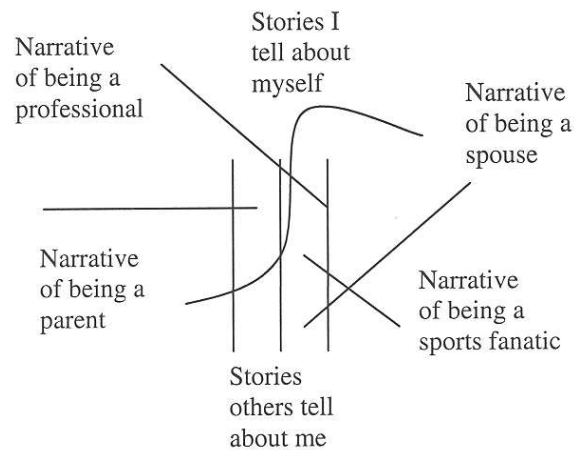


Figure 2: Ricoeur's Model of a Distributed Narrative Self

I have argued elsewhere¹² that Ricoeur's model of the narrative self, which involves a diachronic and complex structure that depends on reflective experience and on factors that are conceptual, emotional, and socially embedded, allows for conflict, moral indecision and self-deception, in a way that would be difficult to express in terms of Dennett's abstract center of narrative gravity.

In this chapter I propose to deepen Ricoeur's already rich account of the narrative self by exploring both internal and external conditions for self-narrative. By internal conditions I mean certain structural and functional aspects of self-narrative that reflect psychological and neuropsychological factors. By external conditions I mean those factors that depend upon embodied action and social relations. In addition, I want to show how these conditions are related to recent neurobiological accounts. In this regard the claim I want to defend is that an enriched notion of the narrative self, developed along the lines indicated by Ricoeur, remains consistent with a neuroscientifically informed materialist account, but in a way that is neither reductionistic, nor deflationary, nor internalistic.

Internal Conditions for Self-Narrative

Normal generation of a self-narrative depends on the proper functioning of a variety of cognitive capacities that reflect the internal conditions for self-narrative. These include capacities for short-term temporal processing (working memory), self-awareness, episodic memory, and reflective metacognition. I will discuss each of these in turn. Let me begin, however, by suggesting that self-narrative, in contrast to the narratives of others, has a certain primacy. Although my own self-narrative is greatly influenced by what others say

¹² Shaun Gallagher, "Philosophical Conceptions of the Self: Implications for Cognitive Science," *Trends in Cognitive Sciences* 4 (2000): 14-21.

about me, and is more generally constrained by the kinds of things that *can* be said, and that *are* said about persons in my culture, it has, from a first-person perspective, a priority in shaping my self-identity. What someone else says about me will have an effect on my self-identity, and will *matter*, only if it is something that I can recognize as applying to me, and only to the extent that it fits, positively or negatively, into my own self-narrative. Furthermore, my relations with others will take shape only within a framework that is defined, to some degree, by my own self-narrative. Thus, how I understand myself, my actions, and other persons will depend on the formation of my self-narrative.

There are at least four internal factors that condition the formation of self-narratives. They may be described in terms of the following cognitive or psychological capacities.

- (1) A capacity for temporal integration of information;
- (2) A capacity for minimal self-reference;
- (3) A capacity for encoding and retrieving memories episodically;
- (4) A capacity for engaging in reflective metacognition.

These capacities reflect basic structural features of self-narrative. If, in any particular instance, any one of these conditions becomes problematic, we would expect that failure to be expressed in the subject's self-narrative, and to represent a pathology in the subject's psychological capacities.¹³ In the following sections we examine each of these capacities in turn, exploring questions that ultimately throw light on the nature of the narrative self.

Temporal Integration

In general, narrative involves a twofold temporal structure. First, there is always a time frame that is internal to the narrative itself, a serial order in which one event follows another. It is this internal time frame that allows for the composition of narrative structure. Although in some way each event in the narrative is something new and different (and this is related to what Ricoeur calls “discordance”), in another way each event is part of a series (constituting a “concordance”), determined by what came before and constraining what is to come, allowing the story to advance.¹⁴ Such configuration is what makes possible the basic structure of narrative, the plot. There is no such thing as a synchronic plot; plot always depends upon an extended time frame.

In contrast to this internal time structure, there is by necessity an external temporality that defines the self-narrator's temporal relation to the events of the narrative. Even if this relation is left unspecified (“Once upon a time . . .”) it is usually open to a specification that these events happened in the past, or that they have not yet happened but will happen in the future, relative to the narrator's present. Although it is possible that if the events never happened and never will happen (as in the case of fictional events) then they do not have a specifiable place in time relative to the narrator, this cannot be the case with respect to self-narrative. Even if the event in question never did happen (for example, an event falsely remembered) or never will happen (for example, a planned event that never comes to fruition), in self-narrative it is still set in a temporal relation to the narrator.

¹³ Good examples of such failures are to be found in schizophrenia. Indeed, there is convincing evidence that for some schizophrenic subjects, more than one, and possibly all of these mechanisms fail. See Shaun Gallagher, “Self-Narrative in Schizophrenia,” in *The Self and Schizophrenia: A Neuropsychological Perspective*, ed. Anthony S. David and Tilo Kircher (Cambridge: Cambridge University Press, 2003 -- in press).

¹⁴ Ricoeur, *Oneself as Another*, 141ff; Ricoeur, *Time and Narrative*.

The external time frame is perspectival in a way that the internal time frame is not. The external time frame is defined relative to the narrator who exists in the present. From the perspective of the narrator telling the story, the events of the story, which internally maintain their serial relations (A happens before B), may be more or less remotely past (or future). In self-narrative it is always the case that the narrator is related to the events in the narrative in a perspectival way. The events recounted in the narrative are part of the narrator's immediate or remote past, or are projected to be part of the narrator's future, or are happening to the narrator in the present. Those who are incapable of maintaining a normal perspectival frame of reference (for example, very young children, and subjects who suffer from specific kinds of pathologies) are unable to generate a self-narrative properly specified in its temporal framework.

Capabilities related to temporal integration and the linear ordering of events within a temporal framework are essential to the formation of the narrative perspective and to the sequential order that characterizes narrative. With respect to what I have termed the internal time perspective, the proper functioning of working memory¹⁵ is a necessary condition for capabilities that are directly relevant to the formation of self-narrative. At least in part, this condition for self-narrative can be cashed out neurologically. Temporal integration and self-temporalization are linked to frontal brain structures.¹⁶ The sequencing of events in proper order is disrupted after frontal lobe lesions, especially in damage to the left frontal lobe.¹⁷ Integration of experienced content in the time domain, the temporal integration of sensory information (intermodal binding) in behavioral and linguistic sequences, and the proper functioning of working memory depend on brain activity in the prefrontal cortex.¹⁸ The external time frame, in contrast, is closely linked with the neurological underpinnings of episodic memory and metacognition (see below).

Minimal Self-Reference

Self-narrative is possible only if one is able to refer to oneself by using the first-person pronoun. To use the first-person pronoun, however, depends, developmentally, on a primitive sense of differentiation between self and non-self. Experiments on neonate imitation clearly show that the sense of this differentiation is present in the human infant from birth.¹⁹ This sense of differentiation is basic to a minimal self -- a self that is accessible

¹⁵ Working memory is a familiar concept in psychology and cognitive science. Those more familiar with phenomenological discussions can think of Husserl's analysis of the retentional-protentional structure of time-consciousness without loss of meaning.

¹⁶ Ernst Pöppel, "A Hierarchical Model of Temporal Perception," *Trends in Cognitive Sciences*, 1 (1997): 56-61; Nicole von Steinbüchel, Marc Wittmann and Ernst Pöppel, "Timing in Perceptual and Motor Tasks after Disturbances to the Brain," in *Time, Internal Clocks, and Movement*, ed. Maria A. Pastor and Julio Artieda (London: Elsevier, 1996), 281-303.

¹⁷ Pöppel, "Time Perception"; Brenda Milner, "Hemispheric Specialization: Scope and Limits," in *The Neurosciences: Third Study Program*, ed. Francis O. Schmitt and Frederic G. Worden (Cambridge, Mass: MIT Press, 1974), 75-89.

¹⁸ Kai Vogeley, Martin Kurthen, Peter Falkai and Wolfgang Maier, "The Human Self Construct and Prefrontal Cortex," in *Schizophrenia* (The Association for the Scientific Study of Consciousness, Electronic Seminar, 1999, at <http://www.phil.vt.edu/assc/esem.html>); Joaquim M. Fuster, "Commentary on 'The Human Self Construct and Prefrontal Cortex,'" in *Ibid.*; Joaquim M. Fuster, Mark Bodner, and James K. Kroger, "Cross-modal and Cross-temporal Association in Neurons of Frontal Cortex," *Nature* 405 (2000): 347-351; Patricia S. Goldman-Rakic and Lynn D. Selemon, "Functional and Anatomical Aspects of Prefrontal Pathology in Schizophrenia," *Schizophrenia Bulletin* 23 (1997): 437-458.

¹⁹ Andrew N. Meltzoff and Keith Moore, "Imitation of Facial and Manual Gestures by Human Neonates," *Science* 198 (1977): 75-78; Shaun Gallagher and Andrew N. Meltzoff, "The Earliest Sense of Self and Others: Merleau-Ponty and Recent Developmental Studies," *Philosophical Psychology* 9 (1996): 213-236.

to immediate and present self-consciousness. It is this minimal self which is then extended and enhanced in the self-narrative. Without the basic sense of differentiation between self and non-self, I would not be able to refer to myself with any specification, and self-narrative would have no starting-point.

The precise way that one is able to use the first-person pronoun provides a secure point of reference for creating a self-narrative. Indeed, it has been argued that I cannot be mistaken in my use of the first-person pronoun “as subject” (that is, when I refer to myself as the subject of experience), and that such uses of the first-person pronoun in self-reference are immune to error through misidentification.²⁰ For example, if I say, “I think Laura was here yesterday,” I may be entirely wrong about it being Laura (it may have been Julia), but I cannot be wrong about the I. I cannot say ‘I’ and mean to identify someone else by that word. If I say, “I see that Elaine is at her desk,” I can be wrong about it being a desk; I can be wrong about it being Elaine; and I can even be wrong about my cognitive act (it may be hallucination rather than visual perception). It would be nonsensical, however, to ask me “Are you sure that *you* are the one who sees that Elaine is at her desk?”

Importantly, even in cases where I misidentify myself “as object” (e.g., if, having released my arrow simultaneously with Elaine's arrow, I mistakenly claim that I am the one who hit the target, when in fact it is Elaine who hit the target), I cannot be mistaken in my self-reference.²¹ When I say ‘I,’ I necessarily refer to myself, even though I may be wrong about who hit the target. Indeed, we can consider this to be a case of misidentification only because I have correctly self-referred. The use of the first-person pronoun, which necessarily constitutes reference to the minimal self, then, is an extremely secure anchor for self-narratives.

A variety of candidate processes have been proposed as the neurobiological basis for our ability to self-refer, but there is as yet no consensus among neuroscientists on this score. What is the self to which I refer when I say “I”? Damasio suggests that “I,” as a minimal or core phenomenon, refers to a “neural self,” which is an integrated representation of one's bodily states (a body image), generated in complex networks of integrated neural connections, with some basic role played by the reticular formation.²² This “core self” is distinguished from what Damasio calls the autobiographical self, which is the self that is filled out by narrative. But the beginning point for a narrative must be the core or minimal self.

An alternative model is suggested by Ramachandran.²³ Even if the neural matrix that constitutes the body image plays some role in constituting our sense of self, Ramachandran suggests that what we experience as our self must include some active or executive component that is generated in the limbic system. “This would be a process involved in connecting motivation and emotion with the choice of action to perform (. . .) very much the sort of thing which the self was traditionally supposed to do.”²⁴ This set of control processes in the brain would be sufficient to generate a sense of the self that could anchor the use of the first person pronoun. For Ramachandran, however, that's all there is to the self, and as such, it is relatively inconstant. This sense of self comes and goes (not unlike

²⁰ Ludwig Wittgenstein, *The Blue and Brown Books* (Oxford: Basil Blackwell, 1958); Sidney Shoemaker, *Identity, Cause, and Mind* (Cambridge: Cambridge University Press, 1984).

²¹ Peter F. Strawson, “The First Person and Others,” in *Self-Knowledge*, ed. Quassim Cassam (Oxford: Oxford University Press, 1994), 210-215.

²² Damasio (*The Feeling of What Happens*, 274) gives the reticular formation an important role in generating this neural self, and specifically in “managing body states and representing current body states.”

²³ Ramachandran and Hirstein, “Three Laws of Qualia.”

²⁴ *Ibid.*, 109.

Strawson's minimal self) as behavior requires, or as the context requires us to use the first-person pronoun. In this regard, the traditional notion of self as something real and substantial is an illusion. Quite literally, the self is an inconstant illusion, generated by certain brain processes, to serve specific behavioral control functions.

More complex pictures of a self system in the brain are proposed by others. Panksepp, for example, views the self to be more constant and primitive, generated in subcortical sensory-motor and emotional circuits.²⁵ Regardless of the theoretical diversity, one thing we can note in these accounts is a certain tension between a reductionist-internalist perspective (the self is nothing more than a certain experience generated in the brain) and indications that point in the direction of embodied action. The self is tied to representations of bodily movement, embodied feeling, or executive processes that control responses to environmental demands. These issues are quite complex and we shall return to this point below.

Episodic and Autobiographical Memory

Both the capacity for temporal integration and the capacity for minimal self-reference seem necessary for the proper working of episodic and autobiographical memory. Indeed, these capacities are related to two aspects normally understood to define episodic-autobiographical memory: the recollection of the specific past time when an event occurred, and self-attribution, the specification that the past event involved the person who is remembering it.

There is a long philosophical tradition, starting with Locke, which holds that just such memories form the basis of personal identity.²⁶ It seems clear that narrative identity is primarily constituted in narratives that recount past autobiographical events. If there is any degree of unity to my life, as Ricoeur suggests there is, it is the product of an interpretation of my past actions and of events in the past that happened to me (all of which constitute my life history). If I am unable to form memories of my life history, or am unable to access such memories, then I have nothing to interpret, nothing to narrate that would be sufficient for the formation of self-identity.

The importance of episodic memory for the construction of narrative is also recognized in neuropsychology. Maguire et al. point out that the successful use of stories to convey and acquire information depends on two factors: that the story makes sense, and that the person who hears the story has access to prior knowledge.²⁷ The coherence of the story depends on these factors. In the construction of self-narrative, episodic-autobiographical memory provides the prior knowledge out of which the coherent narrative is formed. Vogeley et al. suggest that the encoding and retrieval of narrative just is the encoding and retrieval of episodic or autobiographical memory, and that such memory provides the contextual environment of former experiences.²⁸

It is not uncommon to find philosophical and neuropsychological discussions of memory framed in terms of 'encoding' and 'retrieval,' under the conception that memory is a form of information storage. There is an alternative model, however, that demonstrates

²⁵ Jaak Panksepp, "The Periconscious Substrates of Consciousness: Affective States and the Evolutionary Origins of the Self," in *Models of the Self*, ed. Shaun Gallagher and Jonathan Shear (Exeter: Imprint Academic, 1999), 113-130.

²⁶ John Locke, *An Essay Concerning Human Understanding*, 2d ed., ed. Alexander C. Fraser (New York: Dover, 1959).

²⁷ Eleanor A. Maguire, Chris D. Frith and Robin G. Morris, "The Functional Neuroanatomy of Comprehension and Memory: The Importance of Prior Knowledge," *Brain* 122 (1999): 1839-1850.

²⁸ Vogeley et al., "The Human Self Construct and Prefrontal Cortex."

a consistency between hermeneutical accounts and those of neuropsychology. On this model, memory is not simply a matter of reproduction or retrieval of stored information, but a reconstructive process.²⁹ In this sense the narrative (and self-narrative) process is not simply something that depends on the proper functioning of episodic (and autobiographical) memory, but in fact contributes to the functioning of that memory. Just to the extent that the current contextual and semantic requirements of narrative construction motivate the recollection of a certain event, that recollection will be shaped, interpreted and reconstructed in the light of those requirements. For this reason a variety of issues involved in a discussion of memory are also relevant to the discussion of metacognition (below).

On the basis of brain imaging studies, scientists have been able to define more precisely the complex network of brain structures involved in episodic memory,³⁰ although more generally, there is a growing consensus that almost all regions of the brain are involved in memory, and that episodic memories are distributed throughout the neocortex.³¹ For example, neuropsychological studies of brain-damaged subjects show that the hippocampus, the medial temporal cortex, and the prefrontal cortex play essential roles in the construction of episodic memory.³² Damage to the medial temporal cortex and hippocampus impairs the acquisition of new memories. The right prefrontal cortex is activated when there is an intention to retrieve episodic information and again when that retrieval is successful.³³ There is also activation of the right prefrontal cortex in recall involving specific personal qualities in contrast to contents with impersonal qualities.³⁴

²⁹ Craig R. Barclay and Peggy A. DeCooke, "Ordinary Memories: Some of the Things of which Selves are Made," in *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*, ed. Ulric Neisser and Eugene Winograd (Cambridge: Cambridge University Press, 1988); Shaun Gallagher, *The Inordinance of Time* (Evanston, Ill.: Northwestern University Press, 1998); Marya Schechtman, *The Constitution of Selves* (Ithaca, N.Y.: Cornell University Press, 1996).

³⁰ Paul C. Fletcher, Chris D. Frith and Michael D. Rugg, "The Functional Neuroanatomy of Episodic Memory," *Trends in Neuroscience* 20 (1997): 213-218.

³¹ Joaquin M. Fuster, "Network memory," *Trends in Neuroscience* 20 (1997): 451-458.

³² Hippocampal activation has been associated with the creation of associations during encoding of memory and with the experience of conscious recollection. See Daniel L. Schacter, Nathaniel M. Alpert, Cary R. Savage, Scott L. Rauch and Marilyn S. Albert, "Conscious Recollection and the Human Hippocampal Formation: Evidence from Positron Emission Tomography," *Proceedings of the National Academy of Sciences of the United States of America* 93 (1996): 321-325; Katrin Henke, Alfred Buck, Beno Weber and Heinz Gregor Wieser, "Human Hippocampus Establishes Associations in Memory," *Hippocampus* 7 (1997): 249-256. Imaging studies confirm that activation of the hippocampus correlates to the encoding of visuospatial and object memory. Retrieval and recognition of visual information involves the right hippocampus. Encoding for unfamiliar faces and spatial locations activates the left prefrontal cortex, and lesions of the prefrontal cortex affect the retrieval of contextual features of already formed memories. In addition, the network for episodic memory may involve a prefronto-thalamo-cerebellar network and retrieval may further involve the parietal cortex and the precuneus. See Nancy C. Andreasen, Daniel S. O'Leary, Ted Cizadlo, Stephan Arndt, Karim Rezai, Laura L. Boles Ponto, G. Leonard Watkins, and Richard D. Hichwa, "Schizophrenia and Cognitive Dysmetria: A Positron-Emission Tomography Study of Dysfunctional Prefrontal-Thalamic-Cerebellar Circuitry," *Proceedings of the National Academy of Sciences of the United States of America* 93 (1996): 9985-9990; Gereon R. Fink, Hans J. Markowitsch, Mechthild Reinkemeier, Thomas Bruckbauer, Joseph Kessler, and Wolfgang D. Heiss, "Cerebral Representation of One's Own Past: Neural Networks involved in Autobiographical Memory," *Journal of Neuroscience* 16 (1996): 4275-4282.

³³ Fletcher et al., "The Functional Neuroanatomy of Episodic Memory."

³⁴ Fink et al., "Cerebral Representation of One's Own Past."; Vogeley et al., "The Human Self Construct and Prefrontal Cortex."

Reflective metacognition enables the interpretational process that ordinarily shapes episodic memories into a narrative structure. An isolated life event is not usually meaningful in itself, but depends on a narrative structure that lends it context. To form a self-narrative, one needs to do more than simply remember life events. One must see in such events a significance that goes beyond the events themselves; to reflectively consider them, deliberate on their meaning, and decide how they fit together semantically. Metacognition allows for this reflective process of interpretation. It also enhances the product delivered by episodic memory. As Ricoeur points out, narrative identity “must be seen as an unstable mixture of [con]fabulation and actual experience.”³⁵ It is not unusual to construe certain events in a way that they did not in fact happen, for the sake of a unified or coherent meaning. To some degree, and for the sake of creating a coherency to life, it is normal to confabulate and to enhance one's story. Self-deception is not unusual; false memories are frequent. Whether metacognition contributes to this process, or limits it, it is clearly essential for the interpretive process that produces the self-narrative.

Neurological explanations of metacognition are more speculative than the other conditions considered above. Some of the speculation, however, is closely tied to experimental evidence. Based on experiments with split-brain patients, Gazzaniga suggests that a specific left-hemisphere mechanism, which he calls the 'interpreter,' is responsible for generating narratives, and that the interpreter compensates for missing or distorted information. He describes the generation of narrative as the mixing of fact and fiction by a left-hemisphere brain function he terms the 'interpreter.'³⁶ The interpreter weaves together autobiographical fact and inventive fiction to produce a personal narrative that enables the sense of a continuous self. Gazzaniga suggests 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.'³⁷ This also suggests that we cannot help but enhance our personal narratives with elements that smooth over discordances, discontinuities, and discrepancies in our self-constitution. The greater the discordances, discontinuities and discrepancies, the more our personal narratives involve deliberations, abstractions, withdrawals, and confabulations.

External Conditions for Self-Narrative

Several things should be clear from the above considerations.

- First, in regard to the notion of a self-narrative, and the constitution of a narrative self, there is some contributory role to be played by a minimal self.

³⁵ Ricoeur, *Oneself as Another*, 162.

³⁶ Gazzaniga, *The Mind's Past*; Michael Gazzaniga, “Consciousness and the Cerebral Hemispheres,” in *The Cognitive Neurosciences*, ed. Michael Gazzaniga (Cambridge, Mass: MIT Press, 1995), 1391-1400. In split-brain patients, the left hemisphere (LH) has no internal access to right-hemisphere (RH) experience because the corpus callosum has been severed. As a result, in properly designed experimental circumstances, the LH devises interpretations for actions and emotions produced by the RH. These interpretations, however, show consistency with the experiential context belonging to the LH rather than with the RH context. For example, the LH, ignorant of the content or cause of an emotion generated in the RH, generates an interpretation of the event in terms relevant to the content available to the LH, but unrelated to the proper RH context.

³⁷ Michael Gazzaniga and Shaun Gallagher, “The Neuronal Platonist,” *Journal of Consciousness Studies* 5 (1998): 713.

- Second, the various internal conditions that constrain the structure of self-narrative correlate to a set of cognitive capacities that account, in part, for the generation of the self-narrative.
- Third, these cognitive capacities are open to neurological explanation.
- Fourth, the neurological picture is extremely complex. Despite the high degree of uncertainty in regard to the precise neural mechanisms for these various capacities, it is clear that the narrative self is the product of a widely distributed set of functions involving many diverse brain areas and types of neural functioning.
- Finally, although there is a strong tendency in neuroscience to present a reductivist-internalist picture of the generation of the self, and to see the self as a fiction or illusion, a product reducible to brain processes, there are very important indications that this picture will not capture everything necessary for this generation.

I want to explore this last point further by following out some of the lines of tension found in the neurological accounts of the internal conditions for self-narrative. These lines move us in the direction of embodied action and intersubjective interaction, and suggest that the self is mapped out on a larger and more intricate scale than that drawn in purely neurobiological accounts. In turn, this larger map also suggests an even more elaborate neurobiological picture of how the self is generated.

Embodied Action

The minimal self, of which I have an immediate experience, and which I can express directly by using the first-person pronoun, has a structure that can be articulated into several elements. These elements indicate that even the minimal self is an embodied self, and is not reducible to a purely psychological phenomenon, or to a complex brain event. Furthermore, if the minimal self is something like the starting-point for constructing a narrative self, the latter depends upon embodied action for its development.

Developmental and behavioral studies suggest that the minimal self involves at least two distinguishable aspects: a differentiation between self and non-self, and a proprioceptive-kinesthetic sense of one's own body.³⁸ In the case of movement, the latter aspect may be specified further to include both a *sense of ownership* for movement (that is, a sense that I am moving) and a *sense of agency* (that is, a sense that I am the one causing the movement).³⁹ As indicated above, scientific studies of neonate imitation suggest that the differentiation between self and non-self is present from birth. Infants less than an hour old are capable of imitating the facial gestures presented by others, and are able to correct their own gestures to match the gestures presented. This implies that the newborn infant can discriminate its own proprioceptively felt body from the body of the other as it is presented in visual perception. Bermúdez, on the basis of just such developmental evidence, shows that pre-linguistic infants who have not yet acquired use of the first-person pronoun, nonetheless have a non-conceptual sense of themselves already framed in terms

³⁸ José Bermúdez, "The Moral Significance of Birth," *Ethics* 106 (1996): 378-403; Shaun Gallagher, "The Moral Significance of Primitive Self-Consciousness," *Ethics* 107 (1996): 129-140.

³⁹ For the distinction between sense of ownership and sense of agency, see Gallagher, "Philosophical Conceptions of the Self"; Shaun Gallagher, "Self-Reference and Schizophrenia: A Cognitive Model of Immunity to Error Through Misidentification," in *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-experience*, ed. Dan Zahavi (Amsterdam and Philadelphia: John Benjamins, 2000), 203-239; and Shaun Gallagher, "Sense of Agency and Higher-Order Cognition: Levels of Explanation for Schizophrenia," *Cognitive Semiotics* 2 (2003): in press.

of the bodily movements and actions of which they are capable.⁴⁰ The experience of bodily movement thus involves the sense of ownership -- the sense that *my* body is moving, or that it is *my* movement. This is present even in the case of involuntary action, as when someone pushes me from behind. In this case, however, there is no sense of agency, that is, the sense that I caused my movement, a sense that one does have in the case of intentional action. In the immediate phenomenology of intentional action, agency is not represented as separate from the action. It is rather felt as an intrinsic property of the action itself and is experienced as a perspectival and embodied source.⁴¹ Experimental studies of non-pathological subjects suggest that the sense of agency is based on pre-motor neurological processes that precede the action and that translate intention into movement, rather than on sensory feedback from movement itself or from peripheral effort associated with such movement.⁴²

Prior to any ability to use the first-person pronoun, the minimal self is already a complex structure involving non-conceptual self / non-self discrimination, proprioceptive-kinesthetic embodiment, and senses of ownership and agency for action. The first-person pronoun, which comes to anchor the self-narrative, has an embodied referent that is something more than a deflated pronoun, grammatical structure or piece of vocabulary. When we become capable of self-narrative (probably not before the age of four years when capabilities for episodic memory develop), the self that we each call 'I,' and that we begin to narrate, has always already been an embodied, enactive self with a central nervous system capable of instantiating just the right structural features for intentional action. Neither the narrating *I*, nor the narrated *I*, is a Cartesian or transcendental ego. Rather, it is an *I* embedded in a biological life, and in environmental contexts that call for action.

For the construction of self-narrative the sense of agency, and the ability to attribute action to oneself, are essential. Ricoeur, following Bremond, makes it clear that in narrative a character is either (or alternatively) an agent or a sufferer. In self-narrative the construal of the character (the self) as agent or sufferer depends on the ability of the narrator to self-attribute action, and such self-attribution is necessarily based on the sense of agency and ownership (in the case of agentive action) and the sense of ownership (in the case of the sufferer). This means that even if other aspects of the minimal self are intact, the person who lacks a sense of self-agency or self-ownership would be incapable of self-narrative.⁴³ Or more positively, as Ricoeur indicates, the “elementary sequences of a narrative already contain [the] correlation that allows us to ascribe action to an agent.”⁴⁴

⁴⁰ José Bermúdez, *The Paradox of Self-Consciousness* (Cambridge, Mass.: MIT Press, 1998).

⁴¹ Anthony J. Marcel, “The Sense of Agency: Awareness and Ownership of Actions and Intentions,” in *Agency and Self-Awareness*, ed. Johannes Roessler and Naomi Eilan (Oxford: Oxford University Press, 2003). Shaun Gallagher and Anthony J. Marcel, “The Self in Contextualized Action,” *Journal of Consciousness Studies* 6 (1999): 4-30.

⁴² Marcel, “The Sense of Agency.”; Pierre Fourneret and Marc Jeannerod, “Limited Conscious Monitoring of Motor Performance in Normal Subjects,” *Neuropsychologia* 36 (1998): 1133-1140. Other research correlates initial awareness of action with recordings of the lateralized readiness potential and with transcranial magnetic stimulation of the supplementary motor area. See Patrick Haggard and Martin Eimer, “On the Relation between Brain Potentials and the Awareness of Voluntary Movements,” *Experimental Brain Research* 126 (1999): 128; Patrick Haggard and Elena Magno, “Localising Awareness of Action With Transcranial Magnetic Stimulation,” *Experimental Brain Research* 127 (1999): 102.

⁴³ Just such disruptions in the neurological underpinnings of the sense of agency may be responsible for the problems that schizophrenics have with delusions of control and thought insertion, and that are reflected in their self-narratives. See Gallagher, “Self-Narrative and Schizophrenia.”

⁴⁴ Ricoeur, *Oneself as Another*, 144; Ricoeur cites Claude Bremond, *Logique du récit* (Paris, 1973).

What is already contained in the narrative sequences is the embodied sense of self-agency and self-ownership that makes my action the action of my self-narrative.

In addition, action is not something that happens in the abstract. It is precisely this realization that makes all accounts of personal identity that are purely in terms of psychological continuity or purely in terms of physical brain states abstract and artificial. Action happens in rich pragmatic and social contexts, and such contexts make up the substance of self-narratives. Moreover, there is good neurological evidence to suggest that socially contextualized action has a priority over actions that are describable in abstract or even pragmatic or instrumental terms.

Social Context

Self-narrative has the capacity to invent, construct, and extend the reality in which we live, and to carry us beyond the material elements that make up organism and environment, precisely because it is anchored in the real actions of people. As Ricoeur notes, the narrative structure conjoins the personal-level processes of action and character.⁴⁵ Accounts of the self that are cast in the sub-personal terms of brain states simply do not come up to the proper level of description. The person, as character in the narrative, is not an entity distinct from his or her contextualized experiences in which a concordant set of actions are often disrupted by discordant events in a dialectic that forms the plot of the narrative. If, as Ricoeur claims, it is the narrative that “constructs the identity of the character,”⁴⁶ this means also that it is the action and the suffering of the character in the real embodied contexts of life that shape her identity and her (moral) character. Importantly, the actions that matter for shaping the self and its character, are socially contextualized actions, since, as Ricoeur suggests, “action is interaction.”⁴⁷

The idea that action is interaction is reinforced by recent neuroscientific discoveries that verify the connection between our own intentional action and the action of others. Brain imaging experiments show that there are a number of brain areas (including the supplementary motor area, the dorsal premotor cortex, the supramarginal gyrus, and the superior parietal lobe) that are activated in common when a subject

- Engages in intentional action,
- Observes others engaging in such action,
- Consciously simulates (or imagines) performing such action, or
- Prepares to imitate such action.⁴⁸

This intriguing set of shared neuronal activations suggests that the overlap of brain areas may contribute to an explanation of how we come to understand others. That is, we

⁴⁵ Ibid., 146.

⁴⁶ Ibid., 147.

⁴⁷ Ibid., 145; for a sustained argument for the primacy of socially contextualized action, see Gallagher and Marcel, “The Self in Contextualized Action.”

⁴⁸ Sarah-Jayne Blakemore and Jean Decety, “From the Perception of Action to the Understanding of Intention,” *Nature Reviews: Neuroscience* 2 (2001): 561-567; Jean Decety, Thierry Chaminade, Julie Grèzes and Andrew N. Meltzoff, “A PET Exploration of the Neural Mechanisms Involved in Reciprocal Imitation,” *Neuroimage* 15 (2002): 265–272; Vittorio Gallese, “The Acting Subject: Towards the Neural Basis of Social Cognition,” in *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, ed. Thomas Metzinger (Cambridge, Mass.: MIT Press, 2000), 325-333; Julie Grèzes and Jean Decety, “Functional Anatomy of Execution, Mental Simulation, Observation, and Verb Generation of Actions: A Meta-Analysis,” *Human Brain Mapping*, 12 (2001): 1–19; Marc Jeannerod, *The Cognitive Neuroscience of Action* (Oxford: Blackwell Publishers, 1997).

activate parts of our own motor and cognitive systems in a simulative way, and this neural reverberation produces in us a sense of what the other person's experience must be like. Those areas that are non-overlapping are also of importance for explaining our ability to distinguish our own activities from those of others, and may contribute to a sense of self-agency.⁴⁹ It seems evident that activation of certain brain areas is important for explaining how we relate our own actions to the actions of others. Clearly, however, these brain processes refer to and depend upon something more than just more brain processes. That is, such brain processes would not occur unless we were acting within social contexts.

Actions that are elicited in social contexts have a certain primacy over purely cognitive, and purely instrumental activities. We can garner scientific evidence for this claim from several areas of research. First, on the basis of neuropsychological studies of patients recovering from brain damage, Anthony Marcel and I have distinguished between abstract, pragmatic, and socially contextualized actions.⁵⁰ Patients with motor impairments, for example, show significant improvement in various aspects of motor control and fluency in impaired behaviors when such actions are performed as pragmatically meaningful actions, in contrast to their performance when elicited as abstract behaviors. More significantly, in almost all of these cases further measurable improvement is found when nominally the same movements are performed in a socially contextualized situation, in which the movements constituted actions with personal and culturally derived signification.

For example, a woman who had difficulty in grasping, in lifting, and in motor fluency when asked in an experimental [abstract] situation to lift a cylinder of the weight and size of a glass of liquid and to move it toward or away from herself, showed clear improvement in her performance when spontaneously drinking during a meal. This same woman was even more proficient, almost normal, in the very same movements when serving cups of tea to guests in her home [a social context], although not when clearing up the cups [a pragmatic or instrumental context].⁵¹

These and other observations suggest that performance in socially contextualized situations is not reducible to performance in non-social, pragmatic or abstract contexts.

This view is reinforced by experiments conducted with a deafferented subject, IW. IW lost proprioception and the sense of touch below the neck, and as a result, has profound difficulties in controlling locomotive and instrumental movement.⁵² To reach and pick up a glass to take a drink, IW is required to keep his hands in his visual field, and to think through his movements. The tasks that are taken up by normal body schematic processes in the normal case of intentional action are transformed for IW into cognitive tasks. IW has to turn his body into an object in order to do the things that he needs to do in ordinary pragmatic situations. Yet, in spite of this difficulty with pragmatic or instrumental movement, IW is remarkably capable of normal gesticulation in the course of socially communicative practices with others. IW does not have to think about his hand gestures in order for them to take shape in the normal way. And even when his hands are out of

⁴⁹ Marc Jeannerod, "Neural Simulation of Action: A Unifying Mechanism for Motor Cognition," *Neuroimage* 14 (2001): 103-109.

⁵⁰ Gallagher and Marcel, "The Self in Contextualized Action."

⁵¹ *Ibid.*, 8.

⁵² Jonathan D. Cole, *Pride and a Daily Marathon* (Cambridge, Mass.: MIT Press, 1995); Shaun Gallagher and Jonathan D. Cole, "Body Schema and Body Image in a Deafferented Subject," *Journal of Mind and Behavior* 16 (1995): 369-390.

sight, he is capable of close to normal gesture.⁵³ This case, and others like it suggest that social and communicative interaction is not reducible to the cognitive or the pragmatic realm.⁵⁴

I want to suggest that the distinctions between cognitive, pragmatic, and social aspects of experience are important ones to consider, not only in regard to questions about the status of action, or with respect to how we understand one another, but in terms of constructing self-narratives. Even if normally these aspects of experience are fully integrated, still, how we understand ourselves as persons and in our interactions with one another is not reducible or fully explainable in purely pragmatic or purely cognitive terms. This realization not only supports the emphasis that Ricoeur places on the social aspects of self-narrative,⁵⁵ but, without reducing the significance of neurological processes, it suggests that the narrative self is irreducible to such processes.

Conclusion

The capacity for self-narrative and the construction of the narrative self clearly depend on a large number of complex brain processes and cognitive functions. The latter include capacities for temporal integration of information; for minimal self-reference; for encoding and retrieving memories episodically; and for engaging in reflective metacognition. The neurological processes that underpin these capacities involve massively complex networks that integrate functions of the prefrontal cortex, the reticular formation, the limbic system, and other sub-cortical areas, the hippocampus, the medial temporal cortex, parietal cortex, the precuneus, the supplementary motor area, the premotor cortex, the supramarginal gyrus, and various other brain areas. If neuronal processing in these areas allows for the proper functioning of memory systems, perceptual-motor systems, and behavioral control systems, and if these constitute necessary conditions for the narrative constitution of the self, all of them together are still insufficient to generate that self. Narrative happens at a personal level. We do not narrate brain processes or cognitive-behavioral systems. In self-narrative, the narrator-narrated depends on a body that is capable of action, and on embodied action that is socially contextualized. Indeed, one could argue on good evidence that the conditions of embodied and socially contextualized actions generate the brain processes described, and drive both phylogenetic and ontogenetic development of a brain that is capable of spinning narratives.

⁵³ Jonathan D. Cole, Shaun Gallagher and David McNeill, "Gesture Following Deafferentation: A Phenomenologically Informed Experimental Study," *Phenomenology and the Cognitive Sciences* 1 (1): 49-67.

⁵⁴ Shaun Gallagher, Jonathan D. Cole and David McNeill "Social Cognition and the Primacy of Movement Revisited," *Trends in Cognitive Sciences* 6 (2002): 155-156.

⁵⁵ Thus, Ricoeur, writes: "in our experience the life history of each of us is caught up in the histories of others. Whole sections of my life are part of the life history of others." Ricoeur, *Oneself as Another*, 161.