

Cognitive Neuropsychiatry 8: 237–242, 2003.

Book review

When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts

By G. Lynn Stephens and George Graham.

(Cambridge, Massachusetts: MIT Press 2000. Pp *xii* + 198)

Dr. Jakob Hohwy

Department of philosophy

University of Aarhus

Denmark

filhohwy@hum.au.dk

The field of philosophical psychopathology is basically the philosophical study of mental disorders such as schizophrenia, bipolar disorder, depression, autism, as well as more specific symptoms and signs such as Capgras' delusion (the delusion that your spouse, for example, is an impostor) or the anarchic hand sign (where your hand seems to act on its own intentions). This simple epithet covers a multitude of approaches: how can philosophy help to explain mental disorder? What does mental disorder tell us about consciousness, cognition, emotion and 'self'? What does the study of mental disorder tell us about phenomenology? What does philosophical phenomenology tell us about mental disorder? What do mental disorders tell us about reasoning, rationality and belief formation? What are the particular ethical aspects of mental disorder and its treatment? If philosophical

psychopathology can lead to interesting answers to some of these questions it deserves a place among recognised philosophical disciplines. Similarly, one can hope that this will be an area where philosophy can be of relevance to ongoing scientific research.

Though discussion of mental disorders is not unknown in the philosophical tradition (one example is Maurice Merleau-Ponty), it is only within the last decade or so that a recognisable and sizeable field has emerged. One indication of this is the increasing number of research articles and books concerned with philosophical psychopathology, another is the news that textbooks and handbooks are forthcoming (e.g., George Graham is editing a companion to philosophy of psychiatry for Blackwells and Philip Gerrans is writing on cognitive neuropsychiatry for MIT Press). One reason for this surge in interest is no doubt the booming field of cognitive neuroscience, cognitive neuropsychology and cognitive neuropsychiatry, which affords masses of new data and models for thinking about mental disorders. G. Lynn Stephens and George Graham (LSG) are among the pioneers in this field, and it is very much through their work that it has become respectable for philosophers of all stripes to, as it were, incorporate mental disorder into their thinking.

Philosophical psychopathology is apt to be an interdisciplinary enterprise. So ideally, a researcher doing philosophical psychopathology masters contemporary philosophy of mind, cognitive neuroscience, clinical psychopathology and has a keen sense of phenomenology. *When self-consciousness breaks* comes close to the ideal. It focuses on two psychiatric symptoms, associated primarily with schizophrenia, viz. the Schneiderian symptoms of auditory hallucinations and thought insertion. In auditory hallucinations the subject hears voices that often, but not always, are critical of her actions. In thought insertion, the subject has thoughts that she attributes to other people in the sense that she thinks, bizarrely, that others are thinking the thought in her mind.

LSG conceive of these mental disorders as examples of alienated self-consciousness. They argue that, quite apart from their cognitive and neuroscientific underpinnings (p.158), these alien episodes are best understood as caused by disturbances in the sense of *agency* (not the sense of subjectivity).

After describing the symptoms, LSG proceed to a lengthy discussion of Ralph Hoffman's influential theory that auditory hallucinations arise because inner speech has a voice-like character, and that hallucinating subjects confuse inner speech with someone else's speech (chapters 4-5). I will not say anything about their interesting criticism of this proposal here, instead I jump to their discussion of Chris Frith's influential theory of verbal hallucinations and thought insertion (chapter 6) (Frith 1992), and after that I discuss their own positive proposal.

LSG agree with Frith's three initial claims: (i) some verbal hallucinations have no sensory component; that is, they are not distinctly auditory; (ii) there is a close affinity between verbal hallucinations and passivity experiences such as thought insertion: they involve a failure to recognise that some activity is self-initiated; (iii) both inner speech and thought have action-like features inasmuch as we experience them as intended and associated with a sense of effort.

But LSG disagree with the theory Frith proposes on basis of the three claims. Frith begins with the plausible claim that we need the ability to distinguish between changes in the environment that are due to ourselves and changes that are due to others (I need to know whether the deep pressure on my back is due to someone prodding me, or me leaning against a wall). The implication is that there must be some system for monitoring our actions. A failure to monitor inner speech would lead to

attribution to the external environment; in this case, other speakers, even when there are none. This explains auditory hallucinations.

But it is also important for the individual to be able to distinguish actions at another level. Some of my actions are stimulus-driven (a ball comes towards me and I catch it), other actions are willed in the sense that they are associated with my longer-term desires (I pick the carrot over the ice cream because I have a certain long term desire for being fit and healthy, say). So there must be a monitoring system for stimulus intentions and willed intentions too. This system could be based on a principle concerning efference copies borrowed from oculomotor theory (when we move our eyes we know that it is we and not the world that is moving because copies of the motor commands to the muscles controlling the eye are dispatched to a 'monitor'). In the 1992 version of Frith's theory, efference copies of our willed intentions are sent to the monitor which then can tell whether this was a stimulus-driven or a willed action.

Frith's theory is that a defect in the monitoring system of willed intentions would lead the subject to experience actions, such as thinking a given thought, as unintended and hence develop passivity experiences such as thought insertion and delusions of control (another Schneiderian symptom where the subject's bodily movements are attributed to another person's agency).

This theory seems plausible for delusions of control, and has been empirically supported in an experiment that showed that people with schizophrenia are poor at rapid error correction (Frith and Done 1989). I think it is much less straightforward to transpose the theory from the motor control area to the area of conscious thought. One major problem is that we do not seem to have intentions

to think thoughts before we think them – having the intention to think the thought *P* already involves thinking *P*.

LSG has other reasons to reject Frith's theory. Before presenting their arguments I shall briefly mention two worries about their discussion. First, they reject the theory without really accounting for the empirical evidence in its favour. Second (and understandably in the light of the publication date), they reject an out-dated version of the monitoring theory. After advances in computational motor control theory, it is the efference copy and not the intention which now takes centre place (in later iterations of Frith's theory; see, e.g., Blakemore et al 2002).

The first argument against Frith is this (p.141): If the role of the intention monitor is to distinguish willed action from stimulus-driven action, why don't subjects just conclude that the thought in question was stimulus driven, rather than jumping to the bizarre conclusion that someone thought the thought in them? Here it is worth thinking about what a stimulus-driven thought might be. Presumably the kind of thought that occurs as a result of the influence of others (consider for example your non-pathological thought of ice-cream after seeing an ice-cream commercial). However, no sharp divide seems possible between the thoughts that are stimulus driven and willed (think of the thoughts you have after hearing an inspiring lecture on your favourite subject – are they stimulus-driven or willed?) Nevertheless, this objection is important because there is good evidence that inserted thoughts are *not* experienced as thoughts that are influenced by others.

However, the objection overlooks the detail of Frith's original proposal (1992, p.82) according to which the monitor receives efference copies of *stimulus* intentions also. If the monitor detects

nothing at all this means that there is no stimulus-driven action either. In that case it is not open for the subjects to conclude that the action was stimulus driven.

LSG add to their objection that Frith has no story to explain why the errant thought is experienced as *alien*, rather than just a thought influenced by someone else, in the sense explained above (p.142). The answer here must be that it is not experienced as influenced by others because it is not experienced as stimulus-driven, *pace* the response to their initial objection. And if the errant thought is neither experienced as willed nor as stimulus-driven, then things are so weird that it seems a fair bet that it is alien.

In the light of this, it is worth re-appraising Frith's theory. As I mentioned above, we need to attend to the phenomenology and logic of thought and intention. But we also need to consider issues concerning experience and rationality. Do delusions of thought insertion arise as rational or as irrational responses to unusual experiences (thoughts without owners), or do they perhaps arise in one step, as a result of an experience of a thought *as* thought by someone else (for a review and analysis of this crucial debate, see Coltheart and Davies 2000)?

LSG's positive proposal of alienated self-consciousness is set out in Chapters 7 and 8. It begins with the important observation that people who suffer from verbal hallucinations or thought insertion do not seem to have problems with the sense of subjectivity, or with their ego boundary. They know that the experiences occur in their own minds. But this gives rise to a conceptual problem. How can a subject undergoing delusions of thought insertion or verbal hallucination maintain that a mental episode is another's thought while acknowledging that it occurs in her own mind (p. 146)? That is, how can something acknowledged to be mine not be mine? LSG's answer

basically lies in disambiguating the notion of ‘mine’ so that it means ‘subjectively mine’ in one sense and ‘agentually mine’ in another. Their main task is then to make this move plausible. I shall present three stages of their argument, and then raise some problems for it.

First, they invoke Harry Frankfurt’s notion of identifying oneself with one’s thoughts (chapter 7). Sometimes I refrain from attributing my bodily movements to my own agency. That is, I do not identify with them. This happens for example when a muscle suddenly twitches. Likewise, some thoughts are not thoughts that *I think*, I just *find* them occurring in me. This happens, for example, when a jingle keeps popping up in my mind, or when I feel a satisfying *schadenfreude* even though I don’t see myself as the kind of person who rejoices in other’s misfortune. So I may have a thought that I do not identify with in the sense that I do not think it is *my* doing. I may, as it were, disavow the thought if it doesn’t fit with who I am conscious of being.

Second, LSG flesh this out in terms that derive from, for example, Dennett’s notion of the narrative self – the idea that the agent’s self is constituted by her theory or conception of her internal psychology (p. 162). The self is not some kind of Cartesian point, it is the tale we spin about our beliefs, desires and intentions (or rather the tale that spins us, as Dennett puts it (1991)). There is denial of agency, then, when the occurrence of a thought cannot be explained in terms of the agent’s theory of ‘self’ (p. 165). Importantly, LSG are neutral on what may cause this inexplicability, and it could therefore well be a Frithian deficit in monitoring.

Third, LSG address the objection that this doesn’t explain why the subject believes that the thoughts are performed by *someone else*. Why not simply form the belief that ‘I was off-line for a moment there’? Their answer is that the contents of the delusions and hallucinations are experienced as

intentional – they are the intentional type of thing – and that therefore it is unavoidable that the subject expects an agent to be behind them (the contrast might be the experience of seemingly random whirrs and clicks). The subject is therefore presented with a choice to *either* revise their theory of self *or* to form the belief that someone else is doing it. If it is too hard, or uncomfortable, or impossible for other (perhaps neurophysiological) reasons to revise the self-theory, then there is only the latter option left, resulting in alienated self-consciousness (pp. 164-5)

The first worry I have about this is that it may rest on a false dichotomy. Sometimes, when really weird things are observed to happen, it may be rational to neither revise the theory nor blame the observation. It may be rational to do nothing. Sit on it and go on with your business until everyone forgets or something else happens. Surely this is a better option than jumping to the bizarre belief that my thoughts are under someone else's control.

The second worry concerns some results from studies of delusions of control that are not predicted by the LSG account. Delusions of control are highly relevant for their account given their insistence on a very close analogy between the agential aspects of bodily movements and mental episodes. In one study (Spence et al 1997), schizophrenic subjects were asked to perform some joystick movements, they then formed the intention to move the joystick, and acted on the intention. During the movement they experienced delusions of control. This case is important because it demonstrates that in delusions of control the subjects know what they intend: delusions of control are not like the anarchic hand sign where the subjects do not know what the hand intends to do (Frith et al 2000).

The crucial point is that it is implausible that something I know I intend is deemed inexplicable in terms of my theory of self. Surely the theory of my beliefs, desires and intentions that constitute my

self incorporate the intentions I know I have. But if that is the case then the actions I perform as a result of having those intentions *are* explicable in terms of the theory, and yet, in the experiment described above, the agency was ascribed to someone else. *Mutatis mutandis* for the case of thought insertion and verbal hallucination (though the corresponding experiment is very hard to perform for inserted thoughts; see Cahill and Frith 1996 for experiments pertaining to hallucinations). If this is right, then LSG's narrative self story of alienated self-consciousness seems wrong.

In my view, we should respond to this situation by going back to the monitoring theories and re-cast them in the light of the recent work on the motor control system and abnormalities of agency and control, such as delusions of control. Importantly, however, this re-casting must be made in the light of the phenomenological and conceptual differences between thought and action.

Despite my critical remarks about Lynn Stephens and Graham's positive proposal I want to emphasise that their contribution to philosophical psychopathology appears to me to be very important: it criticises various models of self-consciousness and refines our sense of the phenomenology of self, using the best of contemporary philosophy of mind and a keen understanding of pathological conditions, advancing our conception of these issues considerably. It is thus an exemplary case of philosophical psychopathology.

References:

Blakemore, S.-J., Wolpert, D., Frith, C. 2002. Abnormalities in the awareness of action. *Trends in Cognitive Sciences* 6: 237-242.

- Cahill, C. and Frith, C. 1996. False perceptions or false beliefs? Hallucinations and delusions in schizophrenia. In *Methods in Madness*, P. W. Halligan & J. C. Marshall (eds.), Hove: Psychology Press, pp. 267-292.
- Coltheart, M. and Davies, M. 2000. Introduction: Pathologies of Belief. *Mind & Language* 15: 1-46
- Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.
- Frith, C. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hove, East Sussex: Lawrence Erlbaum Ass.
- Frith, C. D. & Done, D. J. 1989. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological medicine* 19: 359-363.
- Frith, C. D., Blakemore, S.-J., Wolpert, D. 2000. Abnormalities in the awareness and control of action. *Phil.Trans. R. Soc. Lond. B* 355: 1771-1788.
- Spence *et al.* 'A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control)', *Brain* 120 (1997), pp. 1997-2011