

G. Lynn Stephens & George Graham, *When Self-Consciousness Breaks – Alien Voices and Inserted Thoughts*. Cambridge: MIT Press, 2000, 198 pp.

In this book, Stephens and Graham search for an explanation of verbal hallucination and thought insertion. In such cases, a subject experiences an episode of his or her mental life as attributable to another person, hence the terms ‘alien’ and ‘inserted’. The major part of the work consists in the examination of hypotheses. Let me summarize the main work done.

Two principal hypotheses are forwarded. The first is that both phenomena are to be conceptualised as a loss of ego boundaries or as an internal/external confusion. This hypothesis is rejected because the subjects in question recognize that they are the subject *in* whom those episodes occur. What is lacking is a sense of agency, as distinguished from the sense of *subjectivity*. What remains to be explained is that somebody else seems to be doing the thinking in the subject’s head.

The second hypothesis is that the self produces the message but misattributes the voice. Reference to self-conversation or inner speech helps to explain the verbal quality of a voice. This hypothesis can be elaborated in three ways.

- (a) The subject talks silently to himself. This hypothesis is rejected, because a ‘genuine theory’ need not to suppose private events.
- (b) Therefore, the speech can be subvocal and not naturally audible (whisper hypothesis). But subjects experiencing verbal hallucinations do not have the impression that they are *hearing* voices. Moreover, deaf subjects can experience verbal hallucinations. Therefore, the whisper hypothesis seems doubtful.
- (c) A third possibility is that the subject receives information other than hearing a voice, and that voices are *ex post facto* explanations or confabulations. Nevertheless, this does not explain the apparent verbal qualities which the subjects report.

In general, the second possibility is endorsed, because the subjects report the characteristics of a voice. “In any case, the key methodological point is that subjects themselves firmly believe that there is more to hearing voices than merely acquiring information and telling a story about it. (...) The hypothesis that subjects are aware of self-generated inner speech provides an explanation of the verbal quality of voices that is

plausible in light of what they themselves say or believe about how voices seem to them.' (p. 30-31)

But besides the *verbal* quality, the *alien* quality of the voices must be accounted for (the misattribution aspect).

(a) A first possibility here is the auditory-hallucination model: there is a similarity between voices and genuine auditory perception. Nevertheless, it is not the *auditory* character which can explain the alien character.

(b) A second possibility is that the apparent externality of what is internal constitutes the alien quality. This is also rejected, because many hallucinators know very well that their voices are hallucinations.

(c) Another hypothesis for explaining the alien character is put forward: the apparent *unintendedness* of voices, whereas ordinary inner speech is not unintended.

Unintendedness is an indicator of nonself origin, of impressions from the outside world. Nevertheless, unintended speech is a common phenomenon. Therefore Hoffman, the author they refer to, appeals to discourse plans, which specify high-level goals (which must not be at the forefront of conscious attention) and lower-level strategies for realizing the high-level goals. These result in tacit expectations. Verbal imagery is strongly intended when it is consciously decided upon, and weakly intended when it is consonant with consciously accessible goals. Weakly intended inner speech does not trigger the experience of unintendedness and does not activate the nonself inference. But in the case of strongly unintended speech reality testing might fail to cancel the initial nonself inference. This is due to a breakdown in discourse-planning: the subject produces unintended inner speech in circumstances in which the backup process is not looking for it.

But the planning argument produces a regress problem: discourse planning itself requires planning. The authors avoid this problem by referring to another regress problem: the inner speech act intention must also be self-attributed and so on. The conclusion is that we need some account of psychological self-attribution in addition to the account of inner speech of Hoffman. This account could be used to explain self-attribution of inner speech, such that Hoffman's account may be unnecessary.

Nevertheless, the authors keep on referring to Hoffman, who continues in terms of 'action-like' attributes of voices, and not in terms of sensory or phenomenal similarities between voices and genuine instances

of auditory perception. But as voices are not in general auditory hallucinations, the hypothesis that subjects mistake voices for perceptual experience does not provide a *general* explanation of the alien quality of voices. This tells not only against Hoffman's account, but against any version of the auditory-hallucination model. Thus Hoffman is not able to give an adequate *general* account of the alien quality of verbal hallucinations. Moreover, the authors consider his idea of reality testing too simple. Therefore, the subject does not arrive at the nonself conclusion by reality testing.

As a result, the alien character of *inserted thoughts* remains unexplained. Moreover, Stephens and Graham ask whether there is a distinction between verbal hallucinations and inserted thoughts. What Hoffman's account amounts to, after all, is that people possess a sense of themselves as agents in at least some episodes in their mental lives. Nevertheless, because of some problems with his account, the authors need to go beyond Hoffman, examining thought insertion and looking for its possible explanatory connection with voices. This is the first genuine task the authors set for themselves. Their alternative does not rely on discourse plans or reality checking mechanisms. Instead, the sense of mental agency figures in the explanation of the nonself attribution. In thought insertion there is no issue of the subject's confusing the thought with an auditory perception, so that whatever explains the alien character of inserted thought may likewise capture the alien quality of 'nonsensory' or all voices. Again, ego-boundary confusion is no adequate explanation. What is needed is a sense of 'mine' in which I can acknowledge that a thought occurs in me and is mine but deny that it is *my* thought, something that *I* think. The *subjectivity* of the thought is acknowledged, it is not a loss of ego boundaries, but a violation of ego boundaries: something alien has been placed within it. The authors appeal to Frith's unified account, which, like Hoffman's, emphasizes the action-like features of thought and inner speech and (un)intendedness. Frith describes voices and thought insertion as failures of our monitoring ourselves, or rather, our intentions to act. The result of this is that the subject experiences his own intended actions as if they come about as a result of forces outside or independent of his control.

But how does this cause the subject to think that *someone else* thinks or causes the thought in her? According to Stephens and Graham, it is the *content* of the thought that provides the experiential or epistemic basis for

attributing it to another person. Moreover, alienation works the same way in thought insertion and nonauditory voices. The difference would be that patients offer different descriptions, depending on the particular life story, cultural or educational background etc. Nevertheless, they do not exclude the other possibility that the experiences really are different, and not only the description of it.

As Stephens and Graham are examining the processes by which verbal hallucinations and inserted thoughts may occur, neurological data would enrich their research significantly. The acknowledgement, by the authors, of this shortcoming unfortunately does not considerably modify this deficiency. Stephens and Graham give abundant and subtle pros and cons and they have managed to be conceptually clear throughout all the arguments. On the other hand, the fact that one third of the book consists in Hoffman's account, is indicative of the amount of novel ideas in it. It seems that their last sentence, 'There is a lot left to learn', is very suitable.

Helena De Preester
Research Assistant Fund for Scientific Research – Flanders (Belgium)
Universiteit Gent