

Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass*. 2 (2): 347–357 (<http://www.blackwell-synergy.com/doi/full/10.1111/j.1747-9991.2007.00067.x>)



This is a pre-print. Click here to obtain official citation information.
<http://www3.interscience.wiley.com/journal/117982788/abstract>

The natural philosophy of agency

Shaun Gallagher
Philosophy and Cognitive Sciences
University of Central Florida
gallaghr@mail.ucf.edu

Abstract: A review of several theories and brain-imaging experiments shows that there is no consensus about how to define the sense of agency. In some cases the sense of agency is construed in terms of bodily movement or motor control, in others it is linked to the intentional aspect of action. For some theorists it is the product of higher-order cognitive processes, for others it is a feature of first-order phenomenal experience. In this paper I propose a multiple aspects account of the sense of agency.

My aim in this paper is to investigate both the phenomenology and science of agency. In its proper sense, I understand agency to depend on the agent's consciousness of agency. That is, if someone or something causes something else to happen, that person or thing is not an agent (even if they might be a cause) if they do not know in some way that they have caused it to happen. A hurricane may cause the electric system to fail, but we would not attribute agency to the hurricane in what I take to be the normal use of the term. The kind of conscious knowledge involved in agency does not have to be of a very high order; it could be simply a matter of a very thin phenomenal awareness, and in most cases it is just that. If this is correct, the questions are, what do we mean by an experience or a sense of agency, and how is it generated. The remainder of this paper addresses these questions and attempts to show how complex these questions actually are.

The phenomenology of the sense of agency at two levels

One way to get at the concept of a sense of agency is to distinguish it from a sense of ownership for movement. It is possible to say that I am moving, and therefore that it is my movement, and thus have a sense of ownership for it, in cases where there is no sense of agency for the movement, for example in reflex or involuntary movements. Distinctions between self-agency and ownership of bodily movement may be found both in first-order phenomenal experience and higher-order consciousness. In regard to the latter, for example, Graham and Stephens (1994) work out their account of introspective alienation in schizophrenic symptoms of delusions of control in terms of two kinds of self-attribution.

—*Attributions of subjectivity* (ownership): the subject reflectively realizes and is able to report that he is moving. For example, he can say, "This is my body that is moving."

—*Attributions of agency*: the subject reflectively realizes and is able to report that he is the cause or author of his movement. For example, he can say "I am causing this action."

This distinction is consistent but not identical with a similar distinction made in regard to the level of first-order phenomenal consciousness (Gallagher 2000a&b).

—*Sense of ownership*: the pre-reflective experience or sense that I am the subject of the movement (e.g., a kinaesthetic experience of movement).

—*Sense of agency*: the pre-reflective experience or sense that I am the cause or author of the movement (e.g., an experience that I am in control of my action).

It seems reasonable to say that the higher-order attributions of ownership or agency may depend on the first-order experience of ownership or agency. For example, if I reach to pick up a glass, I may actually have a sense of control over the movement and so have a sense of agency for this movement; if I am then asked, did I reach for the glass, I can correctly attribute agency to myself: 'Yes, I was the one who reached for the glass'. Graham and Stephens (1994; Stephens and Graham 2000), however, suggest that the sense of agency may in fact be generated at the higher level of attribution. Following Dennett and Flanagan, they propose an explanation of the sense of agency in terms of 'our proclivity for constructing self-referential narratives' which allow us to explain our behavior retrospectively: 'such explanations amount to a sort of theory of the person's agency or intentional psychology' (1994, p. 101; Stephens and Graham, 2000, p. 161). In regard to thinking, for example, if we understand thinking to be an action on our part, then I would have a sense of agency for that thinking derived from a reflective attitude toward it.

[W]hether I take myself to be the agent of a mental episode depends upon whether I take the occurrence of this episode to be explicable in terms of my underlying intentional states (1994, p. 93).

This 'very top-down' account¹ depends on an approach according to which we reflectively make sense of our actions in terms of our beliefs and desires. So, if a

¹ I use the term 'very top-down' to distinguish the sort of account that involves reflectively conscious cognitive processes (as we find here in Graham and Stephens) from what are sometimes

subject does or thinks something for which she has no intentions, beliefs, or desires – mental states that would normally explain or rationalize such actions – the first-order movements or thoughts would not appear as something she intentionally does or thinks. Thus, whether something is to count for me as my action

[...] depends upon whether I take myself to have beliefs and desires of the sort that would rationalize its occurrence in me. If my theory of myself ascribes to me the relevant intentional states, I unproblematically regard this episode as my action. If not, then I must either revise my picture of my intentional states or refuse to acknowledge the episode as my doing. (1994, 102).

On this approach, non-schizophrenic first-order phenomenal experience appears the way it does because of properly ordered second-order interpretations, and schizophrenic first-order experience appears the way it does because of a second-order *mis-interpretation*.

[T]he subject's sense of agency regarding her thoughts likewise depends on her belief that these mental episodes are expressions of her intentional states. That is, whether the subject regards an episode of thinking occurring in her psychological history as something she does, as her mental action, depends on whether she finds its occurrence explicable in terms of her theory or story of her own underlying intentional states. (Graham and Stephens 1994, 102; see Stephens and Graham 2000, 162ff).

It would follow that the sense of agency results from an inference made on the basis of higher-order introspective or perceptual self-observations: 'what is critical [in the case of delusions of control or thought insertion] is that the subject find her thoughts [or actions] inexplicable in terms of beliefs about her intentional states' (Graham and Stephens 1994, 105).²

In contrast to this sort of very top-down account, a bottom-up approach would suggest that the sense of agency originates in neural processes responsible for the motor aspects of action. One such account proposes that efference signals or certain forward motor control mechanisms generate a phenomenal experience of agency (e.g., Blakemore, Wolpert & Frith 2002; Frith, Blakemore & Wolpert 2000; Gallagher 2000a&b; Marcel 2003; Wolpert & Flanagan 2001). Problems that develop at the neuronal level could lead to (1) the loss of the actual experience of

called 'top-down' processes in neuroscience. In neuroscience 'top-down' does not imply conscious processes. E.g., Tsakiris (2005) speaks about non-conscious body-schematic (neural) representations as being top-down processes.

² This kind of very top-down explanation ignores first-level phenomenology and has nothing to say about neurological processes that may be involved. This is not an uncommon type of analysis. See, e.g., Hoffman (1986).

agency, and (2) the generation of an actual experience of the movement or thought as alien (i.e., as caused by someone or something else), as in schizophrenic delusions. In support of this kind of account a number of neuroscientists have attempted to find the neural correlates of the sense of agency (e.g., Chaminade and Decety 2002; Farrer and Frith 2002, and Farrer, Franck, Georgieff, Frith, Decety, and Jeannerod 2003).

Experimenting with the sense of agency

A close reading of these experiments, however, raises some troubling, but nonetheless interesting questions. *Troubling* in the sense that the experiments sometimes seem confused about what they are testing. *Interesting*, nonetheless, because the experimental designs raise a question about how we should understand the sense of agency. The question is: Should we think of the sense of agency as belonging to the realm of motor control and body movement, or as belonging to the realm of the intentionality of intentional action?

In making the distinction between the sense of agency and sense of ownership (Gallagher 2000a&b), which is referenced in all of these experiments, I used the logic of involuntary movement to make the following suggestion. Since in the case of involuntary movement there is a sense of ownership and no sense of self-agency, and because my awareness of my involuntary movement comes from afferent sensory-feedback (visual and proprioceptive/kinaesthetic information that tells me that I'm moving), but not from motor commands that I issue to generate the movement (so, no efference), it seems natural to suggest that in ordinary voluntary movement the sense of ownership might be generated by sensory feedback, and the sense of agency might be generated by efferent signals that send motor commands to the muscle system. In a recent paper, Tsakiris and Haggard (2005; also see Tsakiris 2005) provide empirical evidence to support this division of labor. In their paper they also criticized the already mentioned set of experiments that attempt to identify the neural correlates of the sense of agency. In the brain imaging experiments by Chaminade and Decety (2002), Farrer and Frith (2002), and Farrer et al. (2003) the experimenters designed tasks that were meant to discriminate the sense of self-agency from the sense that someone else is the agent of an action. In these experiments, however, one common element is that subjects are required to move in each trial in order to accomplish a task. The Tsakiris-Haggard objection is that since the subjects are moving in each trial, efferent processes must be generating a sense of agency for each trial, as well as a sense of ownership. A close consideration of these experiments and the Tsakiris-Haggard objection raises the issue of whether the sense of agency is tied to voluntary bodily movement or to the accomplishment of a task.

Chaminade and Decety (2002)

For example, in the PET study by Chaminade and Decety (2002), subjects moved a joystick to control an icon on a computer screen in order to accomplish one of two tasks.

Task A (Leader): Subject moved their own circle icon and observed another subject's circle following it.

Task B (Follower): Subject followed another subject's icon with their own.

The authors write:

Investigation of the neural basis of agency was performed using a paradigm in which the subjects either led (A) or followed (B) the other, in a computerized environment free of explicit reference to body parts. ... The sense of ownership, related to motor control, and the sense of agency, related to the intentional aspect, can be segregated in the analysis. (Chaminade and Decety 2002, 1977)

The reason for this latter claim is that movement was required in each case (so the sense of ownership would be constant – it is my movement in both cases), but the sense of agency would be different for A vs B – the assumption being that A (leading) would generate a sense of agency, while B (following) would not. These distinctions led the researchers to identify activation in the pre-supplemental motor area (SMA) and the *right inferior parietal cortex* as responsible for generating the sense of self-agency, and in contrast, activation of the *left inferior parietal cortex* and the right pre-central gyrus as responsible for attributing agency to another. They note the relevance to schizophrenia – ‘abnormal increased activity in the right inferior parietal cortex has been observed in schizophrenic patients experiencing passivity phenomenon’ (p. 1978).

One objection, of course, is that in both cases (A and B) the subject may have a sense of agency for the intentional aspect – i.e., accomplishing the task. The subject might say, ‘My task in A is to lead, and I have done so, and my task in B is to follow, and I have done so. I am the agent of both of these actions, respectively (leading and following).’ So differential activation of these brain areas may be for something other than agency,³ and may even be for something other than the difference between self-agency and other-agency. The Tsakiris-Haggard objection, however, is somewhat different. It is that the subject will necessarily have a sense of agency in both A and B because in both the subject moves his hand to control the joystick. Tsakiris and Haggard understand agency to be directly tied to motor control, and in that regard, to efference signals (like a number of other authors mentioned above). In contrast, Chaminade and Decety associate the sense

³ Indeed, the results of this experiment don't fully match up with results found in other experiments, including the experiments involving passivity phenomena in schizophrenics which involve activation of the right IPC (indicating other-agency) rather than the left IPC suggested in this experiment. See Farrer and Frith (2002, p. 597): ‘In previous studies attribution of actions to another has been consistently associated with activity in the right inferior parietal lobe. Patients with delusions of control who erroneously attributed their actions to another showed abnormally high activation in this region (Spence *et al.*, 1997). Subjects imagining someone else acting showed greater activity in this area than when they imagined themselves making the action (Ruby and Decety, 2001)’. The experiments in Farrer and Frith (2002) suggested that the right IPC is activated for other-agency.

of agency with the intentional aspect of action – the accomplishment of the task. So even if we think that the sense of agency is generated in first-order phenomenal experience rather than at a higher-order cognitive level, there is still a question about whether it is generated by motor processes or by some awareness of the intentional aspect of action.

Farrer and Frith (2002)

In the fMRI experiment by Farrer and Frith,

Subjects manipulated a joystick [to drive a colored circle moving on a screen to specific locations on the screen]. Sometimes the subject caused this movement and sometimes the experimenter. This paradigm allowed us to study the sense of agency without any confounding from the sense of ownership. To achieve this subjects were requested to execute an action during all the different experimental conditions. By doing so the effect related to the sense of ownership (I am performing an action) would be present in all conditions and would be canceled in the various contrasts. (2002, 597).

Like Chaminade and Decety, Farrer and Frith cite the distinction between sense of agency and sense of ownership, but associate the sense of agency with the intentional aspect of action, i.e., whether I am having some kind of effect with respect to the goal or intentional task. Again, the claim is that the sense of ownership remains constant while the sense of agency changes. But again, the Tsakiris-Haggard objection is that since in each case the subject is required to move the joystick, a sense of agency for that movement must result. In this experiment, however, the subject is asked to perform the task regardless of whether what happens on the screen is known to the subject to be his action or the action of someone else. In some sense, when the subject knows that it is not he who is controlling the screen, his action short-circuits in the movement of the joystick – he knows that his movement doesn't accomplish anything, and his *right inferior parietal cortex* is activated. Supposedly he does not have a sense of agency for what happens on the screen (I say supposedly, because in none of these experiments are the subject's asked). When he does know that he is causing the action on the screen, his *anterior insula* is activated bilaterally, and this is what the experimenters identify as the correlate of the sense of agency.⁴ The sense of agency is thus tied to the intentional task (what happens on the screen) rather than to the bodily movement (which happens whether the subject controls what is on the screen or not).

Now, one might think that this is rather straightforward. Farrer and Frith clearly think of the sense of agency as something tied to the intentional aspect of

⁴ Interestingly, when the subject was not told that it was another person controlling the screen (finding out only in the last second when the circle moved differently from what he intended), no significant differences were found. Apparently, in that case his sense of agency for the task was intact. But this is not made clear by the experimenters.

action and not to mere bodily movement. In that regard they would claim to sidestep the Tsakiris-Haggard objection. Curiously, however, when it comes to *explaining why* the anterior insula should be involved in generating the sense of agency, Farrer and Frith revert to a explanation more consistent with the Tsakiris-Haggard objection, that is, they explain the involvement of the anterior insula in terms of motor control.

Why should the parietal lobe have a special role in attributing actions to others while the anterior insula is concerned with attributing actions to the self? The sense of agency (i.e., being aware of causing an action) occurs in the context of a body moving in time and space. Damasio (1999) has suggested that the sense of agency critically depends upon the experience of such a body. There is evidence that both the inferior parietal lobe and the anterior insula are representations of the body the anterior insula, in interaction with limbic structures, is also involved in the representation of body schema One aspect of the experience of agency that we feel when we move our bodies through space is the close correspondence between many different sensory signals. In particular there will be a correspondence between three kinds of signal: somatosensory signals directly consequent upon our movements, visual and auditory signals that may result indirectly from our movements, and last, the corollary discharge [efference signal] associated with motor commands that generated the movements. A close correspondence between all these signals helps to give us a sense of agency. (601-02).⁵

If this is the case, then the fact that for each task the subject was moving does indeed complicate things – as the Tsakiris-Haggard objection contends.

Farrer, Franck, Georgieff, Frith, Decety and Jeannerod (2003)

The third study that sets out to address the neural correlates of the sense of agency is of a different sort, even though its authors include Decety, Frith, and Farrer. In this experiment subjects do provide a report on their experience; however, all questions about agency were focused on bodily movement. Subjects were not given any intentional task to carry out other than the task of moving their hand, and the focus of their attention was directed towards a virtual (computer image) hand that either did or did not represent their own hand movements, although at varying degrees of rotation relative to true position of the subject's hand.⁶ Interestingly, the

⁵ To make things worse, they cite well-known evidence that the inferior parietal cortex, which they are associating with a sense of other-agency, is responsible for a sense of body ownership – ‘patients with right parietal lesion do not recognize their limbs as their own and perceive them as belonging to others’ (601).

⁶ ‘Four main conditions were used: (1) a condition where the subject had a full control of the movements of the virtual hand, (2) a condition where the movements of the virtual hand appeared rotated by 25° with respect to the movements made by the subject, (3) a condition where the movements of the virtual hand appeared rotated by 50°, and (4) a condition where the movements of the virtual hand were produced by another person and did not correspond to the subject's movements’ (Farrer et al. 2003, 324).

same brain areas involved in the Farrer and Frith experiment were activated in this experiment. Subjects moved their own hand, but saw a virtual hand projected on screen at veridical or non-veridical angles. The virtual hand was either under their control, or not. Subjects were asked about their sense of agency for their bodily movements and the visual (virtual hand) representation of those movements. The less the subject felt in control, the higher the level of activation in the *right inferior parietal cortex*. The more the subject felt in control (the stronger the sense of agency), the higher the level of activation in the *right anterior insula*.

There seems to be a clear shift from the previous experiments, in which the sense of agency was construed in terms of an intentional task that went beyond mere bodily movement, to this one, in which the sense of agency is construed in terms of bodily movement and motor control. In this case the Tsakiris-Haggard objection seems to hold. If the sense of agency is generated by mere bodily movement rather than task-related action (at least a kind of purposive action that goes beyond simply moving one's hand for the sake of an experiment) – and bodily movement does seem to be the only thing at stake in this experiment – then the fact that the subject moves his own hand in all trials in this experiment certainly does not provide any way to discriminate the sense of ownership from the sense of agency, and suggests that the subject should have a sense of agency for all movements of her body.

One could argue that the sense of agency in this experiment is being construed, not in terms of efference signals (since they are present in each case), but in terms of sensory (visual) feedback, or more precisely, in terms of the integration of visual and proprioceptive feedback. When there is incongruency between vision of the virtual hand and proprioception of the subject's real hand, there is a disruption in the sense of agency. But the idea that the sense of ownership for movement may depend on precisely the same kind of sensory feedback, and the fact that in this experiment we get the same results as in the second experiment, where there was some extra-experimental/theoretical indication that these brain areas may also involve sense of ownership, puts the conclusions about sense of agency into doubt. Indeed, the authors cite even more evidence along this line: 'Lesions of the inferior parietal cortex, especially on the right side, have been associated with delusions about the patient's limb that may be perceived as an alien object or as belonging to another person' (Farrer et al. 2003, 329).⁷ Such delusions are about ownership rather than agency. In addition, in Tsakiris and Haggard (2005), activity in the insula was also found in the absence of movement, which implies that this area may in fact reflect body-ownership rather than agency.

⁷ The authors suggest something slightly different, while admitting the possibility that it may be sensory discordance. 'We have proposed the activity seen in inferior parietal cortex relates to the feeling of loss of agency associated with the discrepancy between intended actions and sensory feedback. However, from the experiment discussed so far it is possible that the activity in this region relates solely to the sensory discordance. The feeling of agency might relate to activity in other regions. We think this is unlikely on the basis of various pathological cases in which the primary disorder concerns the feeling of agency rather than sensory discordance' (Farrer et al. 2003, 329). But the pathological studies they cite indicate something about delusions of ownership rather than agency.

Theoretical integration

Where does all of this leave us? First, I want to argue that intentional action is not about mere bodily movement. Consistent with phenomenological theories of embodiment, in everyday engaged action afferent or sensory-feedback signals are attenuated, implying a recessive consciousness of our body (see e.g., Tsakiris and Haggard 2005). We do not attend to our bodily movements in most actions. We do not stare at our own hands as we decide to use them; we do not look at our feet as we walk, we do not attend to our arm movements as we engage the joystick. Most of motor control and body schematic processes are non-conscious and automatic. It still may be the case, however, since action is embodied, that just such processes contribute to a sense of agency, and without the feeling of the embodied nature of action our sense of agency would be very different. But I also want to suggest that the sense of agency is not reducible to just these embodied processes. Nor is it something that is simply in our Cartesian minds, the product of higher-order reflection. If our descriptions and explanations of what we are doing in action are cast at the highest pragmatic level of description – ‘I’m helping my friend’, or ‘I am on my way to the pub’, or whatever, rather than ‘I’m moving my hand’, or ‘I am walking’, then clearly our sense of agency for the action will be tied to that intentional aspect, and that aspect is where our attention is directed – *in the world*, in the project or task that we are engaged in. So clearly a form of *intentional feedback*, which is not afferent feedback about our bodily movements, but some perceptual sense that my action is having an effect, must contribute to the sense of agency.

I suggest, then, that the sense of agency, at the first-order level of experience, is complex because it is the product of several contributory elements: efferent signals, sensory (afferent) feedback, and intentional (perceptual) feedback. If any of these contributory elements fail, or fail to be properly integrated, then we can get a disruption in the sense of agency.

This also suggests that the loss of the sense of agency in various cases – including schizophrenia, anarchic hand syndrome, obsessive-compulsive behavior, narcotic addiction, etc. – may in fact be different sorts of loss. In any particular case the sense of agency might be disrupted in different ways depending on what contributory element is disrupted.

In general, we can identify three concepts of the sense of agency (SA).

- SA as first-order experience linked to intentional aspect (task, goal, etc.) (Chaminade & Decety 2002; Farrer & Frith 2002)
- SA as first-order experience linked to bodily movement (Farrer et al. 2003; Gallagher 2000a&b; Tsakiris & Haggard 2005).
- SA as second-order, reflective attribution (Graham and Stephens 2000)

This also gives us four possibilities for explaining the pathological loss of the sense of agency.

- **Very Top-down:** SA may be disrupted by problems with introspective higher-order cognition (Graham and Stephens) -- this may very well be the case in advanced and involuted symptoms of schizophrenia.
- **Bottom-up:** SA may be disrupted by problems with motor control mechanisms -- efference signals (Tsakiris and Haggard) or the integration of sensory and motor signals in the anterior insula (Farrer et al.)
- **Intentional theory:** perceived lack of concordance between intention and effects of action -- (Chaminade & Decety, Farrer & Frith)
- **Multiple aspects:** SA is complex, and based on the integration of efferent, afferent, and intentional feedback (some sense that my action is having the intended effect on the world)

The multiple aspects option is the one I am defending in this paper.

References

- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn Sci* 6(6): 237-242.
- Chaminade, T. and Decety, J. 2002. Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport* 13 (1528): 1975-78.
- Farrer, C., Franck, N. Georgieff, N. Frith, C.D. Decety, J. and Jeannerod, M. 2003. Modulating the experience of agency: a positron emission tomography study. *NeuroImage* 18: 324–333
- Farrer, C. and Frith, C.D. 2001. Experiencing oneself vs. another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage* 15: 596-603.
- Frith, C. D., Blakemore, S., & Wolpert, D. M. 2000. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res Brain Res Rev*, 31(2-3), 357-363.
- Gallagher, S. 2000a. Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science* 4 (1): 14-21.
- Gallagher, S. 2000b. Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. In D. Zahavi (ed.), *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-experience* (pp. 203-39). Amsterdam & Philadelphia: John Benjamins.
- Graham, G. and Stephens, G. L. 1994. Mind and mine. In G. Graham and G. L. Stephens (eds.), *Philosophical Psychopathology* (pp. 91-109). Cambridge, MA: MIT Press.
- Hoffman, R. 1986. Verbal hallucinations and language production processes in schizophrenia. *Behavioral and Brain Sciences* 9: 503-517.
- Marcel, A. 2003. The sense of agency: Awareness and ownership of action. In J. Roessler & N. Eilan (Eds.), *Agency and Awareness* (pp. 48-93). Oxford: Oxford University Press.
- Stephens G. L. and Graham G. 2000. *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge MA: MIT Press.

- Tsakiris, M. 2005. On agency and body-ownership. Paper presented at Expérience Subjective Pré Réflexive & Action (ESPRA) Conference, CREA, Paris. December 2005.
- Tsakiris, M. & Haggard, P. 2005. The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1): 80-91.
- Wolpert, D. M., & Flanagan, J. R. 2001. Motor prediction. *Current Biology* 11(18), R729-