

Neurocognitive Models of Schizophrenia: A Neurophenomenological Critique

Shaun Gallagher

Department of Philosophy, Cognitive Sciences Program, University of Central Florida, Orlando, Fla., USA

Key Words Schizophrenia W- Frith's model - Delusions of control - Thought insertion - Agency - Ownership - Top-down explanations - Bottom-up explanations - Neurophenomenology

Abstract This paper argues that Frith's (1992) account of the positive symptoms of schizophrenia in terms of a disruption of metarepresentational self-monitoring is inadequate in several specific ways. More generally, this paper argues against top-down explanations for the loss of the sense of agency in such symptoms. In addition, even if delusions of control might be explained by problems involved in motor control mechanisms involving efference copy and comparators, there are good reasons why the same model cannot explain thought insertion. In place of such neurocognitive explanations, the author develops a neurophenomenological explanation for the loss of the sense of agency and the misattribution of actions and thoughts to others in such symptoms.

In the past dozen years, a number of theoretical models of schizophrenic symptoms have been proposed, often inspired by advances in the cognitive sciences, and especially cognitive neuroscience. Perhaps the most widely cited and influential of these is the neurocognitive model proposed by Christopher Frith [1]. Frith's influence reaches into psychiatry, neuroscience, and even philosophy. The philosopher John Campbell [2], for example, has called Frith's model the most parsimonious explanation of how self-ascriptions of thoughts are subject to errors of identification. 'On reflection, it also seems that this is not just one possible theory; it is the simplest theory which has any prospect of explaining the sense of agency, and we ought to work from it, introducing complications only as necessary' [2].

Not everyone agrees. In their recent analysis of alien voices and inserted thoughts in schizophrenia, Stephens and Graham [3] offer a critique of Frith's model. Their criticism, however, although serious, is neither deep nor extensive. They outline three points: first, Frith fails to provide an adequate account of why a subject who experiences thought insertion would misattribute that thought to some other agent; second, Frith does not clarify the distinction between thought insertion and thought influence, and third, Frith fails to explain how a subject can claim both that he is thinking the thought and that the thought is someone else's thought. I suggest that this critique is not sufficiently deep in the sense that Stephens

and Graham [3] do not challenge Frith's cognitive framework, only certain failures within that framework. Indeed, as I will indicate below, Stephens and Graham's [3] own cognitive model shares certain important features with Frith's.

In this paper, I propose a deeper and more extensive critique of Frith's model challenging the cognitive framework that Campbell [2] defends as the most parsimonious one. This critique is based on a phenomenological approach that is informed by neuroscience, and in this sense it can be called a neurophenomenological analysis [4, 5]. The paper is divided into three sections. In the first, I provide a brief outline of Frith's analysis; I explicate the phenomenological critique of this analysis in part two, and in the final part, I outline a neurophenomenological approach that addresses the various problems found in Frith's account.

Frith's Model

Frith's account of positive symptoms such as verbal hallucinations, delusions of control, and thought insertion moves from an explanation of motor behavior to an explanation of cognitive experience. This strategy involves a transposition of certain neurological mechanisms responsible for motor control to the level of cognitive mechanisms responsible for monitoring and controlling thought processes. Across these different levels, the problem for the schizophrenic is the same: a failure in basic self-monitoring processes. To explicate Frith's model, it will be helpful to use a distinction between the sense of agency and the sense of ownership that Frith later adopts [6].

Prereflective self-awareness may include a *sense of agency* (a sense of being the initiator or source of a movement, action, or thought) and a *sense of ownership* (a sense that it is I who am experiencing the movement or thought [3, 7, 8], or what some phenomenologists refer to as the 'myness' of the experience [9]). In the normal phenomenology of voluntary or willed action, the sense of agency and the sense of ownership coincide and are indistinguishable. In the case of *involuntary* or unintended movement, however, they are clearly distinguished. I may acknowledge ownership of a movement – for example, I have a sense that I am the one who is moving or is being moved. I can thus self-ascribe it as my movement. I may rightly claim, however, that I am not the author of the movement, because I do not have a sense of causing or controlling the movement, i.e. I have no sense of agency for it. The agent of the movement, for example, is the person who is pushing or pulling or manipulating my limbs.

Likewise, in the case of involuntary *cognitive* processes, I may acknowledge that I am the one who is thinking, but claim that the thoughts are not willfully generated by me. For example, certain unbidden thoughts or memories may impinge on my consciousness, even if I do not intend for them to do so, or even if I resist them [10, 11]. Melodies stay in our heads when we would rather think of something else, for example. In such cases, of course, we may not want to say that there is a specific agent for thinking. It is not that I think someone else is causing my thoughts or my experience. Nonetheless, my claim of ownership (my self-ascription that I am the one who is undergoing such experiences) may be consistent with my lack of a sense of agency.

Not unlike unintended movement and thought, positive symptoms of schizophrenia such as delusions of control and thought insertion involve a disruption in the sense of agency, but not in the sense of ownership. Frith's cognitive model of schizophrenia as a disruption of basic self-monitoring processes can be understood to be a possible explanation of how self-ascriptions of thoughts and actions may be subject to errors of identification in regard to agency.

Frith is led to this view from observations of motor behavior in schizophrenia [1]. Chronic schizophrenic patients suffer from a variety of movement disorders, and often make mistakes about the agency of various bodily movements. Not only do they fail to recognize their own agency in such cases, but schizophrenics suffering from delusions of control misattribute agency for their movements to someone or something else. Frith provides an example: 'the force moved my lips. I began to speak. The words were made for me' [1]. The motor action responsible for the speech is in fact the patient's own motor action, and the patient acknowledges that they are his lips that are moved, but he makes an error of identification concerning who produced this motion. Here the sense of agency, rather than the sense of ownership, is disrupted. That is, the patient knows that they are his lips and that he speaks, but seemingly his lips were moved, and the words were generated by someone else.

To explain this disruption in the sense of agency, Frith turns to a classic theory of motor control which involves a hypothetical brain mechanism termed the *comparator*. The comparator model states that when a motor instruction to move is sent to a set of muscles, a copy of that instruction, the efference copy, is also sent to a comparator or self-monitoring system. Held [12] suggested that the efference copy sent to a comparator is stored there, and then compared to reafferent

(proprioceptive or visual) information about the movement that is actually made. In this model, confirmation that I moved would seemingly have to wait for sensory feedback; it would come after the fact as a verification that it was I who did the

moving. In the case of the schizophrenic's motor experience, what explains his lack of a sense of self-movement, in this model, is the absence of a match between movement and efference copy at the comparator.

According to this *sensory feedback* model, if something seems to be going wrong with the action, it is quite possible to correct for it on the basis of this sensory monitoring. However, Frith realized that this model alone is not sufficient to explain motor control. A second component involved in the normal control of movement can also be interpreted in terms of a comparator model. In this respect, the comparator is understood to be part of a forward premotor system delivering information *prior to* the actual execution of movement and *prior to* sensory feedback [13, 14]. The forward model derives from an explanation of visuomotor coordination. The visual system can distinguish between movements on the retina that are due to movements in the world, and movements on the retina that are due to movements of the perceiver's own body. In the latter case, stability of the visual image is achieved by the motor system sending an efference copy 'to some monitoring system at the same time as a message is sent to the eye muscles' [1]. The efference copy alerts the visual system to compensate for self-generated or self-initiated movement. Frith refers to this as a monitoring of *intentions to move*. The forward comparator thus functions to match the efference copy with motor *intentions*, anticipating the movement, which allows for quicker and automatic corrections of movements prior to any sensory feedback. This forward system not only assists in the generation of controlled action, it also generates a conscious sense that one is generating the action [15, 16]. Preaction neuronal processes anticipate the actual motor performance and provide an online sense of agency that correlates with the sense of movement. In one's immediate phenomenology during action, agency is not represented as separate from the action, but is an intrinsic property of action itself, experienced as a perspectival source [17].

Experiments have demonstrated that schizophrenics have problems with the forward monitoring of movement, but not with motor control based on sensory feedback. Nonschizophrenic subjects required to use a joystick to follow a target on a computer screen are able to use either visual feedback or the preaction, more automatic process for correction of movement. In the first case, to correct a mistaken movement one depends on visual perception of one's hand. When subjects recognize the error, they move to correct it. If visual perception of the hand is unavailable, however, the normal subject makes a quicker and smoother correction of errors. Such rapid error corrections presumably depend on the forward monitoring of motor intention, prior to any sensory feedback. Schizophrenic patients, however, find it difficult to monitor their own motor intentions at this level [18]. Like normal subjects, they correct their errors when visual feedback is provided, but, unlike normal subjects, they often fail to correct their mistakes when deprived of visual feedback [19].

In delusional experience of control, the schizophrenic feels that he is not the agent of his movement. This loss of a sense of agency is explained by Frith as a disruption of the efference copy in the central monitor, a dysfunction of the forward, preaction aspect of the motor system. This makes good sense with respect to normal involuntary action: where a subject does not intend the action, there would be no preaction preparation, no intention registered at the comparator. In the case of the schizophrenic's delusion of control, however, since he is in fact the agent, the disruption of the sense of agency is explained by something going wrong with the efference copy or the forward comparator mechanism.

According to the experimental data, the schizophrenic does seem able to correct movement errors if sensory feedback is available. Likewise, in the case of involuntary action, the sensory feedback system still seems to do its job, providing a sense that 'I am moving'. Although things are likely to be more complicated at the neurological level, this suggests that the distinction between sensory feedback control and forward preaction control corresponds to the distinction between sense of ownership and sense of agency, respectively. That is, both in the schizophrenic's delusional experience of control and in the normal subject's experience of involuntary action, the sensory feedback system tells the subject that it is he who is moving or being moved (providing a sense of ownership for the movement). With a disruption of efference copy at the forward comparator, however, the system will fail to register a sense of agency, a sense that it is the subject himself who is the willful generator of the movement.

Can something like this model of motor control be applied to cognition? Frith [1] and his colleague [20] think that it can. Frith points out that this sort of self-monitoring mechanism is traditionally used in explaining motor, perceptual, and linguistic behavior [21, 22]. Following Feinberg [23], however, he postulates a similar mechanism for cognition – specifically, for thought and inner speech. He suggests that defects in such mechanisms can explain phenomena such as thought insertion and auditory hallucination.

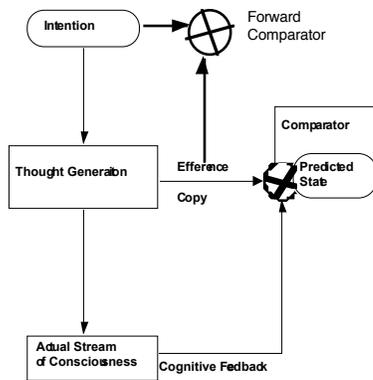


Fig. 1: Match at a feedback comparator identifies thought as happening in one's own system; match at the forward comparator generates a sense of agency.

Frith proposes the following explanation of thought insertion.

Thinking, like all our actions, is normally accompanied by a sense of effort and deliberate choice as we move from one thought to the next. If we found ourselves thinking without any awareness of the sense of effort that reflects central monitoring, we might well experience these thoughts as alien and, thus, being inserted into our minds [1, p. 81].

Frith's model assumes not only that thinking is a kind of action, but that, as in the case of a motor action, we experience an effortful intention. The intention to think, according to Frith, is the element that bestows a sense of agency for the thought. Normally, an efference copy of the intentional generation of thought is sent to a comparator or central monitor, which also registers the occurrence of the actual thought, thus matching up intention and thought (fig. 1). So, if the efference copy is somehow blocked from reaching the central monitoring mechanism, thought occurs which seems not to be generated by the subject. If the efference copy is blocked or goes astray, or is not properly generated, thinking still occurs, but it is not registered as under my control – it appears to be an alien or inserted thought. There is no match between intention and thinking.

We now have the basic framework that Frith uses to explain the delusional symptoms of schizophrenia. The various mechanisms of control are meant to be cashed out in neurological terms. To these mechanisms, however, and as part of the way they work, Frith would add the operations of a higher-order cognitive level [1]. He speaks of a *conscious* feeling of effort for a willed intention to think, and he equates this with a *conscious* monitoring of an efference copy. Frith's analysis relies, not just on an intention to act (to move or to think), but on an *awareness* of the intention to act, and he defines this awareness as a case of 'metarepresentation'. Metarepresentation is a second-order reflective consciousness, 'the ability to reflect upon how we represent the world and our thoughts'. This is part of what it means to monitor our actions and, he claims, it is precisely what is missing or is disrupted in the schizophrenic's experience. In schizophrenic delusions of control and inserted thought, there is a failure of metarepresentation; the metarepresentational monitoring of movement or thought breaks down.

A Phenomenological Critique of Frith's Model

Intentions to Think

Several aspects of Frith's model are phenomenologically problematic. A first set of problems pertains to Frith's characterization of the intention to think. In the case of thinking or conscious experience, what role does something like an intention to think play? It is difficult to conceive of an intention to think prior to thinking itself, unless it is entirely a conscious preparation, as when I might decide to sit down and start thinking about a certain issue. In that case, however, the intention to think is itself a thinking, and an infinite regress begins to loom: do I require an intention to think in order to intend to think?

Stephens and Graham [3] consider this objection, but not in relation to Frith's analysis. Rather they discuss an objection raised by Akins and Dennett [24] against Hoffman's [25] account of alien voices. Akins and Dennett suggest that the idea of having an intention to think leads to a 'never-beginning regress of intentions to form thoughts' [p. 517]. Stephens and Graham [3] point out, however, that Hoffman's [25] account concerns only the kind of thought that might be called inner speech, and just as we might accept the idea that someone may have an intention to say something, it is not so odd to accept the idea that someone may have an intention to say something in inner speech. They agree, however, that

not all intentional thinking could be described as inner speech, and that the idea that there is an intention to think in every case would be unacceptable. Yet in Frith's model, an intention to think, and indeed, an awareness of that intention to think is required to provide a sense of agency for thoughts.

Campbell [26] suggests, as a corrective to Frith's characterization, that efference copy (and thus the intention to think that it represents) is not itself available to consciousness. In his view, efference copy is part of a *subpersonal* and nonconscious process that generates an awareness of effort as thought itself is generated. To get the phenomenology right, Campbell [26] moves the intention to think underground to a subpersonal mechanism. As thoughts are generated (something that might happen in any number of ways), an efference copy is normally sent to a comparator. Thought generation involves subpersonal processes that subtend the thinking process. Ultimately this is to be cashed out in terms of neuroscience rather than phenomenology.

At the very least, then, we should say that most cases of normal thinking are neither prefaced by conscious intentions to think, nor followed by an introspective awareness of that intention. In the normal phenomenology, at least in the large majority of cases, there is not first an intention and then a thinking, nor thinking plus a concurrent but separate awareness of intention to think.

Unbidden Thoughts

What Frith calls 'intention to think', then, may be part of a subpersonal, nonconscious process. When it works properly, it generates a sense of agency for our thoughts. Does the disruption of this process, however, explain the phenomenology of inserted thoughts?

As with movement, it is not always the case that my thinking is characterized as something I intend. There are unbidden thoughts that, as Frankfurt [10] puts it, 'strike us unexpectedly out of the blue; and thoughts that run willy-nilly through our heads' [10, p. 240]. I may just find myself thinking of something. Without a conscious act of recollection, particular memories may come to the fore, invading or disrupting my thoughts with elements of my past that may or may not be relevant to my present circumstance. Is there any kind of intention to think in such cases? If in such cases I do not have a sense of agency for my thoughts, still, I do not attribute my thoughts to someone else in the way that the schizophrenic does. This means, however, that the lack of an intention to think (or the disruption of an efference copy) is not doing all the work that Frith would like it to do in the case of schizophrenia. The absence of an intention to think, and hence, the lack of an efference copy generated to represent that intention at the comparator level do not explain anything more than a relatively normal lack of a sense of agency. As Stephens and Graham [3] point out in their critique of Frith's theory, it does not explain the attribution of thought to another agent, or why some thought might seem inserted for the schizophrenic. At best, in the absence of an intention to think, we have unbidden thoughts, but not inserted thoughts.

Redundancy of Efference Copy

If the efference copy/comparator mechanism fails to explain the phenomenology of inserted thoughts, specifically their misattribution to others, we can push things a bit further by asking whether we need the efference copy/ comparator mechanism at all in the case of thinking. Campbell [26] argues that Frith's model provides the most parsimonious explanation of the sense of agency for cognition. Yet Frith's model would add to the cognitive system a separate mechanism to generate a sense of agency for thinking. Is this necessary in a system that is already conscious?

In the case of visuomotor control, the efference copy serves a pragmatic, executive function rather than a verificational one. In effect, one system informs another system to make adjustments, with very practical effects, such as stability of the visual field, or postural balance. The function of the efference copy is to inform the visual and vestibular systems that the organism, rather than the world, is moving. Even in the case where the motor system is simply updating and correcting itself, the purpose of the efference copy is for motor control. Its purpose is not to verify (simply for the record?) that movement is taking place, nor is it *primarily* an information stream that discriminates between intended movement and nonintended movement. Its purpose is rather to instruct the motor or sensory system to make important adjustments. Is there anything like this happening in the thinking process?

One can certainly distinguish different cognitive systems, such as memory systems, or perceptual systems. However, efference copy, as described by Frith, does not play a communicative role among these systems. Rather, in the Feinberg-Frith model, consciousness seems to be sending *itself* messages. Thus, Campbell [26] suggests, following Feinberg [23], that efference copy has the primary function of keeping thoughts on track, checking 'that the thoughts you actually execute form coherent trains of thought' [26, p. 616]. To keep thoughts coherent and on track, however, could only mean to keep them on a semantic track, that is, on a certain track of meaning or on a line of logical reasoning. It seems odd, however, to assign this task to a subpersonal, nonsemantic mechanism when, simply put, we are already consciously

aware of our thoughts and can keep track of them, and keep them on track, at a conscious level. Indeed, it seems more appropriate to grant this task to the kind of metarepresentational introspection that Frith talks about. I may indeed require a reflective introspection to make sure that my logical reasoning is on track, or that my thoughts form a relatively coherent whole. But all of this seems quite a different task from verifying that the thoughts that I am thinking are self-generated.

Is there a relationship between maintaining a sense of coherency in thought and a sense of agency? One version of a positive answer to this question is given by Stephens and Graham [3].

We propose that a person's sense that he is a thinker or agent of his *mental* activity – of his conscious thoughts and feelings – likewise depends on his conviction that his occurrent mental episodes express his intentional states. That is, whether a person regards a thought (subjectively) in him as something that he *thinks*, rather than a mere episode in his psychological history, depends upon whether he finds it explicable in terms of his conception of what he believes and desires [3, p. 165].

In this respect, Stephens and Graham [3] share part of the same framework that Frith constructs. Specifically, they agree that the sense of agency in some way depends on higher-order cognitive processes. For Frith, the sense of agency depends, at least in part, on the monitoring of a metarepresentational introspection of first-order phenomenal experience. Stephens and Graham [3] seem able to do without Frithian subpersonal mechanisms, but they too require a second-order level of introspection to guarantee the agentive sense for thought.

[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 [11, p. 93].

Their top-down account involves a theory of mind approach in which we reflectively make sense of our actions and our thoughts in terms of our beliefs and desires. So, if a patient does something or thinks something for which he has no beliefs or desires, mental states that would normally explain or rationalize such actions, the first-order movements or thoughts would not appear as something he intentionally does or thinks. 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 [11, p. 102].

The problem with this top-down approach is that there are many kinds of thoughts that may not be coherently integrated with a subject's system of beliefs or desires, but may nonetheless not seem to be inserted; for example, some instances of unbidden thoughts, or in the case of compulsive disorders, compulsive thoughts which do not express the patient's beliefs or desires and yet are not claimed by the patient to be inserted by others [27].

Frith proposes a model that is more hybrid than the one outlined by Stephens and Graham [3]; that is, Frith's model involves both higher-order (metarepresentational) cognition and subpersonal mechanisms. It is not clear, however, how precisely these two components fit together. The fact that the comparator mechanism for conscious thought is subpersonal means that the matching process seemingly involves both conscious and nonconscious elements. Campbell [2] is right to say that efference copy is not something of which we are conscious. But in Frith's model, we need to match that efference copy to the stream of thought itself, which is already a matter of consciousness. Supposedly the *outcome* of the match, the sense of agency, must also be conscious in some way – a conscious sense that I am the one who is thinking. One might suppose, however, that the match process itself in the subpersonal comparator would not be conscious (otherwise there would have to be a consciousness of something that remains nonconscious – efference copy).

You have knowledge of the content of the thought only through introspection. The content of the efferent copy is not itself conscious. But it is match at the monitor between the thought of which you have introspective knowledge and the efferent copy that is responsible for the sense of being the agent of that thought. It is a disturbance in that mechanism that is responsible for the schizophrenic finding that he is introspectively aware of a thought without having the sense of being the agent of that thought [2].

This kind of hybrid process seems less than parsimonious when the particular thing that we are supposedly monitoring – namely, our thought process – is already consciously available. On top of this, as we have seen, nothing in this hybrid system seems to explain why the schizophrenic who lacks a sense of agency goes on to misattribute agency to someone or something other than himself.

Hyperreflexivity

Campbell [2] describes the comparator process as involving a form of introspection: 'it is the match between the thought detected by *introspection*, and the content of the efferent copy picked up by the comparator, that is responsible for the sense of ownership of the thought' [2]. Frith invokes the notion of metarepresentation, a fullfledged act of reflection, as part of the monitoring process. But a metarepresentational introspection again threatens infinite regress. Since metarepresentation is itself a thought, if I want to attribute this metarepresentation to myself, I would need to monitor it with a meta-metarepresentation. I would have to ask, in a further metarepresentational thought: is this my metarepresentation? An extra level of consciousness would be added to the comparator's verification process. My intention to metarepresent would have to generate its own efference copy, to be matched up on top of the original match. As I will suggest below, in the case of thinking, which is already prereflectively conscious, even one level of metarepresentation for maintaining a sense of agency, or even one level of efference copy for purposes of keeping track of thoughts is redundant.

Not only does the need for metarepresentation in the normal case not seem phenomenologically parsimonious, but in regard to the claim that metarepresentation fails in the case of schizophrenia, it runs counter to some clinical accounts. In contrast to Frith's notion that there is a lack of metarepresentational self-monitoring in schizophrenia, some clinicians suggest that there may be too much of it. In principle, it seems possible for metarepresentation to go wrong in at least two ways. First, as Frith suggests, it might fail in such a way that the schizophrenic can be left without the ability to monitor his own experience. Second, however, as Sass [28–30] suggests, metarepresentation can be generated in a hyperreflexive attitude, and as a result, the schizophrenic can overmonitor aspects of his own experience. It also seems possible that the disruption of metacognition could be selectively Frithian and Sassian. As a result of a failure to monitor certain aspects of his experience, the schizophrenic may enter a hyperreflexive attitude about what is absent from or odd about his experience. A schizophrenic may have great difficulties with attention, not because of a complete lack of attention, but because he is attending in a high degree to certain aspects of his experience that are different. The failure of self-monitoring may in some instances be that there is too much of it going on [9, 29]. Exactly how we should think of the role of higher-order cognition (whether it fails due to a lack of metarepresentation, or due to a surplus of hyperreflexivity) is still uncertain. We return to this point below.

The Episodic Nature of Positive Symptoms

If, in the case of schizophrenia, one of the comparator mechanisms described by Frith breaks down or is put out of operation, why do not all thoughts seem alien? If either efference copy fails to generate with the initiation of thought, or the central monitor fails to register efference copy, how do we explain that these mechanisms do seem to work normally sometimes, since not all of the schizophrenic's thoughts are experienced as inserted thoughts? I will refer to this as the *selectivity problem* or the *problem of the episodic nature of positive symptoms*. When a particular thought seems inserted, there are other aspects of conscious experience that do not seem inserted. Either simultaneously with, or immediately after the thought that seems inserted, there is the sense or feeling or realization that it is inserted. But this sense or feeling or realization does not itself feel inserted. That is, the subject, in recognizing a thought as inserted, does not claim that his recognition of this fact is also inserted. He is rather speaking in his own voice when he complains about the inserted thought. That this is the case is clear, not only from empirical reports by patients, but by logical necessity. The subject's complaint that various thoughts are inserted depends on a necessary contrast between thoughts that seem inserted and those that do not seem inserted – and at a minimum, the thoughts that constitute the subject's complaint cannot seem inserted. If all thoughts were experienced as inserted by others, the subject would not be able to complain 'in his own voice', so to speak. He would not maintain a sufficient sense of ownership for his cognitive life, or a sense of cognitive 'space' by which to define an insertion.

This selectivity problem cannot be explained by a failure of a comparator, since such a failure should also affect the sense of recognition that the thought is inserted. A theory that would credit a higher-order cognitive process (a metarepresentation, or an evaluative introspective judgment) founders on this same problem. It would need to explain why a higher-order cognition that fails to generate a sense of agency for a particular thought or experience is itself experienced as self-agentive. Why is it that a subject can have a sense of agency for one thought, but not for the other? Quite obviously the phenomenology here needs to constrain the cognitive explanation.

The Problem of Specificity

A further problem of Frith's model involves the specificity of positive symptoms. In this regard, in cases of

thought insertion, specific kinds of thought contents, but not all kinds appear to be inserted. I suggest that no explanation that remains totally on the subpersonal level will be able to explain this specificity. It is not simply that in experiences of thought insertion patients only occasionally experience thoughts coming into their minds from an outside source. Rather, their experiences are very specific, and are sometimes associated with specific others. For example, a schizophrenic may report that thoughts are being inserted by a particular person, and that they are always about a specified topic. In auditory hallucination, the voice always seems to say the same sort of thing. Such specificity phenomena seem to have a semantic and experiential consistency and a personal-level complexity that cannot be adequately explained by the disruption of subpersonal mechanisms alone.

Global Problems

A more general problem with Frith's analysis concerns the global nature of schizophrenia and its various symptoms. Schizophrenics have problems, not only with movement and self-reference, but also, among other things, with working memory, episodic and autobiographical memory, and narrative construction [31]. Is it possible to reduce all of these problems to one central difficulty with self-monitoring, or to one comparator mechanism which is neurologically ill-defined? If the subject experiences both delusions of control and inserted thoughts, does this mean that two separate comparators – one in the motor system and one in the conscious thought system – malfunction independently? Or is the implication that one comparator covers both movement and thought? Or is something more global involved? At best, Frith proposes similar mechanisms for motor action and cognition, but does not clarify how they might be related, or whether the embodied processes of motor control and cognition are connected.

A Neurophenomenological Approach

I do not propose to solve all of the problems that I have outlined above. I do want to suggest, however, that we can begin to address some of these problems by clarifying the different levels of explanation involved, and different explanatory approaches. The following analysis assumes that there are distinctions between (1) first-order phenomenal experience, (2) higher-order cognition (which supports the ability to make attributive judgments about one's own experience) and (3) nonconscious, subpersonal processes that are best described as neuronal or brain processes. There may be some intermediate levels of description (syntactical or representational) that are also understood as nonconscious, but for purposes of this paper I leave this possibility aside. Also, I do not deny that there may be unconscious mental states, but I will also leave this idea aside. I also assume that experience is not simply cognitive, but also emotional and embodied. Is it possible to explain positive symptoms of schizophrenia, like thought insertion and delusions of control, within the phenomenological- neurological framework defined by these assumptions?

There are two general and basic approaches to explain how these different levels relate to one another. A 'topdown' approach argues that higher-order cognition holds explanatory power, and perhaps has causal power, for what happens (or seems to happen) in first-order phenomenology. A 'bottom-up' approach argues that neurological processes contribute in important ways to the shaping of first-order phenomenal experience, and that second-order cognition has reportive and attributive capabilities that work in either a veridical or nonveridical way. It is also possible to combine these approaches to form a hybrid explanation, as one finds in Frith. The contrast in approaches can be seen in how one might explain the distinction between the sense of ownership and the sense of agency. Stephens and Graham [3] take a top-down approach. They explain the distinction as a production of higher-order cognition that is read into the level of phenomenal consciousness. They are concerned primarily with the attributions of ownership and agency. Thus, what Stephens and Graham [3] characterize as a sense of agency that is 'a normal component or strand in our experience of thinking', but 'normally phenomenologically intertwined with introspective awareness as well as with the sense of subjectivity' [p. 9] turns out, in their view, to be 'constituted by our self-referential narratives or conceptions of our underlying intentional states' [p. 183]. In contrast, a bottom-up approach characterizes the sense of agency as belonging to a first-order phenomenology; it involves a prereflective (nonconceptual) sense that is implicit to experience, and generated by specific neurological processes.

It is possible that many important aspects of human experience are best explained by a top-down approach (indeed, the most rational of them may be the best candidates for such explanation). Following the critical analysis presented above, however, I want to suggest that schizophrenic experience is not one of them. There is, however, something to be said in support of hybrid approaches.

First, even the most top-down explanation should not ignore the contribution of neuronal processes, since the very possibility of attributive introspection or metarepresentation relies on them. Second, even the most bottom-up explanation should not ignore the complications introduced by personal-level processes like introspection. In working out a hybrid account, however, one must still determine whether it is best to start at the bottom or at the top. I propose that the best approach to schizophrenia is a hybrid one that starts by pursuing a bottom-up neurophenomenological analysis. A successful account of the positive symptoms of delusions of control and thought insertion needs to explain 6 things.

- 1 The neurological failure to generate a sense of agency.
- 2 The lack of a sense of agency in first-order phenomenal experience.
- 3 The lack of a higher-order attribution of self-agency.
- 4 The fact that the subject misattributes the agency to another person (or thing).
- 5 The selective or episodic nature of positive symptoms.
- 6 The specificity of delusional and inserted thoughts.

In what starts as a bottom-up approach, an account of 1 helps to explain 2, and accounts of 1 and 2 help to explain 3 and 4. The majority of the explanation has to be worked out at the neurological and phenomenological levels, where the latter involves an account of what it is like for the subject – an analysis of the basic first-order phenomenal experience involved in the senses of ownership and agency or their disruption. Once these analyses are in place, what happens at the cognitive level of self-attribution (the attribution of ownership but not of agency and the misattribution of agency to another) is straight-forwardly explained as an effect of the lower-order disruptions. 5 and 6, however, present special problems that may require personal-level explanations.

Problems with self-agency that manifest themselves as thought insertion and delusions of control are generated on a neurological level. The neurological picture is complex, but recent results of brain imaging studies suggest the importance of two neuronal areas in generating a sense of agency for movement.

Farrer and Frith [6] have shown contrasting activation in the right inferior parietal cortex for perception of action caused by others, and in the anterior insula bilaterally when action is experienced as caused by oneself. One possible explanation for the involvement of the right inferior parietal cortex in the discrimination between self-agency and other-agency is suggested by Jeannerod [32]. Namely, actions performed by others are perceptually mapped in allocentric coordinates. Farrer and Frith [6] note that ‘there is strong physiological evidence that the inferior parietal cortex [involves this] kind of remapping ... to generate representations of body movements in allocentric coordinates [p. 601].

In contrast, they suggest, the role of the anterior insula in providing a sense of self-agency involves the integration of three kinds of signals generated in self-movement: somatosensory signals (sensory feedback from bodily movement, e.g. proprioception), visual and auditory signals, and efference copy associated with motor commands that control movement. ‘A close correspondence between all these signals helps to give us a sense of agency’ [p. 602].

Neuroimaging studies by Decety et al. [33] and Chaminade and Decety [34] provide more evidence for the involvement of the inferior parietal lobule in discriminating agency. They have shown more activation in the left inferior parietal lobule when a subject imitates another person (and thus has an appropriate sense of agency for such actions) compared to more activation in the right inferior parietal lobule when the other person imitates the subject. They conclude that activity in the inferior parietal lobule may be critical for distinguishing actions generated by the self from those initiated by others.

Other studies show that lesions in the right parietal lobe can lead to difficulties in attributing actions. In psychiatric and neurological patients, self-awareness disorders have been linked to metabolic abnormalities in this brain area. In schizophrenic patients, the feeling of alien control (delusions of control) during a movement task has been associated with an increased activity in the right inferior parietal lobe [35].

Of course, things are likely more complicated than this, in both normal and schizophrenic experience. As suggested in the Frithian analysis, the sense of agency for motor action may depend on a preaction, forward motor control mechanism that matches motor intention and efference copy of the motor command. The proper timing of such a match may depend on the proper functioning of the supplementary motor area, the premotor, and prefrontal cortexes, and such functions are known to be disrupted in schizophrenic subjects with delusions of control [15, 18, 36–38]. For reasons outlined above, however, mechanisms that explain motor action may not be the mechanisms that best explain disruptions in the sense of agency for cognition. Indeed, there may be a more general or basic disruption of neuronal processes that affect not just the sense of agency for motor action, but disrupt the sense of agency for cognitive experience (e.g. thought insertion). The sense of agency for thought may depend on the anticipatory aspect of working memory, which is manifested

in the prereflective structure of phenomenal experience [39]. Just such problems with working memory are to be found in schizophrenic subjects with delusions of control [40– 43].

Clearly, both the supposition of neuroscience and the most reasonable explanations of how experience is generated suggest that what phenomenal experience is like depends (at least in part) on the proper functioning of the brain. In the case of the schizophrenic who suffers from delusions of control and thought insertion, neurological problems generate a first-order experience that does not include a sense of agency for certain actions and thoughts. Although it is important to sort out the specific nature of the neurological problem, for purposes of this essay, it is sufficient to say that whatever the precise nature of the neurological disruptions responsible for delusions of control or thought insertion is, some such processes clearly generate a first-order phenomenal experience that lacks a sense of agency.

If this in fact is an accurate characterization of schizophrenic experience at the level of first-order phenomenal consciousness, then in some respects it seems clear that the subject's second-order report on such experience is not mistaken. The subject correctly reports his experience, that he experiences himself as the owner of, but not the agent of the movement or thought; he correctly attributes ownership but not agency for the thought. This is not a mistake or misinterpretation, as suggested by Stephens and Graham [3], but a report of what the subject actually experiences. The experience, summarized in a first-person, albeit abstract way, is this: 'I am not the agent of my movement or thought.' Even if we are reticent to accept the schizophrenic's own reports of such experience, of which there are many, behavioral studies support this interpretation [19, 44]. But, of course, this is not a complete statement, for the patient goes on to say that 'someone else is controlling my movement or thought'. If the only problem were a lack of agency, then there would be no difference between the phenomenology of passive, involuntary movement, or unbidden thoughts, and schizophrenic delusions. The explicit delusional component involves the subject's misattribution of agency to another person (or machine, or thing). In this regard, pace Frith, a lack of a sense of self-agency does not add up to an attribution of action to others. Here there appear to be two possible explanations of the misattribution.

(a) The cause of the misattribution is based on inferences made at the higher-order level of attributive consciousness.

According to a top-down account, of the sort found in Stephens and Graham [3], in normal, nonpathological experience, our cognitive, emotional, and motivational lives are relatively integrated, and thoughts and actions seem to belong to us when they fit into this integrated whole. In schizophrenia, however, this integration is disrupted, to the extent that thoughts and actions come to be disjointed, the unity of the self breaks down, and one begins to feel alienated from one's thoughts and actions. The cause of this breakdown is seemingly the introspection itself, which is either excessive or pathological. In this case, the subject introspectively misinterprets his experience as caused by someone else. The lack of a sense of agency is filled in by a productive narrative that projects into first-order experience the idea that someone or something else must be causing this action.

(b) Some neurological component responsible for the differentiation between self and other is disrupted, and as a result, some sense of alterity is already implicit in the first-order experience.

In this case, the attribution of agency to another is not the result of a misinterpretation or cognitive accounting that would force the subject to infer that since he is not the agent, someone else must be, or a supplemental account generated by introspection, the odd result of a productive narrative; rather, it is a genuine report of what is truly experienced. This is not to rule out the fact that odd, paradoxical, and wildly delusional narratives are often generated as the illness develops. The initial motivation for such narratives, however, may be shaped by processes that start out as completely rational at the second-order level – that is, a completely correct report of what the subject experiences.

Here I will cite only one piece of evidence for this view. Recent brain imaging studies (using PET or fMRI) show activation of a partially overlapping cortical and subcortical network that includes structures directly concerned with motor execution (motor cortex, dorsal and ventral premotor cortex, lateral cerebellum, basal ganglia) and areas concerned with action planning (dorsolateral prefrontal cortex and posterior parietal cortex) in the following situations: during my action, during the observation of another person's action, and during the imaginary simulation of my own or another's action [32, 45, 46]. The brain areas activated when I engage in specific intentional actions are in large part the same brain areas that are activated when I observe someone else engaged in the same activity [47]. A number of researchers suggest that just such overlapping or shared representations may play some part in our ability to simulate the thoughts and attitudes of

others [48–51]. This suggests that if something goes wrong with this neuronal ‘who’ system [15], our own movement, or even our own thoughts, may be experienced as initiated by someone else. There is good evidence that this is what happens in some schizophrenic patients [52].

These observations suggest that whatever higher-order cognition is like in schizophrenia (whether it is *hypo* or *hyper*), it is not led solely on its own (that is, by its own mistaken inferences or misinterpretations or corrupt theories) to misattributions of agency. Rather, the ipseity [9], which includes the sense of agency that is normally implicit in any intentional act, is disrupted at the first-order phenomenal level, and the effects of this disruption become an object of reflective introspection. This is the point at which to evoke hyperreflexivity – not as a higher-order cognition that disrupts ipseity and generates alterity, but as a reflection that discovers an alterity that is already manifesting itself on the first-order phenomenal level, generated by neuronal processes that are abnormal.

Two problems remain: the selectivity and the specificity of delusional and inserted thoughts. The selectivity problem may simply depend on the neurological disruptions being occasional. Delusional states may be generated when the shared representations that allow for a discrimination between my movement and someone else’s movement, or my thought and my simulation of someone else’s thought, are disturbed. This may be infrequent early in the illness, and more frequent as the illness develops.

The specificity problem, that is, the fact that the schizophrenic delusions refer to specific people or things or situations or content, is likely more complex. In part, it may be simply an accident of when and in what context the shared representations are disrupted, and who happens to be around. In part, it may depend on the subject’s own personal history, emotional attitudes and states, and relations with others [4, 8]. And in part, it may be intensified by an introspection that shades off into confabulation.

Such considerations do not amount to a complete explanation of schizophrenia, or even of the positive symptoms that we have been discussing. There is much more to be said about the nature and temporal structure (including an anticipatory structure) of first-order, prereflective self-consciousness, the *ipseity* that gets disrupted by neurological processes that dysfunction [4, 8, 39]. There is also much more to say about the working of higher-order cognition, and the generation of peculiarly schizophrenic narratives [31]. My claim, however, is that the framework provided by this neurophenomenological approach can sustain these other analyses better than a framework designed around a purely cognitive or a Frithian neurocognitive approach.

References

- 1 Frith CD: The Cognitive Neuropsychology of Schizophrenia. Hillsdale, Erlbaum, 1992.
- 2 Campbell J: Immunity to error through misidentification and the meaning of a referring term. *Philos Top* 1999;26:89–104.
- 3 Stephens GL, Graham G: When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts. Cambridge, MIT Press, 2000.
- 4 Gallagher S, Varela F: Redrawing the map and resetting the time: Phenomenology and the cognitive sciences. *Can J Philos*, in press.
- 5 Varela FJ: Neurophenomenology: A methodological remedy for the hard problem. *J Conscious Stud* 1996;3:330–349.
- 6 Farrer C, Frith CD: Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *Neuroimage* 2001;15:596–603.
- 7 Gallagher S: Philosophical conceptions of the self: Implications for cognitive science. *Trends Cogn Sci* 2000;4:14–21.
- 8 Gallagher S: Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification; in Zahavi D (ed): *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience*. Amsterdam, Benjamins, 2000, pp 203–239.
- 9 Zahavi D, Parnas J: Phenomenal consciousness and self-awareness: A phenomenological critique of representational theory. *J Conscious Stud* 1998;5:687–705.
- 10 Frankfurt H: Identification and externality, in Rorty AO (ed): *The Identities of Persons*. Berkeley, University of California Press, 1976, pp 239–251.
- 11 Graham G, Stephens GL: Mind and mine; in Graham G, Stephens GL (eds): *Philosophical Psychopathology*. Cambridge, MIT Press, 1994, pp 91–109.
- 12 Held R: Exposure-history as a factor in maintaining stability of perception and coordination. *J Nerv Ment Dis* 1961;132:26–32.
- 13 Wolpert DM, Ghahramani Z, Jordan MI: An internal model for sensorimotor integration. *Science* 1995;269:1880–1882.
- 14 Frith CD, Blakemore S-J, Wolpert DM: Abnormalities in the awareness and control of action. *Philos Trans R Soc Lond B Biol Sci* 2000;355: 1771–1788.
- 15 Georgieff N, Jeannerod M: Beyond consciousness of external events: A ‘who’ system for consciousness of action and self-consciousness. *Conscious Cogn* 1998;7:465–477.
- 16 Jeannerod M: The representing brain: Neural correlates of motor intention and imagery. *Behav Brain Sci* 1994;17:187–245.
- 17 Marcel AJ: The sense of agency: Awareness and ownership of actions and intentions; in Roessler J, Eilan N (eds): *Agency and Self-Awareness*. Oxford, Oxford University Press, in press.
- 18 Malenka RC, Angel RW, Hampton B, Berger PA: Impaired central error correcting behaviour in schizophrenia. *Arch Gen Psychiatry* 1982;39:101–107.
- 19 Frith CD, Done DJ: Towards a neuropsychology of schizophrenia. *Br J Psychiatry* 1988;153: 437–443.

- 20 Blakemore S-J: Monitoring the self in schizophrenia: The role of internal models; in Zahavi D (ed): Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience. Amsterdam, Benjamins, 2000, pp 185– 202.
- 21 Sperry RW: Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol* 1950;43:482– 489.
- 22 von Holst E, Mittelstaedt H: Das Reafferenzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* 1950;37:464–476.
- 23 Feinberg I: Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophr Bull* 1978;4:636–640.
- 24 Akins KA, Dennett D: Who may I say is calling? *Behav Brain Sci* 1986;9:517–518.
- 25 Hoffman R: Verbal hallucinations and language production processes in schizophrenia. *Behav Brain Sci* 1986;9:503–517.
- 26 Campbell J: Schizophrenia, the space of reasons and thinking as a motor process. *Monist* 1999;82:609–625.
- 27 Gallagher S: Sense of agency and higher-order cognition: Levels of explanation for schizophrenia. *Cogn Semiotics*, in press.
- 28 Sass L: Schizophrenia, self-experience, and the so-called negative symptoms; in Zahavi D (ed): Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience. Amsterdam, Benjamins, 2000, pp 149–182.
- 29 Sass L: Schizophrenia, self-consciousness and the modern mind. *J Conscious Stud* 1998;5: 543–565.
- 30 Sass L: *Madness and Modernism: Insanity in the Light of Modern Art, Literature, and Thought*. New York, Basic Books, 1992.
- 31 Gallagher S: Self-narrative in schizophrenia; in David AS, Kircher T (eds): *The Self in Neuroscience and Psychiatry*. Cambridge, Cambridge University Press, 2003, pp 336–357.
- 32 Jeannerod M: To act or not to act: perspectives on the representation of actions. *Q J Exp Psychol A* 1999;52:1–29.
- 33 Decety J, Chaminade T, Grezes J, Meltzoff AN: A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage* 2002;15:265–272.
- 34 Chaminade T, Decety J: Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport* 2002;13:1975–1978.
- 35 Spence SA, Brooks DJ, Hirsch SR, et al: A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain* 1997;120: 1997–2011.
- 36 Fournieret P, Jeannerod M: Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* 1998;36:1133– 1140.
- 37 Haggard P, Eimer M: On the relation between brain potentials and the awareness of voluntary movements. *Exp Brain Res* 1999;126:128– 133.
- 38 Haggard P, Magno E: Localising awareness of action with transcranial magnetic stimulation. *Exp Brain Res* 1999;127:102–107.
- 39 Gallagher S: *How the Body Shapes the Mind*. Oxford, Oxford University Press, in press.
- 40 Daprati E, Franck N, Georgieff N, Proust J, Pacherie E, Dalery J, Jeannerod M: Looking for the agent: An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition* 1997;65:71–96.
- 41 Franck N, Farrer C, Georgieff N, Marie-Cardine M, Daléry J, d'Amato T, Jeannerod M: Defective recognition of one's own actions in patients with schizophrenia. *Am J Psychiatry* 2001;158:454–459.
- 42 Singh JR, Knight T, Rosenlicht N, Kotun JM, Beckley DJ, Woods DL: Abnormal premovement brain potentials in schizophrenia. *Schizophr Res* 1992;8:31–41.
- 43 Vogeley K, Kurthen M, Falkai P, Maier W: The human self construct and prefrontal cortex in schizophrenia. Association for the Scientific Study of Consciousness. Electronic Seminar 1999 (<http://www.phil.vt.edu/assc/esem.html>).
- 44 Posada A, Franck N, Georgieff N, Jeannerod M: Anticipating incoming events: An impaired cognitive process in schizophrenia. *Cognition* 2001;81:209–225.
- 45 Decety J, Grezes J, Costes N, Perani D, Jeannerod M, Procyk E, Grassi F, Fazio F: Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain* 1997;120:1763–1777.
- 46 Grezes J, Decety J: Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Hum Brain Mapp* 2001;12:1–19.
- 47 Jeannerod M, Frak V: Mental imaging of motor activity in humans. *Curr Opin Neurobiol* 1999;9:735–739.
- 48 Blakemore S-J, Decety J: From the perception of action to the understanding of intention. *Nat Rev Neurosci* 2001;2:561–567.
- 49 Chaminade T, Meary D, Orliaguet J-P, Decety J: Is perceptual anticipation a motor simulation? A PET study. *Neuroreport* 2001;12: 3669–3674.
- 50 Decety J: Naturaliser l'empathie. *L'Encéphale* 2002;28:9–20.
- 51 Jeannerod M: Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage* 2001;14:S103–S109.
- 52 Jeannerod M, Farrer C, Franck N, Fournieret P, Posada A, Daprati E, Georgieff N: Action recognition in normal and schizophrenic subjects; in Kircher T, David A (eds): *The Self and Schizophrenia*. Cambridge, Cambridge University Press, 2003.