Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations

Charles Fernyhough*

Department of Psychology, University of Durham, South Road, Durham, DH1 3LE, UK

Abstract

The phenomenon of auditory verbal hallucinations (AVHs) is one of the most intriguing features of the psychiatric literature. Two alternative models of the development of AVHs in both normal and psychotic populations are proposed. In the disruption to internalisation (DI) model, AVHs result from a disruption to the normal processes of internalisation of inner speech. In the re-expansion (RE) model, AVHs result when normal inner speech is re-expanded into inner dialogue under conditions of stress and cognitive challenge. Both models draw on Vygotsky’s (The Collected Works Of L.S. Vygotsky, New York, Plenum Press, 1987) ideas about the development of inner speech. On this view, normal inner speech is considerably abbreviated relative to external speech, and also undergoes some important semantic transformations. In both the DI and RE models, AVHs arise when the subject’s inner speech involves inappropriately expanded inner dialogue, leading the subject to experience the voices in the dialogue as alien. The two models may prove useful in explaining some of the social-developmental evidence surrounding the phenomenon, and also make a number of testable predictions which are suggested as priorities for future research.

Keywords: Auditory verbal hallucinations; Internalisation; Inner speech; Vygotsky’s theory; Psychosis; Cognitive development

*Tel.: +44 191 334 3243; fax: +44 191 334 3241.
E-mail address: c.p.fernyhough@durham.ac.uk (C. Fernyhough).
1. Introduction

The phenomenon of auditory verbal hallucinations (AVHs), where subjects report the perception of speech in the absence of any external stimulus, has proved to be one of the most consistently puzzling in the psychiatric literature. AVHs are seen as a primary positive symptom of schizophrenia, and were classified by Schneider (1959) as first-rank symptoms. More recently, a number of developments in psychiatry, psychology and neuroscience have contributed to a renewed interest in AVHs. Firstly, the move towards a non-Kraepelinian, symptom-based approach to the psychopathology of schizophrenia (e.g. Bentall, 1990, 2003; Frith, 1992) has led to a focus on AVHs as a phenomenon worthy of study in its own right, rather than as a feature of a larger explicandum, namely schizophrenia. Secondly, a growing consensus that hallucinations, verbal and otherwise, are not necessarily signs of pathology (Feelgood & Rantzen, 1994; Johns & van Os, 2001; Pearson et al., 2001; Reese, 1971; Tien, 1991) has contributed to a view of AVHs as potentially part of normal as well as abnormal experience. Thirdly, advances in brain-imaging techniques have meant that researchers have been able to study AVHs in vivo, at the very moment that the subject is experiencing them (e.g. Shergill, Brammer, Williams, Murray, & McGuire, 2000). Finally, a degree of success in the psychological modelling of AVHs (e.g. Bentall, 1990; Frith, 1992; Hoffman, 1986) has led to a new optimism that we might be within reach of a psychological-level account of these experiences.

This resurgence of interest in the psychological modelling of AVHs comes at a time of growing interest in the linguistic and verbal qualities of normal as well as abnormal consciousness. Ever since James (1890) made his observations on the ongoing interplay of verbal images that characterises human thought, the ‘stream of consciousness’ has been considered a primarily verbal phenomenon. Not only does this view have a considerable amount of intuitive appeal, it has also proved useful for a growing number of psychologists and philosophers (e.g. Carruthers, 2002; Clark, 1998; Dennett, 1997; Kinsbourne, 2000) in explaining how the functional properties of natural language can augment the pre-existing cognitive capacities of the individual. Another factor in the resurgence of interest in inner speech is the growing popularity of Vygotsky’s (e.g. 1934/1987) ideas on the phenomenon. On this view, verbal thought develops through the gradual internalisation of external forms of dialogue, with the result that mature inner speech is irreducibly dialogic in character (Fernyhough, 1996; Wertsch, 1991).

My aim in this paper is to situate the study of AVHs within the wider context of the study of inner speech, with particular focus on the normal and abnormal development of inner speech. To date, there have been no developmental-psychological accounts of AVHs. This is partly because the phenomenon has traditionally been seen as part of a broader disorder, schizophrenia, which has proved notoriously resistant to developmental explanation. Consequently, AVHs have come to scholars’ attention as a full-blown, adult symptom requiring explanation in terms of organic deficits and resultant cognitive disturbance. The present view, in contrast, is that any satisfactory theory of AVHs must incorporate
an account of how these experiences develop, and particularly their relation to normally developing inner speech. My starting point is the Vygotskian view that inner speech is irreducibly dialogic. I take a non-Kraepelinian, dimensional approach to psychosis in general, on which AVHs are viewed as part of normal as well as abnormal experience. Two alternative developmental models, essentially based on Vygotskian ideas about inner speech, are presented. Any developmental account must of course be answerable to developmental data, and, in evaluating the present models with respect to competing accounts of AVHs, I attempt to show how these models can make sense of a growing body of evidence from both normal and abnormal development.

2. Explaining auditory verbal hallucinations

A single paradox lies at the heart of the AVH puzzle. Subjects report their experience of an AVH as a perception, internal to themselves, of a voice other than their own. At the same time, the voice is usually accepted to have arisen from within the boundaries of the self. As a number of authors have noted (Hoffman, 1986; Leudar & Thomas, 2000; Stephens & Graham, 2000), this ‘alien yet self’ paradox presents a puzzle for any account of AVHs that depends for its explanatory power on a confusion between self and other. If the subject is simply drawing the boundary between self and other at the wrong location, it is difficult to see how they should also accept the hallucination as being of themselves.1

Any satisfactory psychological theory of AVHs must, then, be able to account for this central paradox. Most of the competing psychological-level models of AVHs have expended at least some explanatory effort in attempting to account for this puzzle. One of the most influential of these models is that of Hoffman (1986), who proposed that the alien quality of AVHs is a result of an inference made by subjects on the basis of experiences of unintendedness in their inner speech. That is, a breakdown in the subject’s normal discourse-planning processes results in some inner speech utterances being experienced as unintended, with the result that they are attributed to an external source. Hoffman’s account has been criticised on a number of grounds. Akins and Dennett (1986) have suggested that the account is open to an infinite-regress objection, namely that any verbal thought must be preceded by an intention, which is in itself a thought, and so on. Stephens and Graham (2000) note that Hoffman’s requirement that at least some inner speech utterances be intended requires that the corresponding discourse-plans also be intended, or risk being taken by the subject as unintended and therefore alien.

---

1It should be noted that a minority of psychiatric patients who report AVHs attribute these experiences to an entirely external source (Slade & Bentall, 1988), and thus do not demonstrate this paradox. Jaspers (1913/1963) distinguished between ‘pseudo-hallucinations’, where alien voices are recognised to come from within the boundaries of the self, and ‘true hallucinations’, where the attribution is entirely external. However, it is now generally agreed that this distinction is of little value either theoretically or clinically (Bentall, 2003).
A different account of AVHs has been put forward by Frith (1992), who suggests that AVHs arise as a result of a breakdown in the corollary discharge processes that normally accompany the instigation of an intentional action. On this view, information about the subject’s own behaviour is normally integrated with an efference copy of the command to instigate that behaviour, and thus accepted by the subject as being of internal origin. In the pathological case, the corollary discharge is missing or degraded, and the subject consequently has no basis on which to accept the behaviour as self-generated. Frith’s account is also open to at least two charges of infinite regress. Firstly, it holds that at least some inner speech utterances are intended, and so is open to the same infinite-regress charge as that levelled at Hoffman’s theory. The second charge relates to Frith’s need to postulate a monitoring system for integrating corollary-discharge and kinaesthetic information about the subject’s behaviour. Presumably, such a system will also need to have a decision-making system embedded within it for overseeing the various inputs into the monitoring system, which will in turn require a further embedded decision-making system for overseeing the various inputs of that first decision-making system, and so on.

Another model that proposes a source-monitoring deficit as a cause of AVHs is that of Bentall (1990, 2003). Drawing on experimental evidence that patients experiencing AVHs have difficulties distinguishing between internally- and externally-generated changes in their perceptual experience (e.g. Johns & McGuire, 1999), Bentall has argued that AVHs arise when subjects mistakenly label elements of their inner speech as coming from an external source. Although less vulnerable to the charges of infinite regress that can be levelled at the theories of Hoffman and Frith, one difficulty for Bentall’s theory is in explaining precisely why AVH-experiencers should err on the side of external rather than internal attributions. My aim in this article is to show that an account of the emergence of AVHs which gives due consideration to the form and function of inner speech in development may increase our understanding of why internal events are sometimes misattributed in this way.

3. A developmental approach to auditory verbal hallucinations

There are a number of reasons why a developmental approach to AVHs might seem attractive. In the first place, such an approach allows us to move beyond the treatment of AVHs as fully fledged symptoms of a pre-existing psychopathology, and instead to see how they might arise during normal and abnormal development. Related to this is the growing appreciation that AVHs can form part of normal experience in both children and adults. At the same time, a developmental approach can help us to make sense of a growing body of evidence relating to the ontogenetic course of the phenomenon.

My approach is to view AVHs as a variety, albeit an unusual and occasionally disordered one, of normal inner speech. The idea that AVHs can be seen as a form of inner speech is not new (e.g. Bentall, 2003; Bick & Kinsbourne, 1987; Leudar &
Thomas, 2000), and has gained empirical support from recent neuroimaging evidence linking the incidence of AVHs with brain processes in normal inner speech (e.g. McGuire, Shah, & Murray, 1993). However, little attention has to date been paid to the development of inner speech as a normal psychological phenomenon, and how this might bear on the disordering of inner speech that occurs in AVHs. To date, the fullest account of the development of inner speech has been that of Vygotsky.

My starting point is an idea implicit in Vygotsky’s (e.g. 1934/1987) writings, although never fully spelled out by him: that inner speech is irreducibly dialogic in character. Inner speech is the end result of a gradual process of internalisation of dialogue which begins with the child’s first entry into linguistic exchanges. Although the speech that is internalised undergoes a number of important syntactic and semantic changes during this process, it retains its dialogic character. In the version of this account spelt out by Fernyhough (1996, 1997, in press), the voices in inner dialogue represent semiotically manifested perspectives on reality, just as the voices in external dialogue represent differing perspectives on the world. Mature inner speech is an ongoing dialogue between these internalised, simultaneously held perspectives. At first blush, then, AVHs should not strike us as a very strange phenomenon. By its very nature, inner speech involves the coordination of multiple voices.

The key to understanding the strangeness of AVHs for those who experience them lies in the syntactic and semantic changes which, Vygotsky claimed, dialogue undergoes in the process of internalisation. In this paper, I present two alternative models for the development of AVHs, both of which draw on Vygotskian ideas about the development of inner speech. In outline, my argument is as follows. In normal development, the syntactic and semantic changes envisaged by Vygotsky ensure that fully internalised inner speech bears little superficial resemblance to the external dialogue from which it was derived. Most importantly, inner speech is considerably abbreviated relative to external dialogue, and only occasionally manifests the dialogic ‘give-and-take’ structure of conversation.

In this view of inner speech development, there are at least two possible situations in which the subjective experience of inner speech will be abnormal. In the first scenario, described by the disruption to internalisation (DI) model, the normal process of internalisation is disturbed, with the result that the adult’s inner speech is incompletely abbreviated, and retains many of the superficial features of external dialogue. In the second scenario, described by the re-expansion (RE) model, fully internalised inner speech is temporarily re-expanded into an inner dialogue which retains the give-and-take structure of external dialogue. In both scenarios, the fact that the resulting dialogue continues to take place in the absence of any external stimulation—that is silently, in inner speech—means that the voices in the dialogue are perceived as having an external origin. In order to understand precisely why the voices in AVHs acquire the characteristics they do, we need to take a closer look at Vygotsky’s theoretical ideas on inner speech.
4. Vygotsky’s account of the development of inner speech

Vygotsky’s (1934/1987) view of the development of inner speech formed part of a broader theory of how individual mental processes are developmentally determined by the interpersonal interactions from which they derive. In his ‘general genetic law of cultural development’ (Vygotsky, 1931/1997), Vygotsky claimed that every mental function appears twice in development: firstly on the interpsychological plane (that is, as a function distributed between more than one individual), and secondly on the intrapsychological plane, as an internalised version of that previously external function.

We can see internalisation at work in the process by which external dialogue (between, for example, a child and her caregiver) is transformed into inner speech (an inner dialogue conducted by the child alone). As an example, consider a child and her mother engaged in the collaborative solving of a jigsaw puzzle. Initially the puzzle-solving process will involve an external dialogue between child and mother, with the mother typically asking the child questions about which piece should be placed next, and the child answering, requesting further assistance, and so on. Later in development, this dialogic pattern of exchange is adopted more and more exclusively by the child, until it is fully internalised into inner speech (Wertsch & Stone, 1985). Crucially, the process of internalisation of external speech passes through a transitional phase where, inter alia, the child asks questions of herself out loud and then proceeds to answer them. This stage of private speech (see Berk, 1992, for a review) represents a waystation on the developmental path between external and inner speech.

Internalisation, on this account, is more than a simple copying of external speech onto the intrapsychological plane. Rather, the interpersonal dialogue that is internalised undergoes a number of important transformations in its conversion to inner speech. The most important structural change is abbreviation, where the psychological subject of the utterance (the ‘given’, in Chafe’s (1974) terminology) is dropped in favour of the psychological predicate (the ‘new’). The syntactic abbreviation of inner speech is responsible for our experiencing it not as a sequence of fully formed utterances, but rather as a fragmentary, condensed series of verbal images. Vygotsky’s views on abbreviation have been supported by a number of findings of abbreviation in children’s private speech, which is seen as the direct precursor of inner speech (Feigenbaum, 1992; Goudena, 1992; Pellegrini, 1981; Wertsch, 1979; Winsler, de Leon, Wallace, Carlton, & Willson-Quayle, 2003).

In addition to the syntactic abbreviation of inner speech, Vygotsky (1934/1987) proposed that inner speech has three semantic properties that distinguish it from external dialogue and private speech. Firstly, the predominance of sense over meaning refers to the way that the personal, private significance of words in inner speech takes precedence over their conventional meanings. Secondly, the process of agglutination involves the development in inner speech of hybrid words signifying complex, subject-specific concepts. Thirdly, the infusion of sense into a word describes the process whereby a word in inner speech becomes loaded with more associations than are inherent in its conventional meaning, and thus acquires a richness for the subject that is not shared by others using the same word.
The result of these syntactic and semantic changes is that inner speech loses most of the acoustic and structural qualities of external speech, and becomes a process of ‘thinking in pure meanings’ (Vygotsky, 1934/1987). In elaborating on Vygotsky’s ideas in this regard, it is possible to sketch out a four-level developmental scheme for the development of inner speech (see Fig. 1). At Level 1 (external dialogue), children and caregivers engage in overt dialogue which displays the characteristic give-and-take of conversation. At Level 2 (private speech), children conduct these dialogues in their own overt (and then gradually subvocalised) private speech. At Level 3 (expanded inner speech), private speech is fully internalised and covert, but the give-and-take of normal conversation is still manifested internally as the process of talking silently to oneself. At Level 4 (condensed inner speech), the syntactic and semantic transformations of internalisation ensure that inner speech retains few if any of the accoutrements of external language, and has become an dialogic interplay between alternate perspectives on reality (Fernyhough, 1996)—that is, the stage of ‘thinking in pure meanings’ described by Vygotsky.

Importantly, this model does not merely describe the stage-like developmental process through which the child internalises dialogue. It also allows for movement between the levels at any given point in development. Of particular importance for

![Diagram](Fig. 1. Stages of internalisation.)
present purposes is the possibility that, under demanding cognitive conditions, there can be a transition from Level 4 inner speech (fully condensed) back to Level 3 inner speech (expanded), and even back to Level 2 (dialogic private speech). It is under these conditions that our normal inner speech is experienced as an expanded dialogue, with all the give-and-take qualities of normal conversation. Under certain extremely stressful conditions, this expanded dialogue may be experienced by previously unaffected and otherwise normal subjects as an AVH. In such circumstances inner dialogue may also be re-externalised, resulting in a form of adult private speech.

5. Auditory verbal hallucinations and the development of inner speech

What can this account of inner speech development offer us in our attempt to explain AVHs? To answer this question, let us return to the paradox with which we began. How is it that individuals who report AVHs experience them as alien, at the same time as typically acknowledging that they arise from within themselves? I suggest that a solution to this paradox may lie in the account of inner speech developed above. Specifically, the internalisation of external dialogue involves taking on the voices of both participants in the dialogue, and thus something of the perspective of the other. Any disturbance to the usual transformation processes that accompany internalisation, or to the normal patterns of transition between the four levels, may result in that ‘otherness’ manifesting itself as an experience of an alien voice.

5.1. The disruption to internalisation (DI) model

The first developmental model of AVHs relates to the psychological consequences of disturbance to the normal processes of internalisation of speech. To see how a disruption of internalisation might result in inner speech that is experienced as alien, let us return to the four-level scheme outlined above. I suggest that one cause of AVHs might be a developmental problem with the transition between Level 3 (expanded inner speech) and Level 4 (condensed inner speech). Specifically, the problem may lie with the processes of syntactic and semantic abbreviation described by Vygotsky. Instead of the usual situation whereby inner speech becomes a process of ‘thinking in pure meanings’, it remains shot through with fragments of incompletely abbreviated dialogue. That is, inner speech remains excessively and inappropriately expanded. Because these fragments manifest themselves when the subject is not exposed to any external speech input, they are experienced as alien voices.

A second possible cause of disturbed internalisation which might lead to AVHs in the adult is when the disturbance occurs at Level 1 of inner speech development. If the child has a relatively impoverished experience of balanced, two-way dialogue with a caregiver, the progression through the subsequent levels of inner speech will
be disturbed. I consider below some indirect evidence that such impoverishment might lead to inner speech pathology in the adult.

Under what circumstances might the normal process of internalisation become disrupted in the ways I have described? It is here that the DI model of AVHs becomes useful in accounting for the emerging data on the development of both normal and pathological AVHs. In answering this question, we also need to consider whether there is any evidence that internalisation is disrupted in subjects who report AVHs. To date the evidence is only indirect. Here I consider two sources of evidence that may have a bearing on this issue.

One factor that might influence the progress of internalisation is the quality of infant–caregiver relationships in infancy and early childhood. A useful construct in this respect is infant–caregiver security of attachment. In their groundbreaking work on attachment classifications, Ainsworth, Blehar, Waters, and Wall (1978) described the parents of securely attached children as being more sensitive to their children’s needs, and more likely to be responsive to their current needs and goals. The distribution of attachment classifications described by Ainsworth et al. has frequently been replicated (see Solomon & George, 1999, for a review).

Among the many likely correlates of the quality of the infant–caregiver attachment relationship is the degree to which the child can engage in balanced, two-sided dialogue with the caregiver. Mothers of insecurely attached children have been shown to be less competent at adopting their child’s perspective in collaborative problem-solving tasks (Meins, 1997), and less likely to treat their infants as individuals with mental states of their own (Meins, Fernyhough, Russell, & Clark-Carter, 1998, Meins, Fernyhough, Fradley, & Tuckey, 2001; Meins et al., 2002). We might therefore expect that children who have not enjoyed secure attachment relationships in infancy will have fewer opportunities for internalising dialogue, and that their experience of dialogue with their caregivers will be relatively one-sided in comparison with their securely attached peers. In other words, we would expect a relatively impoverished experience of balanced, two-sided external dialogue, which in turn may lead to a disturbance to inner speech development at Level 1.

What evidence is there for a relation between AVHs and insecure attachment in infancy? To date there is only very limited evidence linking infantile attachment patterns with adult psychopathology. Carlson (1998) reported a significant positive correlation between a measure of attachment disorganisation in infancy and a measure of psychopathology at age 17. A slightly larger body of evidence connects psychopathology with individuals’ representations of their own childhood attachment relations as recalled in adulthood. Dozier, Stevenson, Lee, and Velligan (1991) found a high prevalence of dismissing attachment styles among a sample of patients diagnosed with schizophrenia. Subsequently, Dozier and Lee (1995) found an association between this attachment style and symptoms such as delusions, hallucinations and suspiciousness. Mickelson, Kessler, and Shaver (1997) reported links between insecure (especially dismissing) adult attachment styles and schizophrenia in large community samples, a finding that has been replicated in a large sample of adolescents (Cooper, Shaver, & Collins, 1998).
Despite this suggestive evidence, we need to be very cautious about interpreting these findings as evidence that disturbed patterns of infant–caregiver interaction have any causal role in the development of psychosis. Firstly, with the exception of one study that was able to draw on data from infancy (Carlson, 1998), the studies mentioned above have relied on obtaining information about attachment relationships from retrospective reports in adulthood. The relative scarcity of autonomous (or secure) attachment classifications may reflect the incoherence associated with schizophrenic thought disorder (Dozier, Stovall, & Albus, 1999). The value of such retrospective attachment reports is further diminished by the finding that some psychotic patients, specifically those with persecutory delusions, have difficulty accessing autobiographical information, with the result that their attachment reports may seem excessively dismissive (Kaney, Bowen-Jones, & Bentall, 1999). Finally, in none of these studies has the psychopathology involved been specifically linked to the occurrence of AVHs.

Another suggestive line of evidence for the DI model is that which relates individuals’ pre-existing communicative difficulties with their predisposition to psychosis. Clearly, speech difficulties in childhood will have implications for the individual’s opportunities to engage in external dialogue. Again, the evidence relating speech difficulties to the development of AVHs is limited and only indirect. Using data from two very large British health surveys, and hospital case notes on those participants who later became psychotic, Jones and Done (1997) reported an association between speech difficulties reported in middle childhood and the later incidence of psychosis. Such an association could, of course, reflect a pre-existing neurological deficit which predisposes individuals both to speech disorder and to psychosis. Such an early neurodevelopmental marker of later psychosis has, however, proved difficult to document (Bentall, 2003), despite attempts to characterise schizophrenia as fundamentally a neurological disorder of language (e.g. Crow, 1997).

To summarise, the DI model holds that disruption to the normal processes of internalisation of external speech results in inner dialogue that is abnormally and excessively expanded. Rather than experiencing inner speech as a syntactically abbreviated version of external dialogue, voice-hearers report hearing fragments of speech which retain the linguistic accoutrements of external speech. Because these experiences arise in the absence of any external linguistic input, they are experienced as alien. To date there is only limited and indirect evidence for the DI model. Adherents to the model would have to demonstrate a link between impoverished experience of external dialogue in infancy and childhood, and later psychosis. To date, however, such evidence has proved very difficult to document.

5.2. The re-expansion (RE) model

The second model of the development of AVHs that can be derived from Vygotsky’s ideas on inner speech relates to the possibility of movement between the four levels of inner speech. Specifically, the RE model holds that AVHs are experienced when normally abbreviated inner speech becomes re-expanded under
conditions of stress and cognitive challenge. This re-expansion can occur in both psychiatrically healthy and psychotic individuals. Re-expansion may involve a move from Level 4 (condensed inner speech) to Level 3 (expanded inner speech), or from Level 4 to Level 2 (external private speech). In the former case, the subjective experience of suddenly expanded dialogue in the absence of any external linguistic input leads to the subject reporting an AVH.

The RE model thus differs from the DI model in holding that patients who report AVHs enjoy normal, condensed inner speech under normal conditions. It is only under conditions of stress and cognitive challenge that re-expansion takes place. On this view, healthy individuals differ from voice-hearers not in their susceptibility to re-expansion, but in their interpretation of the phenomenon when it occurs. Healthy individuals may also experience re-expansion of inner speech under stressful or challenging conditions, but they do not usually interpret the resulting inner dialogue as alien.

The RE model generates at least four testable predictions. The first two of these predictions relate to voice-hearers’ experience of inner speech. To date there has been remarkably little investigation of qualitative aspects of the inner speech of psychotic patients (e.g. Hurlburt, 1990). In particular, the lack of systematic research into the dialogic (or otherwise) nature of such speech means that the following must, for the time being at least, remain highly speculative.

The first prediction that follows from the RE model is that voice-hearers will also, when they are not hallucinating, experience normal condensed inner speech. To recap, the RE model holds that AVH-experiencers are not disordered in inner speech development, but rather make pathological interpretations of the re-expansion of inner speech under challenging conditions. They should therefore report normal condensed inner speech (Level 4). In contrast, the DI model holds that inner speech development is fundamentally disordered, and thus inner speech never reaches the level of condensed inner dialogue. Rather, the DI model would predict that voice-hearers should report all inner speech as AVHs.

The second prediction that arises from these models concerns whether psychiatric patients who present with AVHs also experience ‘normal’, Level 3, expanded inner dialogues. Both the DI and RE models would lead us to predict that such individuals will have no experience of ‘normal’ Level 3 inner speech. In the case of the DI model, a disruption/disturbance to internalisation results in inner dialogue remaining abnormally expanded at Level 3, and thus experienced as alien. In the case of the RE model, the transition from Level 4 to Level 3 is for some reason surprising or distressing to the individual, with the resulting inner dialogue being experienced as alien.

A third prediction arising from the RE model is that AVHs will only be experienced under conditions associated with likely re-expansion, such as cognitive challenge or stress. In contrast, the DI model holds that AVHs will be experienced under all cognitive conditions, not just stressful or challenging ones. Systematic research into the relation between emotional arousal and the occurrence of AVHs among patients with psychosis has to date been very limited, although there has been some suggestive evidence that the onset of auditory hallucinations coincides with
physiological correlates of stress such as skin conductance (Cooklin, Sturgeon, & Leff, 1983). A priority for future research is therefore further careful investigation of the relation between AVHs and cognitive challenge or stress, with a particular focus on which cognitive aspects of stressful situations are associated with the onset of hallucinations.

A fourth prediction that arises from the RE (but not the DI) model is that previously unaffected subjects, as well as psychotic patients, will experience AVHs under conditions of stress. There is limited evidence that this is the case. There have been a number of accounts of AVHs being reported by normal subjects as a result of bereavement (e.g. Reese, 1971). Extreme stress has also been associated with both auditory hallucinations (Balan et al., 1996) and hallucinatory experiences in other modalities (e.g. Belenky, 1979; Comer, Madow, & Dixon, 1967; Siegel, 1984). In such circumstances the inner speech of normal subjects may shift from Level 4 back to Level 3, or even Level 2, and be experienced as a full-blown internal (or, in the case of Level 2, external) dialogue. There is a certain intuitive plausibility to the idea that our experiences of inner dialogue are connected with stressful situations—we only need to think of Hamlet’s inner wranglings to see how this might be true. There is also evidence that the incidence of children’s private speech increases in conditions of cognitive challenge (see e.g. Berk, 1992; Vygotsky, 1934/1987), and it seems likely that such conditions will have a similar effect on the incidence of inner speech and private speech in adults. I suggest that a similar shift may occur when previously unaffected individuals experience AVHs under certain conditions; that is, the inner dialogue becomes temporarily expanded. Because of the processing demands associated with a highly stressful situation, individuals mistakenly attribute the resulting expanded dialogue to alien voices.

The issue of the role of stress in determining AVHs raises a further issue which both the DI and RE models must address, namely, the role of trauma in the ontogeny of AVHs. There is some evidence that the development of AVHs and other psychotic symptoms is related to early trauma, such as childhood sexual abuse (e.g. Hammersley et al., 2003). So far I have said nothing about the content of AVHs. I suggest that one of the roles of trauma in AVHs is to provide, or at least constrain, the content of the hallucination. For example, if AVHs are associated with stress (in both healthy individuals and patients), it is conceivable that the stressful situation will reactivate the memory of trauma and thus furnish the content of the hallucinatory utterance. This association with early trauma would of course help to account for the often negative valence of AVHs among patients (e.g. Honig et al., 1998), and might also help to explain why some hallucinatory voices have very specific personal and acoustic properties (Leudar & Thomas, 2000).

6. Conclusions

The present paper has considered two models of the development of AVHs arising from Vygotsky’s ideas about the ontogenesis of inner speech. On this view, AVHs are seen as an unusual and occasionally pathological form of dialogic inner speech.
On the DI model, AVHs develop in psychotic individuals when there is a disturbance to the normal processes of internalisation of dialogue. Such a disturbance may arise at several points in development. For example, AVHs can develop when there is a disruption to typical patterns of dialogic interaction in infancy and early childhood. Other possible causes of disruption/disturbance to internalisation are speech difficulties which prevent the child from engaging in normal external and private speech, and the failure of syntactic abbreviation processes in internalisation, possibly as a result of the child’s diminished experience of external dialogue and private speech.

In contrast, the RE model holds that AVHs arise, not from any fundamental disturbance to internalisation, but from problems with the process by which condensed inner dialogue is occasionally transformed back into expanded inner dialogue (Level 3), or dialogic private speech (Level 2). On this view, the sudden re-expansion of dialogue under challenging cognitive conditions results in patients experiencing the resulting inner speech as alien. Under very stressful conditions, something similar may happen to previously unaffected individuals.

A strength of the DI and RE models is that they generate clear testable predictions about the developmental course of AVHs. The DI model would be supported by evidence that early experiences that might lead to a disruption of internalisation are associated with the later development of AVHs. To date, only one study (Carlson, 1998) has attempted to relate direct measures of infant–caregiver interaction to later psychopathology, although in this case the presence of AVHs was not directly assessed. Further careful longitudinal research in this area is needed if we are to gain a better understanding of the development of the phenomenon.

The RE model makes several testable predictions that would allow it to be distinguished from the DI model. It predicts that AVH-experiencers will experience normal condensed inner speech, but not normal expanded inner dialogue; that such experience will be associated with conditions of stress and cognitive challenge; and that very stressful conditions may lead to previously unaffected individuals experiencing AVHs. Gathering further data on these issues—through, for example, interview-based assessments of the quality of inner speech among patients, healthy voice-hearers and non-voice-hearers—would seem to be a priority for future research.

One advantage of both the DI and RE models is that they allow a solution to the paradox that voices in AVHs are often acknowledged simultaneously to be both alien and of the self. This is because, in internalising dialogic exchanges, the individual takes on the voice, and thus the semiotically manifested perspective, of the partner in the dialogue. Normal human thought is thus an ongoing interplay between differing perspectives on reality (Fernyhough, 1996). In the pathological case, the voices in internal dialogue are experienced by the subject in an incompletely abbreviated form, as an inappropriately expanded inner dialogue. As these experiences occur in the absence of external speech input, the voices are experienced as alien.

Both models thus allow for continuity in the experience of AVHs between psychotic and normal populations. In the case of the DI model, continuity is ensured
by the fact that internalisation is not an all-or-nothing process, and thus that its disruption may be a matter of degree. In very seriously affected individuals, internalisation may be held up at Level 1 or 2. Less seriously affected individuals may experience AVHs because of problems with the syntactic abbreviation that would usually occur in the transition between Levels 3 and 4. On the RE model, continuity is possible because of the proposed relation between re-expansion and cognitive challenge. In normal individuals, conditions of extreme stress or cognitive challenge can result in re-expansion of inner dialogue, which may be further re-externalised as dialogic private speech. In very extreme cases of stress and trauma, the re-expanded utterances in inner dialogue may be experienced as alien.

In comparison with current theories of AVHs, the DI and RE models enjoy a number of further advantages. Firstly, they avoid getting caught up on the issue of the ‘intendedness’ of inner speech utterances, as discussed by Stephens and Graham (2000) and Akins and Dennett (1986) in relation to Hoffman’s (1986) model. Because inner speech, on these accounts, is fundamentally different to external speech in its syntactic structure and semantic properties, it does not require the same level of discourse planning as external speech. Intended inner speech utterances can occur, of course, but these happen at the level of expanded inner dialogue (Level 3), where discourse-planning demands are presumably similar to those involved in external speech.

Secondly, current theories of AVHs are hard pressed to explain the social-developmental evidence surrounding the phenomenon. How, for example, might Hoffman’s (1986) model of AVHs explain the evidence for disturbed attachment patterns in individuals who later become psychotic? It is similarly difficult to see how such findings can be accommodated within a neuropsychological theory such as Frith’s (1992). That is not to say that there is no common ground between the DI and RE accounts and those theories of AVHs that attribute them to a primary neurological disorder, such as a frontal monitoring deficit (Frith, 1992). Later I consider how such brain systems might be necessary for the internalisation of dialogue, and thus how the DI and RE models can be assimilated into the broader picture of neurological damage in schizophrenia.

Thirdly, the models presented here are less reliant than many of their competitors on a simple confusion between internal and external sources of data, with all the dangers of regress and circularity that come with that form of explanation. Rather, the Vygotskian approach allows us to rethink the dichotomies of self/other and inner/outer as they relate to inner speech and verbal thought. On this account, the inner is always at least partly outer: normal inner speech is shot through with alternative perspectives on reality. At the same time, individual thought is seen as a distributed, essentially social process: as much a collaboration between individuals as a solo endeavour. This view thus has much in common with recent attempts to do justice to the ‘extended’ nature of individual mental functioning (e.g. Clark & Chalmers, 1998; Dennett, 1997; Wertsch, 1991). It also means that the Vygotskian approach is equally appropriate to situations where alien voices are recognised to come from within the boundaries of the self, and to those where they are attributed to an entirely external source (see footnote 1).
To take the Vygotskian view that the internal is always partly external does not, however, explain why some individuals make systematic errors in explaining the provenance of their own thoughts. For most of us, the fact that our inner speech is shot through with other voices does not lead to our perception of these voices as alien. To put it another way, it is difficult to see how the hearing of voices could be construed as anything other than a problem in monitoring the different sources of experiences. What the Vygotskian models presented here can add to, say, Bentall’s (1990, 2003) source-monitoring account is their ability to explain why external attributions are made for internal events. The key lies in the nature of inner speech, and particularly the expanded form of inner dialogue that Vygotsky hinted at in his discussion of the internalisation of linguistic exchanges. That is, certain types of internal event (expanded inner dialogues) have a form that makes them more likely to be attributed to an external source. It may be that individuals with relatively less accurate source monitoring capacities will, when experiencing expanded inner dialogue under conditions of stress or cognitive challenge, be more likely than their healthy counterparts to attribute elements of that dialogue to external sources. Such a view would entail seeing the Vygotskian and source-monitoring accounts as potentially complementary rather than mutually exclusive.

Despite these apparent strengths of the DI and RE accounts, a number of gaps in the theory remain to be filled. Firstly, we need more specificity in the four-level model of inner speech development outlined here. What conditions must be met before movement between levels can occur? Under precisely what circumstances does normal subjects’ inner speech revert from Level 4 to Level 3 or Level 2? What demands do these varieties of inner speech place on the various components of the cognitive system, such as working memory? How important is the presence or absence of external speech input in determining whether a subject will experience AVHs? What is the role of previous trauma in the development of AVHs, perhaps in lowering the threshold of stress or cognitive challenge needed to trigger the re-expansion of inner dialogue? In addition to these issues specific to the DI and RE models, the Vygotskian concept of internalisation will continue to benefit from further clarification, elaboration, and empirical testing. Although some important steps have recently been taken in this regard (e.g. Tomasello, Kruger, & Ratner, 1993), much work remains to be done in specifying the cognitive mechanisms underlying internalisation.

Another problem facing the DI and RE models is the need to explain why the psychiatric consequences of disrupted internalisation only become apparent in adulthood. This problem is particularly acute for the DI model, which would seem to hold that disruptions to internalisation leading to AVHs in adulthood should be detectable at earlier stages of development. It is also problematic for the RE model, which would need to explain why the re-expansion of inner dialogue should often have a specific onset in adulthood. My approach to this problem is to note the gaps in our knowledge relating to the developmental course of AVHs, and the continuity in AVHs between normal and clinical populations. For example, it may be that ceasing to view AVHs as necessarily pathological might lead to us seeing evidence for them much earlier in development. There is already evidence that a later proclivity to
develop AVHs can be detected in adolescence (e.g. Escher, Romme, Buiks, Delespaul, & van Os, 2002). There is also some suggestive recent evidence that normal children can experience what we would otherwise describe as AVHs, particularly in the context of their dealings with imaginary companions. For example, Pearson et al. (2001) reported that exposing 9- to 11-year-old children to ambiguous perceptual stimuli led to hallucination-like experiences, especially when subjects reported imaginary companions. Much more work remains to be done to link children’s experiences of AVHs with the developmental course of private and inner speech. Similarly, future research into developmental precursors of AVHs in particular, and psychosis in general, would do well to consider how individual differences in the emergence of private speech, and its subsequent internalisation into inner speech, are related to later susceptibility to hallucinations.

It is important to note, however, that what may be pathological in adulthood may not necessarily be pathological in childhood. Indeed, the Vygotskian approach to AVHs implies that, at the stage when children are internalising external dialogue and engaging in conversations with imaginary companions, experiences that would otherwise be classified as AVHs might be the norm rather than the exception. These experiences should, perhaps, only be seen as pathological if they continue into adulthood. The DI model holds that AVHs may be a developmental consequence of failures in this process of internalisation, while the RE model holds that AVHs result from abnormalities in the process whereby internalisation is temporarily reversed. It seems likely that, through learning much more than we currently know about the mechanisms of internalisation, we might gain some important insights into what happens when this crucial developmental process goes wrong.

Another challenge for the present model is to account for the overwhelming evidence that the main disorder associated with AVHs, schizophrenia, is a biological disorder with a strong genetic component. How can we square any developmental psychological account of AVHs with such evidence? Firstly, we can point out that these Vygotskian models of AVHs are not models of schizophrenia. Rather, they attempt to explain a symptom which occurs frequently in schizophrenia, but also in bipolar disorder (Potash et al., 2001), and in normal adult and child populations. A second point is that the neurological systems whose breakdown is implicated in schizophrenia may well have a role to play in the process of internalisation. For example, internalising dialogue must minimally involve the integration of one’s own contribution to the external dialogue with the contribution of one’s interlocutor, thus drawing on the same kind of monitoring capacities whose breakdown has been implicated in schizophrenia (e.g. Frith, 1992). It may also be that the cognitive resources that are furthest stretched in conditions of stress and

---

2This argument would appear to be challenged by the observation that many psychiatric patients can date the onset of their voices to a particular point in time, often associated with a trauma or crisis in adulthood. It could be argued that any evidence for such time-specificity of onset would provide a compelling reason for favouring the RE model over the DI one. That said, the Vygotskian approach to AVHs also provides grounds for doubting patients’ self-reports of hallucination onset. If it is correct that hallucination-like experiences are not uncommon in childhood, it would not be surprising if they were not accurately recalled in adulthood. Only careful longitudinal research can settle this issue.
cognitive challenge—the executive functions—are responsible for controlling the re-expansion and re-externalisation of inner speech, and thus that the disturbance to these processes in subjects with AVHs has a fundamentally executive cause.

One further question arises from the Vygotskian approach to AVHs. The foregoing discussion has focused primarily on Vygotsky’s ideas about syntactic abbreviation in internalisation. What of the semantic abbreviation processes described by Vygotsky? One might argue that the processes of semantic transformation described by Vygotsky actually amount to a fairly good description of disordered speech in schizophrenia. If this characterisation of ‘schizophrenene’ is accurate, we would have to conclude that the external language of patients with speech disorder undergoes the same semantic transformations as normal subjects’ inner speech. We could perhaps take this as evidence that the inner speech of such individuals is similarly transformed. What seems certain is that there is something inappropriate in the way these features are not confined, as they are in normal subjects, to inner speech. If schizophrenic speech is like normal inner speech spoken out loud, it is interesting to consider what cognitive features of the disorder might be responsible. It seems plausible that this inappropriate externalisation of inner speech is related to the mentalising and perspective-taking deficits reported in such individuals (see Bentall, Corcoran, Howard, Blackwood, & Kinderman, 2001, for a review).

Two final points need to be made. Firstly, contra Slade & Bentall (1988), the present account has nothing to say about hallucinations in other modalities. Secondly, we should consider whether the DI model has any implications for therapy. I arrive at this question against the background of some recent successes in enabling patients with disturbing AVHs to engage with their voices (e.g. Davies, Thomas, & Leudar, 1999). It may be that encouraging patients to engage with their voices allows them to correct abnormalities in the normal processes of internalisation and re-expansion. In this way, a troubling experience of alien voices might become a true inner dialogue: condensed, abbreviated, semantically transformed, and indistinguishable from normal inner speech.

Acknowledgements

I am grateful to Elizabeth Meins, Sue Leekam and Richard Bentall for helpful comments on an earlier version of this manuscript. This article was prepared when I was a visiting researcher at the Department of Psychology, Macquarie University, Sydney.

References


