

# Narrators and Comparators: The Architecture of Agentive Self-Awareness

Tim Bayne  
University of Oxford and St. Catherine's College  
Manor Road  
Oxford OX1 3UJ  
United Kingdom  
tim.bayne@gmail.com

Elisabeth Pacherie  
Institut Jean Nicod UMR 8129  
1 bis, avenue de Lowendal  
75007 Paris  
France  
pacherie@ens.fr

This is a **DRAFT** of a paper that is forthcoming in *Synthese*. It is **NOT** for quotation. Comments are most welcome.

---

## 1. Introduction

Until recently, neither philosophers nor psychologists had much interest in the awareness of one's own agency. This is no longer the case, and there is now a burgeoning literature on the mechanisms underlying 'the sense of agency'—what we will call agentive self-awareness. Two central strands can be discerned in the literature. Some argue that agentive self-awareness involves the deployment of a holistic, domain-general, central systems mechanism—a narrative module. Other theorists argue that agentive self-awareness involves the operation of atomistic, domain-specific, comparator systems. We argue that neither of these two approaches is entirely satisfactory, and that a full account of agentive self-awareness needs to draw on the resources of both narrator-based and comparator-based approaches to agentive self-awareness.

## 2. Agentive self-awareness

Despite the recent explosion of interest in the topic, the complexity of agentive self-awareness remains under-appreciated. The aim of this section is not to present a complete taxonomy of agentive self-awareness, but to sketch those distinctions that have an important bearing on the architecture of the topic (see Horgan et al 2003; Bayne 2007; Pacherie 2007).

A first dimension along which states of agentive self-awareness can be located concerns their *representational contents*. The contents of such states can range from the 'thin' to the 'thick'. At the thin end of the spectrum, one can experience oneself as the mere author of an event. One can be aware of the movement of one's leg as constituting an action of one's own as opposed to the action of someone else or a mere event. At the thicker end of the spectrum, agentive self-awareness can include not merely the representation of a movement as one's own action, but also a representation of the kind of action it is and one's reasons for performing it. One can be aware of oneself as opening a door, and as opening a door in order to (say) leave a building (as opposed to showing someone how to open the door). As we will see, it is very much an open question just how the thin and thick components of agentive self-awareness are related to each other.

A second dimension along which states of agentive self-awareness can be located concerns the notion of awareness in question. Some states of agentive self-awareness take a judgmental (or doxastic) form. Normally, one's awareness of the kinds of actions that one is performing—opening a door, shaking a colleague's hand, making a cup of coffee—takes the form of belief. Some theorists appear to take agentive self-awareness to be exclusively judgmental in form.<sup>1</sup> This is not our view, however. We hold that the 'vehicles' of agentive self-awareness are often more primitive than judgments. Think of what it is like to push a door open. One might judge that one is the agent of this action, but this judgment is not the only way in which one's own agency is manifested to oneself; indeed, it is arguably not even the primary way in which one's own agency is manifested to oneself. Instead, one *experiences* oneself as the agent of this action. Such states are no more judgments than are visual experiences of the scene in front of one or proprioceptive experiences of the current position of one's limbs. We will reserve the term 'experiences of agency' for this kind of awareness of one's own agency. We use the term 'agentive

---

<sup>1</sup> Note that restricting agentive self-awareness to judgment doesn't settle the question of whether there is a phenomenology of first-person agency, for it is very much an open question whether judgements might have phenomenal character. Some theorists think that they do; others argue that they don't.

awareness' as a generic term that subsumes both agentic experience and agentic judgment.

Why posit agentic experiences in addition to agentic judgments? Consider a well-known case from William James's *Principles of Psychology*. A patient with an anesthetized arm is asked to raise it. The patient's eyes are closed, and unbeknown to him his arm is prevented from moving. Upon opening his eyes, the patient is surprised to discover that his arm has not moved, for he had experienced himself as moving his arm. He now judges that actually he has not moved his arm. Suppose the patient was asked to close his eyes again and raise his arm. Although the patient will judge that he's not moving his arm, he may nonetheless experience himself as moving his arm. Simply judging that one does not move doesn't make the experience go away.

Although some treatments of agentic experience conceive of it in non-representational terms—as involving 'raw feels' that are merely associated with actions—this is not our view. As we conceive of it, agentic experiences have *representational* content. Just as one's proprioceptive experience can represent one's arms as extended parallel to each other, so too one's agentic experience can represent oneself as moving one's arms. And just as proprioceptive experiences can be misleading (illusory), so too agentic experiences can be misleading (illusory).

It is very much an open question how rich the contents of agentic experience can be. A strong case can be made for thinking that the contents of agentic experience can go beyond merely representing oneself as the agent of an event; instead, it might also be possible for information about the degree of control one has over the movement and the degree to which the action is effortful, for example, to be encoded in the contents of agentic experience. But it is less clear whether information about the intentional content of an action, or information about one's aims in performing the action, can also be encoded in agentic experience. Arguably, one can be conscious of this kind of information only by forming beliefs about what it is that one is doing (or trying to do) and why.

To distinguish agentic experiences from agentic judgments is not to deny that they are related. For one thing, agentic judgments are typically grounded in and justified by agentic experiences. In the normal case, we judge that we are the agent of a particular movement on the grounds that we enjoy an agentic experience with respect to it; and, correlatively, we normally judge that we are not the agent of a particular event because we lack an agentic experience with respect to it. Of course, our agentic judgments are not beholden to our agentic experiences. Not only can we deny that we are the authors

of events towards which we have an agentic experience, we can also assert that we are the authors of events for which we lack such an experience. But, arguably, such cases are the exception rather than the rule.

So, to recap. A number of different states can be subsumed under the phrase 'agentic self-awareness'. We suggest that two distinctions are particularly helpful in imposing some order on what threatens to be a welter of confusion. Firstly, there is a distinction between states with relatively thin contents and states with relatively thick contents: in the thin sense, one might be aware of a movement merely as one's own action, in the thick sense, one might be aware of one's movement as realizing a certain intention. Secondly, there are two types of agentic representations: agentic experiences and agentic judgments. A *full* account of agentic self-awareness must explain both how agentic experiences and agentic judgments are generated, and how they are related.<sup>2</sup>

### **3. Two approaches to agentic self-awareness**

The literature on agentic awareness contains two general approaches: a holistic narrator-based approach, and an atomistic comparator-based approach. Let us examine these two approaches in detail.

According to the holistic approach, agentic self-awareness is subserved by a holistic mechanism that is concerned with narrative self-understanding. Our sense of what, if anything, we are up to, is a result of a high-level integrative process that draws on the agent's self-conception and tries to put the best spin on things that it can. We turn Dennett's intentional stance inwards, and treat ourselves as entities whose behaviour needs to be made sense of in light of an implicit theory of ideal agency. As such, there is every chance that agentic self-awareness will contain a large number of half-truths or even outright errors.

Many authors have expressed some sympathy with, and in some cases whole-hearted commitment to, the narrative approach. Interpreting split-brain studies in light of Dennettian (1992) themes concerning the role of narrative in self-interpretation, Roser and Gazzaniga (2004, 2006) have argued that the left hemisphere contains an interpreter, whose job it is to make sense of the agent's own behaviour. Peter Carruthers takes a similar position, claiming that "our awareness of our own will results from turning our

---

<sup>2</sup> We hasten to add that a completely comprehensive account of agentic self-awareness would also account for the awareness that we enjoy of our own *mental* agency. We will not speculate about how mental agency might figure in the account that we develop, but it is an important issue that deserves extended treatment in its own right.

mind-reading capacities upon themselves, and coming up with the best interpretation of the information that is available to it—where this information doesn't include those acts of deciding themselves, but only the causes and effects of those events." (2007: p. 3 in the draft). The psychiatrist Louis Sass has suggested that schizophrenic patients with delusions of alien control no longer feel as though they are in control of their actions because "particular thoughts and actions may not make sense in relation to the whole" (1992: 214). Developing Sass's proposal, Stephens and Graham suggest that a "subject's sense of agency regarding episodes in her psychological history might depend on her ability to integrate them into her larger picture of herself" (Stephens and Graham 2000: 161). Holistic themes also play an important role in Daniel Wegner's influential treatment of agentive self-awareness:

The fact is, each of us acts in response to an unwieldy assortment of mental events, only a few of which may be easily brought to mind and understood as conscious intentions that cause our action. We may find ourselves at some point in life taking off a show and throwing it out a window, or at another point being sickening polite to someone we detest. At these junctures, we may ask ourselves, What am I doing? Or perhaps sound no alarms at all and instead putter blithely along assuming that we must have meant to do this for some reason. We perform many unintended behaviours that then require some artful interpretation to fit them into our view of ourselves as conscious agents. Even when we didn't know what we were doing in advance, we may trust our theory that we consciously will our actions and so find ourselves forced to imagine or confabulate memories of "prior" consistent thoughts. (Wegner 2002: 145f.)

Wegner does not say much about where in cognitive architecture we might find the agent's theory of apparent mental causation, but the holistic nature of this model suggests that it must be centrally located.<sup>3</sup>

---

<sup>3</sup> In the second chapter of *The Illusion of Conscious Will*, Wegner suggests that there are "multiple sources" for agentive self-awareness (what he calls "the conscious will"): a person can derive the feeling of doing from "conscious thoughts about what will be done, from feedback from muscles that have carried out the action, and even from visual perception of the action in the absence of such thoughts or feedback" (2002: 49). So it is clear that Wegner's view is not committed to any straightforward version of the narrative approach. Nonetheless, given the holistic nature of Wegner's theory of apparent mental causation, it is difficult to see how, on his account, the mechanisms responsible for the feeling of doing could be anything but centrally located.

Although the narrative approach permeates much of the recent literature on agentic self-awareness they do not enjoy a monopoly. Jostling alongside the narrative approach, and even inter-mingled with it, is a conception of agentic self-awareness as generated by low-level, distributed mechanisms that have more to do with motor control than narrative self-understanding. We will refer to this as the comparator-based approach, on the grounds that the most influential versions of the view appeal to the role of comparators involved in forward models of action control in accounting for agentic self-awareness (Frith et al., 2000a, 2000b; Blakemore and Frith, 2003). According to this approach, the predictors or forward models are fed a copy of the motor commands and compute estimates of the sensory consequences of the ensuing movements. These predictions can be used in several ways. First, awareness of initiating a movement depends on awareness of the predicted sensory consequences of the movement. Second, the predictions can also distinguish the sensory consequences of self-generated movement from those due to other causes. The job of the comparators is to compare the predicted sensory consequences of the movement with sensory feedback. When there is a match between predicted and actual state, the comparator sends a signal to the effect that the sensory changes are self-generated, and when there is no match (or an insufficiently robust match), sensory changes are coded as externally caused. Third, these predictions can also be used to filter sensory information and to attenuate the component that is due to self-movement. Crucial to the comparator approach is the notion that agentic awareness can be generated by mechanisms that need not—and typically will not—have access to fully-fledged intentions. There is no need for a centralized narrator with access to high-level representations to get into the action.

#### **4. Evaluating the debate between the holistic and narrative approaches**

What should we make of the debate between these two approaches to agentic awareness? We might begin by asking what exactly these models are models of. Unfortunately, the answer to this question is not always clear. Although we have distinguished agentic experiences from agentic judgments, and agentic states with thin contents from agentic states with thick contents, these distinctions are not commonly made, and it is often unclear exactly what phenomena proponents of the above models are attempting to analyse.

How *might* these models most plausibly be taken? On the face of things, the holistic approach would appear to be better suited to accounting for agentic *judgments*. Holistic approaches have some plausibility when it comes to first-person mind-reading of mental states that have a dispositional character. Arguably, we come to think of ourselves as kind, friendly, grouchy, honest, and so on by taking a third-person perspective on our

behaviour. And, perhaps, the self-ascription of propositional attitudes—understood as dispositions to behave in certain ways—can also be accounted for by appeal to the operations of a mind-reading system that operates according to holistic principles.

But it is less plausible to suppose that agentic *experiences* might depend on the operations of a holistic mechanism of narrative self-understanding. In general, experiential states—sensations, affective states, and perceptions—are generated by modular systems that are (reasonably) encapsulated from information represented in the central system. Talk of background principles of rationality or narrative consistency appears to be out of place here.<sup>4</sup>

With this in mind, we approach the debate between the narrative-based and comparator-based approaches with the following tentative assumption: narrative-based approaches are likely to prove most viable with respect to agentic judgments, whereas comparator-based approaches are likely to prove most viable with respect to agentic experiences. Let us turn now to see whether this assumption might be vindicated.

#### **4.1 Limitations to the holistic approach**

Proponents of the narrative approach have marshalled a wide array of evidence in support of their position. They have appealed to developmental evidence. When young children happen to achieve a goal by luck, they will say that they had intended the action that yielded that goal all along (Phillips et al., 1998). Proponents of the narrative approach have also appealed to studies involving patients with brain damage. Patients with anosognosia for hemiplegia say that they are currently raising their arm when, in fact, their arm has not moved. When it is pointed out to the patient that his arm has not moved, he may confabulate an excuse for his inertia (Feinberg et al., 1998, 2000). Split-brain subjects are prone to confabulate accounts of actions that are generated by their right hemisphere (Gazzaniga and LeDoux 1978: 148). Data from subjects in altered states of consciousness has also been taken to support the narrative approach. For example, bizarre behaviours performed in response to hypnotic suggestion are often accompanied by elaborate rationalizations and confabulation on the part of the agent (Moll 1889: 153-4). Finally, the narrative approach derives support from a number of laboratory studies with normal subjects, in which it has been shown that agentic judgments can be modulated by

---

<sup>4</sup> It comes as no surprise that even dyed-in-the-wool theory theorists, such as Gopnik (1993), appear to take a non-holistic approach to the self-ascription of sensations. Gopnik argues that sensations are self-ascribed on the basis of their 'Cartesian buzz'. She doesn't say much about what she takes the Cartesian buzz to be, but it is clear that she doesn't think of it in holistic terms.

priming and various contextual parameters (Aarts et al 2005; Metcalfe & Greene; in press; Wegner & Wheatley 1999; Wegner et al 2004). In one of the most striking experiments of this kind (Johansson et al 2005), normal subjects—that is, subjects without working memory problems—had to make choices between pairs of faces on the basis of attractiveness. On some trials, immediately after their choice they were presented with the picture they had chosen and asked to state the reasons for their choice. Unknown to them, on certain these trials the picture they had chosen was surreptitiously replaced with the other picture. Most participants failed to notice the mismatch between the face they had intended to choose and the face they were presented with and readily offered reasons for why they had chosen that face!

It is clear that these studies reveal something about the structure of agentic self-awareness, but it is not at all clear just what they reveal. Care must be taken in interpreting these studies. Although they do show that agentic experience is ‘inferential’, they do not show that these mechanisms are interpretive in the sense presupposed by the narrative account. In other words, not all of these studies show that the interpretive constraints on agentic self-awareness derive from the agent’s overall self-conception. Merely showing that agentic experience involves interpretive elements does not show that it is subject to narrative—or even doxastic—penetration. After all, perceptual experience is interpretive without being doxastically penetrable. It may be true that the mechanisms of agentic experience obey the general Humean principles of priority, consistency and exclusivity, but it doesn’t follow that they are central and holistic. Local mechanisms can exploit general interpretive principles.

Secondly, we can note that the disruption of agentic self-awareness seems not to be accompanied by any obvious disruption to the agent’s narrative self-conception. As best we know, those whose narrative moorings are cut adrift—such as patients with massive retrograde amnesia—do not display any signs of disruption to their sense of agency. And, conversely, those who have suffered some basic disruption to their agentic self-awareness do not display any obvious signs of disruption to their narrative self-conception. (Those who suffer from schizophrenic delusions of alien control and thought insertion might be a partial exception here.)

Thirdly, impulsive, routine and automatic actions are not preceded by conscious intentions, nor do we usually retrodict conscious intentions for them once performed, yet they are accompanied by a sense of agency (Marcel, 2003; Pacherie, 2001; Spence, 2001). Consider the contrast between the experience of blinking and the experience of feeling one’s eyes twitch. Blinking is experienced as an action, as something that one does, whereas muscular spasms are experienced as things that happen to one. But it is highly

implausible to suppose that one is aware of intentions to blink before one blinks, or that the difference between the experience of blinking and that of muscular spasms might be the result of whether or not one can locate the physical movement in a narrative framework and thus retrodict an intention to blink. We might also note that it seems plausible to suppose that non-linguistic animals—whom we can assume lack a narrative self-conception—might enjoy primitive forms of agentic experience. (We are reminded here of a video of a dog with what appeared to be an anarchic paw, which was recently posted on YouTube.)

A fourth challenge to the narrative approach concerns its ability to explain the contrast between syndromes in which behaviour is similar but agentic awareness is dissimilar. One such contrast holds between the Anarchic Hand Syndrome (AHS) (Marchetti & Della Salla, 1998; Della Sala et al, 1991; Goldberg & Bloom, 1990) and Utilization Behaviour (UB) (Lhermitte 1983, 1986; Archibald et al 2001; Estlinger et al, 1991). The two syndromes have much in common from a behavioural perspective. In both syndromes, the patient engages in stimulus-driven behaviour that is, often, is inappropriate given the agent's social context. For instance, a patient with UB may pick up a pair of glasses and put it on, in response to the glasses being placed in front of him, if a second and then a third pair of glasses are placed in front of him, he will put them on and will end up wearing all three. Similarly, a patient with AHS may pick food from his neighbour's plate with his anarchic hand. Yet the responses of patients to this stimulus-driven agency is very different; indeed, at a behavioural level it distinguishes the two syndromes. Whereas patients with an anarchic hand complain that the hand is out of their control and "has a mind of its own", patients with utilization behaviour make no such complaint and never exhibit surprise or perplexity at their own behaviour.

Prima facie, it is difficult to see how one could account for the discrepancy between the anarchic hand syndrome and utilization behaviour by drawing only on the resources provided by holism. What global constraints might lead the patient with AH to deny that the movement is hers and at the same time lead the patient with UB to incorporate his actions into his on-going self-narrative? The holist might suggest that the differences between the AH and UB reflect pre-morbid individual differences: perhaps those AH patients are pre-disposed to alienate their stimulus-driven actions, whilst UB patients are pre-disposed to self-ascribe their stimulus-driven actions. But this proposal is puzzling, for one would expect pre-morbid differences in 'attributional style' to be evenly distributed between AH and UB patients. More plausibly, the holist might suggest that one or both of these disorders not only involves damage not only to the mechanisms of action-production but also to the mechanisms of narrative self-interpretation. But this

proposal fares no better than the previous one. Not only is it at odds with the fact that the two syndromes have different sites of neurological damage, it is also at odds with a case in which a patient exhibited utilization behaviour with his right hand and anarchic agency with his left hand: the patient was unconcerned about the former but troubled by the latter (Marcel, 2003)!

Fifthly, the holistic approach faces certain challenges in accounting for those pathologies of agentive awareness, such as the delusion of alien control, in which agents fail to experience themselves as the authors of their actions. Why can't these subjects find some way of making sense of their own behaviour? Stephens and Graham suggest that delusions of alien control arise because the patient cannot resolve the tension between what they are (in fact) doing and their self-conception without alienating the action (Stephens and Graham 2000). But there are a number of ways in which an agent can resolve the tension between their behaviour and their self-conception. For one thing, the agent could revise their self-conception. They might come to think of themselves as the kind of person who will act in certain ways. Alternatively, the agent could revise their conception of what it is that they are doing. So, why don't patients with delusions of alien control resolve the tension between their behaviour and their self-conception by adopting one of these two responses? Why don't they either revise their self-conception or find (confabulate) some conception of what they are doing that is in accord with their self-conception? Given the alacrity with which people are inclined to confabulate their intentions, proponents of the holistic approach to delusions of alien control need to explain why patients with delusions of alien control resolve the interpretative challenge posed by their behaviours in the ways that they do. We don't think that this lacuna presents a decisive objection to the holistic approach, but it is an important (and rarely-noted) challenge that needs to be addressed.

We note, in passing, that applications of the holistic approach to the schizophrenic delusion of thought-insertion are particularly problematic. How, according to such accounts, do patients with this delusion become aware of a mental state with content  $\langle P \rangle$ ? It would seem to be part and parcel of the interpretive framework that the mental states detected are one's own. The holistic approach takes there to be a thought with content P on the basis of the interpretive/narrative fit between this thought and the rest of their self-ascribed mental states. As a result, this approach has no room for the possibility of a lack of fit between the content of a state and the rest of the subject's self-conception. In short, one cannot invoke holistic considerations to explain how the subject becomes aware of a certain thought and at the same also invoke holistic considerations to explain why the subject refuses to accept that the thought in question is theirs. Of course, one could

respond to this point by adopting an atomistic account of the first-person epistemology of propositional attitudes, but at least some proponents of the holistic approach to delusions of thought insertion are not inclined to take this view. According to Stephens and Graham, “we infer the existence of propositional attitudes in ourselves in much the same way that we infer their existence in other people” (Stephens and Graham 2000: 89).

A final respect in which the narrative approach is problematic concerns the possibility of error. There is an intuitive case for supposing that it is possible for an agent to judge that they are acting when they are not acting, and vice-versa. Indeed, it is plausible to suppose that the patient with delusions of alien control makes such a mistake when she denies that she is moving her body. But it is not obvious that the narrative approach can accommodate errors of this kind, for according to the narrative approach, the same set of holistic considerations that are employed in the third-person ascription of agency are also employed by the agent themselves in the first-person ascription of agency. So it is unclear how the holistic account can explain how agents could be wrong about what it is that they are doing (or not doing); or indeed, about whether or not they are doing anything at all. Of course, this objection is by no means decisive. One might argue that the agent fails to realize that recognize their actions as their own because they lack all the relevant data that should be admitted as input to the holistic framework; or perhaps the narrative constraints that they employ in interpreting their own behaviour deviate from those that we employ in third-person ascriptions of agency. There is, clearly, much more that could be said here. Nonetheless, the objection indicates that there is clearly more work for the holistic theorist to do.

## **4.2 Arguments in favour of the atomistic approach**

We turn now from problems with the holistic approach to arguments in favour of atomism. One line of evidence for the atomistic approach comes from Libet’s well-known experiments on the timing of agency, in which healthy subjects report initiating a movement between 80-200 ms before the movement actually occurs (Libet *et al.*, 1983; Libet, 1985). In experiments extending Libet’s work, Haggard and colleagues (Haggard & Eimer, 1999; Haggard & Magno, 1999) confirmed both that intention judgments (awareness of the intention to move) and movement judgments (awareness of the movement itself) precede actual movement, and that both types of judgments are unrelated to the general readiness potential but co-vary with the lateralized readiness potential. This suggests that awareness of an intention is tied not to the general aspects of action preparation but to the selection of a specific motor program. In an experiment using Transcranial Magnetic Stimulation (TMS), Haggard and Magno (1999) showed that applying TMS over the primary motor cortex created a large delay of the actual reaction

time (movement onset) but a much smaller delay of the time of awareness of movement, whereas applying TMS over pre-motor areas, specifically the SMA, led to a much smaller delay of actual reaction time but to a greater delay in the awareness of movement. These data support the view that awareness of action onset is generated upstream of the primary motor cortex but downstream of the pre-motor structures. A further study by Sirigu and colleagues showed that patients with parietal damage could report when they started moving but not when they first became aware of their intention to move (Sirigu *et al.*, 2004). This is consistent with independent evidence that the parietal cortex is important in activating and maintaining internal models used to predict the future outcome of a given action (Sirigu *et al.*, 1996; Desmurget & Grafton, 2000).

A second line of evidence for comparator-based models appeals to studies showing that the perceptual consequences of self-generated actions are attenuated (Blakemore *et al.*, 1998; 1999, 2002; Claxton, 1975; Collins *et al.*, 1998). The predictions derived from motor commands can be used to filter incoming sensory information and thus attenuate self-produced sensory stimulation. As a result, self-produced tickling sensations are both phenomenologically and physiologically attenuated, with the amount of attenuation being proportional to the spatial and temporal congruence between the predicted and actual feedback. Notably, schizophrenic patients suffering from delusion of control do not enjoy the same level of attenuation and *are* able to tickle themselves (Blakemore *et al.*, 2000).

A third line of evidence for the comparator approach involves 'intentional binding', a phenomenon in which self-produced causes and their effects are moved closer together in subjective time (Haggard *et al.*, 2002; Haggard & Clark, 2003). More specifically, when a voluntary act (a button press) causes an effect (a tone), the action is perceived by the agent as having occurred later than it would otherwise have been perceived as occurring, and the effect is perceived as having occurred earlier than it would otherwise have been. Intentional binding depends critically on agent's experience of having intentionally produced the effect: when similar movements and auditory effects occur involuntarily rather than by the subject's intention, the binding effect is reversed and cause and effect are perceived as further apart in time than they actually are.

Haggard suggests that intentional binding is best explained in terms of predictive mechanisms of action control: it depends on efferent signals since it does not occur with passive movements and it causes anticipatory awareness of action effects, a shift that suggests prediction. On this predictive account, the conscious experience of action would be constructed at the time of the action itself, as an immediate by-product of the motor control circuits that generate and control the physical movement itself. Haggard and

Clark (2003) tested their predictive account by using TMS to insert occasional involuntary movements of the right finger at a time when the subject intended to press the button but had not yet done so. They found that intentional binding did not occur if the intention was interrupted by an imposed involuntary movement that caused the button press. These results appear to be incompatible with a reconstructive account of the sense of agency, where the existence of a match between intention and action is enough for the agent to retrospectively infer that he was the source of the action.

A fourth line of evidence for the atomistic model comes from work by Sato and Yasuda (2005), which confirms that the degree of congruency between predicted and actual sensory feedback modulates the sense of self-agency. In one of their experiments, subjects had to make self-paced presses on either of two keys. Under the congruent tone condition, the key press evoked the same tone that had been associated with that key in a previous learning session, while in the incongruent tone condition it evoked a tone at a different frequency. Similarly, in the no delay condition, the tone was immediately presented after each button press as it had in the learning session, while in the delay condition, delays of various lengths were introduced. Although the tones were actually produced as a consequence of their actions in all these conditions, the subjects' agentic experience was reduced in conditions in which self-produced tones were unpredictable. In a further experiment, participants experienced a reduced sense of agency when there was a discrepancy between predicted and actual consequences of action, regardless of the presence or absence of a discrepancy between the intended and actual consequences of the action. This result accords with Haggard's finding that a match between an intentional action and its consequence does not suffice for agentic experience.

The four lines of evidence just reviewed converge on the view that the system(s) responsible for agentic experiences are nested within the very mechanisms responsible for motor production. Of course, this research does not show that holistic mechanisms have no role to play in the generation of agentic awareness, but it does suggest that whatever role is played by such processes it will be one that takes a back-seat to those just described.

## **5. Toward an integrated model of agentic awareness**

How might holistic (narrator-based) and atomistic ("comparator-based") process cooperate to produce the rich array of forms of agentic self-awareness that we enjoy? We cannot hope to produce a full model of how these components interact here; instead, we will sketch a framework that might guide future investigations. Many details remain to be filled in.

## 5.1 The cognitive penetration of agentic experience

As we view matters, *agentic experience*—that is, our moment-by-moment sense of ourselves as the agents of various movements—is largely the output of low-level, comparator-based systems. These systems seem to be largely impervious to the agent's beliefs about what they are doing, in much the same way that the output of perceptual systems is largely impervious to the agent's beliefs about the nature of the objects in their perceptual environment. In the same way that one's judgments about the relationship between the two lines of the Müller-Lyer illusion will not change the way that they look, so too one's judgments about whether or not one is an agent (or has free will, or is really in control of one's body) will not, we suspect, have much impact on how one experiences one's own agency.

But perhaps an agent's narrative self-conception might, in *some* situations, have *some* impact on his or her agentic experience. We are certainly open to this possibility. Consider one of the classic papers of New Look psychology, Bruner and Goodman's (1947) finding that poorer children perceived coins as bigger than rich children do. This study suggests that high-level information can affect the contents of visual perception. (Of course, it is not clear that we should think of the information as question as deriving from the children's narrative self-conception, as such.) Similarly, it might be possible for narrative constraints to exert a top-down influence on agentic experience around the margins. But we suspect that, by and large, the contents of agentic experience are as impervious to an agent's narrative self-conception as are the contents of their visual experience.

## 5.2 From agentic experiences to agentic judgments

A second locus of interaction between comparator systems and the narrator concerns the transition from agentic experience to agentic judgment. According to one model of this relationship, agentic judgments are highly dependent on agentic experiences. On this model, agents will typically judge that they are the authors of a movement—that it realizes one of their own actions—if and only if they have an agentic experience with respect to it. This model would suggest that pathologies of agentic self-awareness in which the agent denies that one of their actions is their own would be grounded in pathologies of agentic experience. The aetiological account of such pathologies would, on this approach, be primarily a matter of accounting for disturbances to the agent's agentic experience.

But this is not the only way in which agentic experience and agentic judgment might be related. Instead, the agent's narrative self-conception might place rich and substantive

constraints on whether or not the contents of agentic experiences are to be accepted. In other words, one might regard the narrator as having the ability to veto the deliverances of the mechanisms responsible for agentic experience. In general, experiential states do not compel assent, and there is no reason to think that matters are any different with respect to agentic experiences. It might be the case that agents evaluate their agentic experience (or lack thereof) in light of their narrative self-conception. To put some flesh on this proposal, consider again the delusion of alien control. Perhaps patients with delusions of alien control do have a basic sense of agency with respect to those movements towards which they feel alienated, but that this basic level of agentic experience is overridden by certain holistic constraints. We hasten to say that we don't find this account of delusions of alien control compelling; delusions of control often concern mundane actions such as combing one's hair, raising one's arm or scribbling with a pen, which suggests that in many cases the very actions that the patient experiences as under alien control are not discrepant with his or her prior intentions. Nonetheless, we offer this account of alien control as an illustration of one way in which an agent's narrative self-conception might restructure their agentic self-awareness.

Furthermore, there is reason to think that in many cases the agent won't have any information about agentic experience to draw on in forming agentic judgments. The atomistic approach to agentic experiences sketched above suggests that such experiences are likely to be fragile and short-lived; they leave no trace in memory unless attentional resources are used to probe and consolidate them. When we act our attention is typically focussed on the outside world; it is only rarely that the agent's attentional focus is turned inwards. Given their short life-span, agentic experiences may well have been obliterated by the time we make an agentic judgement. A minute ago I was sitting at my desk, but I now find myself standing by the window deep in thought. Despite the fact that I have no memory of experiencing myself as walking from the desk to the window, I do not hesitate to judge that I did in fact perform this action. (Conversely, upon opening his eyes and seeing that his arm is still on the armrest, James' patient may judge that he did not move his arm despite having had an agentic experience of raising his arm.) In short, there is reason to suspect that there may be 'gaps' between agentic experience and agentic judgment, and that these gaps might be large enough for processes of narrative reconstruction to exploit.

### **5.3 The contents of agentic experience**

We turn finally to a third—and arguably the most robust—way in which narrative self-understanding contributes to agentic self-awareness, namely, by governing (or at least modulating) the self-ascription of intentions and motivations. I am currently aware of

myself as engaged in editing this paper—what processes are responsible for this kind of agentive self-awareness?

Atomistic approaches of the kind that we have been examining appear to be unable to account for our awareness of states with ‘thick’ content. If agentive experience is subserved by the mechanisms of motor control, then it will be restricted to those contents that can be reliably reconstructed from the information available to such mechanisms. These mechanisms might have access to information about the proprioceptive consequences of one’s actions, and, in the appropriate contexts, also to information about the effects of those movements in the outside world (Sato and Yasuda 2005). But it is less clear that information about one’s *conceptually-laden* (as opposed to motor) intentions can be available to motor control processes. Not only can a single (conceptually-laden) intention be implemented by means of multiple motor routines, multiple intentions can be realized by a single motor routine. In light of this, the atomistic approach to agentive experience suggests that information about what it is that one is doing—rather than merely information about how one is moving—will not be encoded in agentive experience. Yet, as we noted in section 2, there is a straightforward sense in which conceptually-laden information about what one is doing is available to awareness: not only am I aware of myself as moving thus-and-so, I am also aware of myself as making a cup of tea. But if awareness of one’s own intentions is not made available by the forward-models of motor control how is it made available to the agent?

One answer to this question—perhaps the standard answer to this question, at least within philosophical circles—is that one’s own intentions are self-intimating. Absent distraction, attentional deficits and the like, one can become aware of one’s intentions simply by attending to them. This proposal is clearly an atomistic one, although it differs from our atomistic account of agentive experience in not locating the awareness of intentions in motor processes. Although we regard this self-intimating account of intentions as a live one, the evidence marshalled by the proponents of the narrative approach suggests that our access to our own intentions is much less direct, and much more interpretive, than philosophers (and the folk!) have traditionally assumed. It could be the case that although intentions are self-intimating, at least in favourable conditions, the representations of them are fragile, and are easily swamped by the narrator’s need to maintain a coherent self-conception. A more radical possibility is that the self-ascription of intentions and motivations is largely under narrative control. On this view, the narrator would receive sparse intention-free information about the ways in which the agent is moving, and the job of the narrator would then be to fit this information with their self-

conception. The evidence is still out on which of these two possibilities is the more plausible, but our money is on the former.

## 6. Conclusion

According to some, agentic self-awareness is fundamentally holistic in nature. It involves the operations of a central-systems narrator, a homunculus who may not know much more about what we are really up to than a well-informed neighbour might. Although it contains an important element of truth, we have argued this picture of agentic self-awareness is importantly mistaken. We certainly do have a self-conception that guides our judgments about what we are up to, but this is not all there is to agentic awareness. Agentic judgments are constructed on a foundation of agentic experiences, where agentic experiences are produced by low-level mechanisms that are grounded in the very machinery responsible for action-production. Agentic self-awareness undoubtedly contains narrative elements, but it is not narrative all the way down. At least, that's the story that we're inclined to tell.<sup>5</sup>

## References

- Aarts, H., Custers, R. & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition*, 14: 439-458.
- Archibald, S. J., Mateer, C. A., & Kerns, K. A. (2001). Utilization behavior: Clinical manifestations and neurological mechanisms. *Neuropsychology Review*, 11, 3: 117-130.
- Bayne, T. (2007). The young person's guide to the phenomenology of agency. *Philosophy Compass*.
- Blakemore, S.-J. & Frith, C. D., (2003). Self-awareness and action. *Current Opinion in Neurobiology*, 13: 219-224.

---

<sup>5</sup> Early versions of this paper were presented at a workshop on delusions at the Institut Jean Nicod, at a workshop on the phenomenology of agency organized by Michael Schmitz at the Institute of Philosophy, School of Advanced Study at the University of London, and at the Macquarie Centre for Cognitive Science. We thank the audiences on those occasions for their very helpful comments. The research was partially funded by an Australia Research Council Discovery Grant, DP0452631.

- Blakemore, S.-J. Wolpert, D. M, & Frith, C. D. (1998). Central cancellation of self-produced tickle sensations. *Nature Neuroscience*, 1, 635-640.
- Blakemore, S.-J. Wolpert, D. M, & Frith, C. D. (1999). Spatiotemporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, 11: 551-559.
- Blakemore, S.-J. Wolpert, D. M, & Frith, C. D. (2000). Why can't you tickle yourself? *NeuroReport*, 11: 11-16.
- Blakemore, S.-J. Wolpert, D. M, & Frith, C. D., (2002). Abnormalities in the awareness of action. *Trends in Cognitive Science*, 6, 6: 237-242.
- Bruner, J. S. & Goodman, C. C. (1947). Value and Need as Organizing Factors in Perception, *Journal of Abnormal and Social Psychology*, 42, 33-44.
- Carruthers, P. (2007). The illusion of conscious will. *Synthese*, xxx-xx.
- Claxton, G. (1975). Why can't we tickle ourselves? *Perceptual and Motor Skills*, 41: 335-338.
- Collins, D. A., Cameron, T., Gillard, D. M., & Prochazka, A. (1998). Muscular sense is attenuated when humans move. *Journal of physiology*, 508: 635-643.
- Della Sala S., Marchetti, C., & Spinnler, H. (1991). Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia* 29, 11: 1113-1127.
- Dennett, D. (1992). The self as a center of narrative gravity. In F. Kessel, P. Cole, & D. Johnson (Eds.) *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum.
- Desmurget, M. & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Science*, 4, 423.
- Estlinger, P. J., Warner, G. C., Grattan, L. M., Easton, J. D. (1991). Frontal lobe utilization behavior associated with paramedian thalamic infarction, *Neurology*, 41: 450-52.
- Feinberg, T.E., Roane, D.M., & Ali, J. (2000). Illusory limb movements in anosognosia for hemiplegia. *Journal of Neurology, Neurosurgery & Psychiatry*, 68, 4: 511-513
- Feinberg, T.E., Roane, D.M., Cohen, J. (1998). Partial status epilepticus associated with asomatognosia and alien hand-like behaviours. *Archives of Neurology*, 55: 1574-1577.
- Frith, C. D., Blakemore, S.-J. & Wolpert, D. M. (2000a). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London B*, 355: 1771-1788.

- Frith, C. D., Blakemore, S.-J. & Wolpert, D. M. (2000b). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, 31: 357-363.
- Gazzaniga, M.S. & LeDoux, J.E. (1978). *The Integrated Mind*. New York: Plenum Press.
- Goldberg, G. & Bloom, K. K. (1990). The alien hand sign. Localization, lateralization and recovery. *American Journal of Physical Medicine and Rehabilitation*, 69: 228-38.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1-15, 90-101.
- Haggard, P. & Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Consciousness and Cognition*, 12: 695-707.
- Haggard, P. & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, 126: 128-133.
- Haggard, P. & Magno, E. (1999). Localizing awareness of action with Transcranial Magnetic Stimulation. *Experimental Brain Research*, 127: 102-107.
- Haggard, P. Clark, S. & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5(4): 382-385.
- Horgan, T., Tienson, J. & Graham, G. (2003). The Phenomenology of First-Person Agency. In S. Walter and H. D. Heckmann (Eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter: Imprint Academic, 323-40.
- Johansson, P. Hall, L. Sikström, S. & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, Oct 7; 310:116-9.
- Lhermitte, F. (1983). Utilization behavior and its relation to lesions of the frontal lobes. *Brain*, 106: 237-55.
- Lhermitte, F. (1986). Human autonomy and the frontal lobes. Part II. Patient behaviour in complex and social situations: the "environmental dependency syndrome". *Annals of Neurology*, 19: 335-343.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8: 529-566.
- Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activities (readiness potential): the unconscious initiation of a freely voluntary act. *Brain*, 106: 623-642.

- Marcel, A. (2003). The sense of agency: Awareness and ownership of action. In J. Roessler & N. Eilan (Eds.) *Agency and self-awareness*. Oxford: Oxford University Press, pp. 48-93.
- Marchetti, C. & Della Salla, S. (1998). Disentangling the alien and anarchic hand. *Cognitive Neuropsychiatry* 3, 3: 191-207.
- Metcalfe, J., & Greene, M.J. (in press). The metacognition of agency. *Journal of Experimental Psychology: General*.
- Moll, A. (1889). *Hypnotism*. London: Walter Scott.
- Moore, J. & Haggard, P. (2007). Awareness of action: inference and prediction. *Consciousness and Cognition*, in press.
- Pacherie, E. (2001). Agency lost and found. *Philosophy, Psychiatry and Psychology*, 8, 2-3: 173-176.
- Pacherie, E. (2007). The sense of control and the sense of agency. *Psyche*, 13, 1.  
<http://psyche.cs.monash.edu.au/symposia/siegel/Pacherie.pdf>
- Phillips, W., S. Baron-Cohen, and M. Rutter. (1998). Understanding intention in normal development and in autism. *British Journal of Developmental Psychology* 16: 337-48.
- Roser, M. & Gazzaniga, M.S. (2004). Automatic brains—interpretive minds, *Current Directions in Psychological Science*, 13/2: 56-59.
- Roser, M.E. & Gazzaniga, M.S. (2006). The interpreter in human psychology, in: T. M. Preuss & J.H. Kaas (Eds.), *The Evolution of Primate Nervous Systems*. Oxford: Elsevier.
- Sass, L. (1992). *Madness and modernism: Insanity in the light of modern art, literature, and thought*. New York: Basic Books.
- Sato, A. & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of ownership. *Cognition*, 94: 241-255.
- Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A. & Haggard, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience*, 7, 1: 80-84.
- Sirigu, A., Duhamel, J.-R., Cohen, L., Pillon, B., Dubois, B., Agid, Y. (1996). The mental representation of hand movements after parietal cortex damage. *Science*. 273(5281): 1564-1568.

- Spence, S. A. (2001). Alien control: From phenomenology to cognitive neurobiology, *Philosophy, Psychiatry, & Psychology*, 8, Number 2/3: 163-172.
- Stephens, G.L. & Graham, G. (2000). *When Self-consciousness Breaks: Alien voices and inserted thoughts*. Cambridge, MA: MIT Press.
- Tsakiris, M. & Haggard, P. (2003). Awareness of somatic events following a voluntary action. *Experimental Brain Research*, 149, 439-446
- Tsakiris, M. & Haggard, P. (2005). Experimenting with the acting self. *Cognitive Neuropsychology*, 22, 3/4, 387-407.
- Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Wegner, D.M., Sparrow, B. & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology*, 86/6: 838-48.
- Wegner, D. M., & Wheatley, T. P. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480-492.