Fear, Inner Speech, and Hostility in Auditory Verbal Hallucinations:

A Model of Fear-Mediated Misattribution

Vadim Saunanen

Independent Researcher, Finland

Email: vadsaunanen@icloud.com

ORCID iD: https://orcid.org/0009-0002-1847-627X

Abstract

Auditory verbal hallucinations are among the most characteristic and clinically significant symptoms of psychotic disorders; however, the mechanisms underlying the formation of their content—particularly its hostile and threatening nature—remain insufficiently understood. The present article proposes a theoretical model that explains the aggressive and threatening content of auditory hallucinations through a mechanism of fear-driven misattribution of inner speech.

According to the proposed concept, impairment in the recognition of authorship of inner speech leads to its perception as alien, thereby activating evolutionarily grounded threat-detection systems. The resulting fear is conceptualized not as a secondary emotional reaction, but as an active generative factor that shapes the content of perceived voices through subconscious simulation of potential danger. These simulations are experienced as external utterances, giving rise to a self-sustaining cycle of "alien voice – fear – hostile content."

The model integrates findings on the role of traumatic experience, individual differences in the thematic content of voices, and the sensory vividness of hallucinatory experiences. The proposed approach supports a continuum-based view of the boundary between norm and pathology and carries important clinical implications, particularly for psychotherapeutic interventions aimed at reducing fear and restoring a sense of authorship over internal experiences.

Keywords

Auditory verbal hallucinations; Inner speech; Source monitoring; Fear; Self-attribution; Psychosis

Introduction

Auditory verbal hallucinations (AVH) are among the most extensively studied yet conceptually complex symptoms of psychotic disorders, particularly schizophrenia. Despite substantial progress in understanding the neurobiological and cognitive mechanisms underlying the emergence of hallucinations, the question of why the content of perceived voices so often takes on a threatening, accusatory, or hostile character remains insufficiently explained.

Contemporary models largely converge on the view that auditory verbal hallucinations are related to disturbances in inner speech processes and self-monitoring of thought. Within cognitive frameworks, it is proposed that internal verbal processes are erroneously attributed to external sources, leading to the experience of voices as “alien” or autonomous agents (Frith, 1992; Bentall, 1990; Jones & Fernyhough, 2007). Neurocognitive models complement this account by pointing to dysfunctions in source monitoring mechanisms and in the prediction of the sensory consequences of one’s own actions (Feinberg, 1978; Ford & Mathalon, 2005).

However, despite a relative consensus regarding the origin of voices, considerably less attention has been paid to the mechanisms underlying the formation of their content. In particular, it remains unclear why, in a substantial proportion of cases, voices are not merely neutral but instead take the form of threats, insults, accusations, or commands to engage in violent behavior (Nayani & David, 1996; McCarthy-Jones et al., 2014). Existing explanations typically appeal to the role of traumatic experience, negative self-schemas, and affective disturbances (Birchwood et al., 2004; Hardy et al., 2005), yet these factors are often treated as background correlates rather than as active mechanisms involved in content generation.

The present article proposes a theoretical model in which fear is conceptualized not as a secondary emotional reaction to the emergence of voices, but as a central generative mechanism shaping their hostile content. The core hypothesis suggests that impairment in recognizing inner speech as self-generated leads to its attribution to an external agent. Within

the framework of human cognitive and evolutionary architecture, the perception of an “alien” agent automatically activates threat-detection systems. This activation, in turn, elicits fear, which begins to structure and populate the content of the perceived voices.

Within the proposed model, threatening verbal expressions are interpreted as projections of the individual’s own fears, anxious expectations, and affectively charged scenarios generated by subconscious processes. Because these processes operate outside conscious control and lack the marker of self-authorship, their products are experienced as externally imposed. In this way, a self-reinforcing cycle emerges in which perceiving the voice as alien amplifies fear, and fear, in turn, intensifies the hostile content of the voice.

Additionally, the model draws on phenomenological parallels with dreaming, in which fear is likewise capable of autonomously generating threatening images and scenarios in the absence of conscious control. It is hypothesized that in auditory verbal hallucinations, similar mechanisms of imagery and affective simulation are activated during wakefulness, but are erroneously interpreted as external and real.

The aim of the present article is to introduce and conceptually substantiate a model of fear-driven misattribution of inner speech that accounts for the frequent hostility and threatening nature of auditory verbal hallucinations. The article is theoretical in nature and seeks to integrate cognitive, phenomenological, and affective perspectives, as well as to formulate hypotheses that may be tested in future empirical research.

Inner Speech as a Normal Cognitive Process

Inner speech is a fundamental component of human thinking and self-regulation. It plays a central role in action planning, reflection, emotional regulation, and the formation of the sense of self. Most individuals are continuously engaged in internal dialogue which, despite the absence of external sound, possesses a structure comparable to overt speech, including syntax, intonation, and semantics (Vygotsky, 1934/1987; Alderson-Day & Fernyhough, 2015).

Contemporary research indicates that inner speech is not a homogeneous phenomenon. It can vary in terms of voluntariness, emotional intensity, and sensory richness. In some cases, inner speech takes the form of spontaneous or intrusive thoughts that arise without conscious

intention and may cause distress; however, a key feature of normative functioning is preserved—namely, the recognition of these thoughts as one's own (Morin, 2009).

The phenomenon of the repetitive melody, commonly referred to as an "earworm," provides a clear example of the autonomous activity of inner speech and imagination. Musical fragments may involuntarily replay in consciousness, often accompanied by the voice of a specific performer, including characteristic timbre and intonation. Despite their vivid sensory qualities, such experiences are not interpreted as external or imposed from outside, but are recognized as internal mental processes, even if not fully under voluntary control (Liikkanen, 2012). This example illustrates that the presence of rich sensory characteristics alone does not lead to the experience of "alienness."

A key mechanism supporting the normal functioning of inner speech is the system of self-reference and source monitoring. The brain continuously differentiates between mental events that are generated by one's own cognitive activity and those that originate from external sources. This process operates automatically and outside conscious awareness, producing a stable sense of authorship over one's thoughts (source ownership) (Johnson et al., 1993).

Disruptions in these mechanisms may give rise to states in which one's own thoughts are experienced as intrusive, unwanted, or incongruent with the self-concept. However, within normative functioning, such states are typically accompanied by preserved metacognitive insight into their internal origin, even when they are subjectively experienced as "alien in content" (for example, in obsessive thoughts) (Poyurovsky et al., 2008).

It is important to emphasize that in normal mental functioning, conscious and subconscious processes do not exist as separate structures but rather form a continuous and dynamic continuum. Subconscious contents may spontaneously enter awareness without losing their connection to the unified structure of the self. Even in cases of emotionally charged or intrusive thoughts, individuals retain a basic sense of ownership over these experiences.

Thus, inner speech under normative conditions is characterized by three key features:

1. spontaneity and variability of content,
2. the potential for sensory richness, and
3. preserved recognition of self-authorship.

According to the model proposed in this article, the loss of the third feature constitutes the critical transition point from normative inner experience to the pathological experience of voices.

Disruption of Authorship Attribution and the Phenomenon of the "Alien Voice"

One of the central phenomena underlying auditory verbal hallucinations is a disruption in the attribution of authorship to one's own mental acts. In cognitive psychology and neuroscience, this process is described as a failure of source monitoring mechanisms, whereby internal speech events are mistakenly attributed to an external agent (Johnson et al., 1993; Bentall, 1990).

Under normal conditions, inner speech is accompanied by an implicit sense of authorship that does not require conscious verification. Individuals do not "check" whether a thought belongs to them; this sense of ownership is embedded within the act of thinking itself. In psychotic disorders, however, this mechanism loses its reliability, resulting in internal speech processes being experienced as autonomous and not belonging to the subject (Frith, 1992).

Importantly, this phenomenon does not involve the emergence of fundamentally new content, but rather a change in the status of already existing cognitive processes. Thoughts, phrases, or images that are ordinarily integrated into the stream of consciousness acquire the quality of "intrusions" when authorship attribution is disrupted. They are experienced as imposed, unsolicited, and beyond voluntary control.

Phenomenological studies of auditory hallucinations indicate that patients frequently describe voices as possessing their own identity, intentions, and even personality traits (Nayani & David, 1996). Within the proposed model, this is explained not by the presence of a separate mental agent, but by the fact that inner speech deprived of authorship is automatically personified by consciousness. Human cognition is evolutionarily attuned to agency: meaningful and goal-directed signals are predisposed to be interpreted as originating from an intentional subject.

Thus, the loss of the sense of self-authorship initiates a process of secondary interpretation: if a thought is not experienced as “mine,” it must belong to someone else. This mechanism becomes particularly salient under conditions of uncertainty and cognitive distress, which are characteristic of psychotic states.

It should be emphasized that this dissociation does not imply a complete separation of conscious and subconscious processes as independent systems. Rather, it reflects a functional disruption in which subconscious contents lose access to markers of self-reference. As a result, consciousness begins to experience these contents as external to the self, despite their internal origin.

In this context, the contrast with obsessive states is especially informative. In obsessive–compulsive disorder, intrusive thoughts may be experienced as alien in content but not in source. Patients generally recognize these thoughts as their own, even when they find them unacceptable or frightening. In auditory hallucinations, this level of metacognitive recognition is impaired, leading to the experience of thoughts as voices.

In summary, disruption of authorship attribution constitutes a necessary but insufficient condition for the formation of auditory verbal hallucinations. It explains why a voice is experienced as alien, but it does not explain why that voice so often adopts a hostile or threatening character. To address this question, it is necessary to examine the role of fear as an active cognitive and affective mechanism, which will be the focus of the following section.

Fear as a Trigger of Hostile Content: The Closed-Loop Hypothesis

Although disruption of authorship attribution of inner speech can explain the emergence of an “alien” voice, it does not address a key question: why the content of these voices so frequently takes on a threatening, accusatory, or aggressive character. Within the proposed model, fear is hypothesized to play a decisive role in shaping such content, arising in response to the perception of inner speech as foreign.

From the perspective of evolutionary psychology, the perception of an external or alien agent automatically activates threat-detection systems. The human brain is tuned to prioritize the processing of potentially dangerous stimuli, particularly those that display signs of agency

and intentionality (Öhman & Mineka, 2001). When an internal voice is no longer recognized as self-generated, it falls into the category of ambiguous and potentially threatening stimuli, thereby initiating a fear response.

Unlike rational fear directed toward a specific external object, fear in psychotic experience lacks a clearly identifiable source. It becomes diffuse and self-sustaining, intensifying cognitive hypervigilance. In this state, consciousness actively seeks explanations for what is occurring and begins to anticipate possible threats emanating from the perceived "alien" agent.

The central hypothesis of this article is that fear does not merely accompany auditory hallucinations but directly shapes their content. Subconscious processes activated by fear begin to generate simulations of potential danger. These simulations include threats, accusations, and commands to harm oneself or others—not because the voice itself is malevolent, but because such scenarios represent cognitive simulations of worst-case outcomes.

A comparable mechanism can be observed in dreams, particularly nightmares, where fear autonomously constructs frightening images and events without volitional control. During sleep, these images are experienced as real due to the attenuation of critical evaluation and self-referential mechanisms. Within the proposed model, auditory verbal hallucinations are conceptualized as a functionally similar process occurring during wakefulness, albeit with partial preservation of conscious awareness.

This gives rise to a closed loop:

1. Inner speech loses markers of self-authorship;
2. The voice is perceived as an alien agent;
3. Perception of an alien agent activates fear;
4. Fear triggers the generation of threatening scenarios;
5. These scenarios are reproduced in the form of voices;
6. Hostile content intensifies fear, thereby closing the loop.

An important aspect of this model is that the content of voices is neither arbitrary nor random. It is structured around individual fears, traumatic experiences, and affectively salient

themes. Accordingly, hostile verbalizations should be understood not as external intrusions, but as internal cognitive–affective constructions deprived of recognizable authorship.

Consequently, the threatening nature of voices may be viewed as a byproduct of psychological defense mechanisms aimed at detecting and anticipating danger. Under normal conditions, these mechanisms serve adaptive purposes; however, when the integration of conscious and subconscious processes is disrupted, they assume a maladaptive form.

Individual Differences: Trauma, Personal Experience, and the Thematic Content of Voices

One of the most consistent empirical observations in research on auditory verbal hallucinations is the pronounced variability of their content. Voices differ in thematic focus, emotional tone, degree of hostility, and perceived "personality." In some individuals they are overtly aggressive and threatening, in others accusatory or demeaning, while in certain cases they may be neutral or even supportive (Nayani & David, 1996; McCarthy-Jones et al., 2014).

Within the proposed model, this variability is explained by individual differences in affective and traumatic experience, as well as in the structure of personal fears. Because the content of voices is generated by fear as an active generative mechanism, those themes that carry the greatest emotional salience for the individual become dominant within the hallucinatory experience.

Empirical studies consistently demonstrate an association between traumatic experiences and the content of auditory hallucinations. Experiences of abuse, neglect, harsh criticism from significant others, or chronic feelings of helplessness are often reflected in the character of the voices that are heard (Read et al., 2005; Hardy et al., 2005). In such cases, the voice may reproduce the intonation, phrasing, and emotional tone characteristic of past traumatic interpersonal interactions.

It is important to emphasize that this does not constitute a literal "replay" of memory. Subconscious processes do not operate on memories in their original form, but rather transform them into generalized threat-related scenarios. These scenarios are subsequently expressed in

the form of voices whose content reflects not specific past events, but enduring expectations of danger and negative evaluation.

The phenomenon of voice personification can also be understood through the lens of individual experience. If the individual has previously been exposed to authoritarian, critical, or frightening figures, the voice often acquires a corresponding timbre, intonation, and communicative style. As with the internal reproduction of musical pieces, the mind is capable of generating highly specific vocal characteristics without an external source; however, when authorship attribution is disrupted, these characteristics are experienced as belonging to an external agent.

Not all “alien” voices, however, are hostile. In cases where dominant affective states involve needs for support, protection, or approval rather than fear, the content of voices may assume a compensatory or comforting character. This observation is fully compatible with the proposed model and underscores that the emotional valence of voices is determined by the prevailing affective state, rather than by the mere presence of disrupted authorship attribution.

Thus, individual differences in the content of auditory hallucinations reflect the unique configuration of fears, traumatic experiences, and affective needs of each person. The proposed model integrates this variability into a unified theoretical framework, treating such differences not as anomalies, but as expected expressions of a common underlying mechanism.

Why Voices Are Experienced as “Real”: The Role of Imagination and Sensory Integration

One of the most compelling characteristics of auditory verbal hallucinations for patients themselves is their strong sense of subjective reality. Voices are often experienced as originating externally, with a clear spatial localization, timbre, loudness, and directionality. In some cases, hallucinations are accompanied by visual imagery, a sense of presence, or even olfactory and tactile components (Waters et al., 2012).

At first glance, such sensory richness may appear to argue against cognitive models that explain voices in terms of inner speech. However, a growing body of evidence indicates that human imagination is capable of generating sensory experiences with a high degree of realism,

particularly under conditions of emotional arousal and reduced metacognitive control (Pearson et al., 2015).

Even outside the context of psychopathology, individuals differ substantially in the vividness and intensity of their imaginative capacities. Some people can vividly recreate sounds, smells, or tastes previously experienced in reality, approaching near-complete sensory simulation, whereas others struggle to visualize even simple images. These differences are considered relatively stable individual traits and have identifiable neural correlates (Cui et al., 2007).

Within the proposed model, the degree of sensory "reality" of voices is assumed to depend on the interaction of two factors:

1. disruption of authorship attribution for inner speech, and
2. individual capacity for sensory imagination.

When these factors co-occur, internally generated cognitive events cease to be marked as thoughts while simultaneously acquiring a high level of sensory detail. As a result, the voice is experienced not as an abstract mental process but as a fully fledged perceptual event.

An additional contribution comes from the weakening of reality-monitoring mechanisms that is characteristic of psychotic states. Under normal conditions, even vivid imaginative imagery is automatically recognized as unreal. When the integration of sensory and cognitive information is disrupted, however, this distinction becomes blurred, strengthening the individual's conviction that the voices have an external origin (Frith, 2005).

Similar mechanisms can be observed in other states, such as hypnagogic hallucinations, vivid dreaming, or dissociative experiences. In all of these conditions, internally generated images and sounds may be experienced as real when conscious regulatory control is temporarily diminished.

Thus, the subjective "reality" of voices does not contradict their internal origin. On the contrary, it points to the involvement of powerful mechanisms of imagination and sensory simulation which, when combined with a loss of authorship attribution, produce a genuine perceptual experience. This framework helps explain why some individuals experience voices exclusively "inside the head," whereas others perceive them as external and spatially localized.

The Boundary Between Norm and Pathology: From Inner Dialogue to Auditory Hallucinations

Traditional clinical thinking often draws a sharp boundary between normal mental processes and pathological symptoms. However, with regard to inner speech and auditory verbal hallucinations, an increasing body of evidence points to the existence of a continuous spectrum rather than a qualitative divide (Johns & van Os, 2001).

As shown in the previous sections, inner speech in healthy functioning can be spontaneous, emotionally charged, and even intrusive. Individuals are capable of hearing voices of other people, distinct intonations, musical fragments, or engaging in internal dialogues without any external auditory input. These phenomena become pathological not because of their content or form, but because of a loss of integration with the sense of self and with the recognition of authorship.

Within the proposed model, the transition from norm to pathology is described not as the sudden emergence of a new mental phenomenon, but as a gradual amplification of three interrelated factors:

1. weakening of self-referential mechanisms,
2. increased affective tension, primarily fear, and
3. heightened sensory vividness of internal imagery.

In a normative state, even emotionally unpleasant or frightening thoughts retain the status of being “mine.” They may evoke anxiety, but they are not interpreted as intrusions from an external source. Under conditions of increased stress, trauma, or neurobiological disruption, this integration begins to weaken. Subconscious contents become more autonomous, while fear amplifies their salience and intrusiveness.

Clinically significant pathology emerges at the point where inner speech loses its affiliation with the unified structure of the self and begins to be experienced as a separate agent. This point can be regarded as a critical threshold beyond which the psyche’s protective mechanisms cease to serve adaptive functions and instead sustain a maladaptive self-reinforcing cycle.

Importantly, similar processes can be observed in subclinical forms. Research indicates that a substantial proportion of the general population experiences episodes of hearing voices at some point in life without developing a psychotic disorder (Johns et al., 2014). This finding supports the continuum view and suggests that auditory hallucinations are not a unique marker of illness, but rather an extreme expression of universal cognitive mechanisms.

Thus, the boundary between norm and pathology does not lie in the presence or absence of inner speech itself, but in the degree of its integration, controllability, and affective tone. The proposed model conceptualizes auditory verbal hallucinations as the result of quantitative, rather than qualitative, changes in mental functioning.

Such an approach has important clinical and theoretical implications. It reduces the stigmatization of patients, shifts the focus from a "foreign symptom" to disrupted integration of otherwise normal processes, and opens avenues for therapeutic interventions aimed at restoring a sense of authorship and reducing fear.

Comparison with Existing Models of Auditory Verbal Hallucinations

The hypothesis proposed in the present work does not aim to replace existing neurocognitive models of auditory verbal hallucinations (AVH), but rather to be considered as a potential extension and clarification of several already established mechanisms. This section offers a conceptual comparison between the proposed approach and the most influential theoretical frameworks in AVH research.

Relation to Inner Speech Source Monitoring Deficit Models

The most widely accepted cognitive models of AVH attribute their emergence to impairments in source monitoring of inner speech, whereby products of one's own thinking are mistakenly attributed to an external agent (Bentall, 1990; Frith, 1992; Johns et al., 2001). Within these approaches, deficits in recognizing the authorship of internal verbal material are considered central.

The present hypothesis is conceptually consistent with this position, insofar as it assumes that the loss of the sense of ownership over inner speech may constitute a necessary

condition for the emergence of the experience of an alien voice. However, unlike classical formulations that primarily focus on the mechanism of misattribution itself, the current work proposes that affective processes—specifically fear—may play a critical role in shaping the content and emotional tone of hallucinatory experience once authorship has been lost.

Comparison with Affective and Trauma-Based Accounts

A number of contemporary models emphasize the importance of affective factors, including anxiety, stress, and traumatic experience, in modulating the content of auditory hallucinations (Morrison et al., 2003; Hardy et al., 2005). Within these accounts, traumatic memories and negative self-beliefs are often viewed as sources of recurrent and emotionally charged verbal phenomena.

The present hypothesis does not reject this contribution, but rather seeks to refine the mechanism through which it may operate. It is proposed that fear arising from the perception of the inner voice as alien and potentially dangerous may act as a trigger for the activation of subconscious scenarios associated with previously experienced threats and affective states. From this perspective, traumatic content in voices may be understood not solely as a direct replay of memory, but also as the result of the psyche's predictive activity aimed at simulating possible danger.

Relation to Predictive and Bayesian Models of Perception

Contemporary neurocognitive theories increasingly conceptualize psychopathological phenomena within the framework of predictive coding, in which perception emerges from the interaction between sensory input and prior expectations (Friston, 2005; Fletcher & Frith, 2009). In the context of AVH, it has been suggested that excessive precision assigned to prior predictions may lead internal signals to be experienced as external perceptions.

The present hypothesis can be interpreted as a specific phenomenological instance within the predictive framework, in which fear functions as an amplifier of negative prior expectations concerning the intentions of an "alien" agent. In this context, the threatening content of voices is conceptualized as the outcome of affectively biased predictions rather than as information originating from an independent external source.

Distinction from Ontological Interpretations of a "Hostile Agent"

In some descriptive and clinical approaches, auditory hallucinations are interpreted—by patients themselves—as possessing autonomous will, intentions, and personal characteristics. The present work makes no ontological claims regarding the nature of these experiences and treats them exclusively as subjective phenomena of consciousness.

The hypothesis advanced here suggests that the perception of voices as hostile may arise secondarily, as a consequence of interpreting one's own internal experience under conditions of impaired authorship attribution and heightened affective arousal. Accordingly, hostility is viewed not as an intrinsic property of the voices, but as a possible product of cognitive–affective dynamics.

Overall Positioning of the Hypothesis

Taken together, the proposed model may be regarded as a hypothetical extension of existing approaches, integrating mechanisms of inner speech misattribution, affective regulation, and predictive processing. It does not claim to offer a universal explanation for all forms of auditory hallucinations, but may be particularly relevant for understanding cases in which threatening and hostile verbal experiences predominate.

Further empirical investigation is required to clarify the explanatory value and clinical applicability of this hypothesis.

Empirically Testable Hypotheses

Given that the proposed model is hypothetical in nature, its value largely depends on the possibility of empirically testing its core assumptions. The present article does not aim to provide an exhaustive list of predictions; however, the proposed framework allows several key hypotheses to be identified that may be examined in future research.

Hypothesis 1. Association between impaired authorship attribution and hostility of voices

It is hypothesized that the degree of impairment in recognizing inner speech as self-generated may be related to the phenomenological characteristics of auditory verbal

hallucinations. Specifically, a more pronounced loss of the sense of authorship over inner verbal experience may be associated with greater perceived hostility, threatening content, and experienced autonomy of voices.

This hypothesis is consistent with existing cognitive models of source monitoring deficits, while extending them by suggesting a possible relationship between the severity of attributional impairment and qualitative features of hallucinatory experience.

Hypothesis 2. Affective role of fear as a potential mediator

Within the proposed framework, fear is hypothesized to play a mediating role between the loss of authorship attribution and the emergence of threatening content in auditory verbal hallucinations. That is, perceiving inner verbal experience as alien and potentially dangerous may be accompanied by affective responses that, in turn, facilitate the generation of negative and hostile interpretations of voices.

Importantly, fear is not conceptualized here as an inevitable consequence of hearing voices, but rather as a potential factor that amplifies or modulates their content. Empirical examination of this hypothesis could contribute to a more precise understanding of the role of affect in shaping the phenomenology of auditory hallucinations.

Hypothesis 3. Individual specificity of voice content

It is further hypothesized that the specific content of hostile auditory hallucinations may be related to individually salient fears, affective schemas, and prior subjective experience. Within this model, hostile verbalizations are not viewed as universal or random phenomena, but as potential reflections of personal cognitive–affective patterns.

This assumption allows variability in the content of auditory hallucinations to be treated as a meaningful phenomenon rather than as a secondary or incidental by-product of psychopathology, and may hold relevance for future phenomenological and clinical investigations.

Concluding remark

It should be emphasized that the hypotheses outlined above do not constitute claims of causality and are not intended to apply universally to all forms of auditory hallucinations. Rather, they represent logical implications of the proposed theoretical model and require further empirical testing using both quantitative and qualitative research methods.

Clinical and Therapeutic Implications

The proposed model of fear-mediated misattribution of inner speech carries several important clinical implications. By conceptualizing auditory verbal hallucinations not as autonomous pathological phenomena but as the result of a disintegration of normally functioning cognitive processes, the model offers a reframed perspective on both therapy and clinician–patient interaction.

First, this framework shifts the therapeutic focus away from direct confrontation with voices toward addressing fear and the interpretation of subjective experience. If hostile voice content emerges as a consequence of fear, attempts to suppress or ignore voices without engaging the affective component may prove ineffective or may even exacerbate symptom severity.

This perspective is consistent with findings from cognitive-behavioral and metacognitive approaches, which suggest that modifying a patient's relationship to voices and reducing perceived threat can lead to decreased intensity and distress, even when voices do not fully remit (Chadwick & Birchwood, 1994; Morrison et al., 2014). Within the proposed model, such effects may be explained by disruption of the self-reinforcing cycle of "alien voice — fear — hostile content."

In addition, the model highlights the importance of restoring a sense of authorship and integrating internal experiences into the self-concept. Therapeutic interventions aimed at enhancing metacognitive awareness, recognizing inner speech, and identifying its emotional origins may reduce the experience of voices as external and autonomous agents.

Work with traumatic experiences acquires particular relevance in this context. If the content of voices reflects individual fears and affective residues of past experiences, trauma-informed approaches may not only reduce emotional burden but also transform the thematic

content of hallucinations themselves. From this perspective, voices cease to be viewed as meaningless symptoms and instead become interpretable as distorted forms of internal emotional dialogue.

The proposed model may also contribute to the reduction of stigma. Explaining voice-hearing experiences as the outcome of universal cognitive mechanisms intensified by fear and disrupted integration allows patients to view their experiences as understandable and meaningful, rather than as indicators of "otherness" or irreversible pathology.

In sum, the clinical value of this model lies not in replacing existing therapeutic approaches, but in conceptually integrating them and deepening the understanding of patients' subjective experience.

Limitations of the Model and Directions for Future Research

Despite the explanatory potential of the proposed model of fear-mediated misattribution of inner speech, several limitations must be acknowledged when interpreting and applying it.

First, the model is theoretical in nature and does not claim to provide a direct account of the neurobiological mechanisms underlying auditory verbal hallucinations. Although it is broadly consistent with existing findings on source-monitoring deficits and affective dysregulation, empirical studies are needed to directly test the role of fear as an active factor shaping the content of voices.

Second, the model does not posit fear as the sole possible source of negative or hostile content. Other affective states—such as shame, guilt, or depressive affect—may also contribute to the thematic structure of voices, particularly in cases where hostility is predominantly intrapunitive. Future research could clarify the relative contribution of different affective states and their interaction with fear.

Third, the model does not encompass all forms of auditory hallucinations. It appears less applicable to cases in which voices are neutral, fragmented, or minimally distressing. This limitation highlights the need for a differentiated approach to the typology of hallucinatory experiences rather than a single explanatory framework.

In addition, the neural correlates of the proposed self-reinforcing cycle remain an open question. A promising direction for future research involves examining interactions between threat-processing systems, self-referential networks, and mechanisms of sensory simulation. Functional neuroimaging studies may help clarify how affective arousal influences the perception of authorship in inner speech.

Finally, empirical validation of the model's therapeutic implications represents a crucial avenue for future investigation. Clinical studies could assess whether targeted interventions focused on fear reduction and reinterpretation of voices lead to changes in their content, intensity, and perceived hostility. Such findings would help determine the applied clinical value of the proposed approach.

In sum, the present model should be regarded as a conceptual framework that opens new directions for research rather than as a complete or exhaustive theory of auditory verbal hallucinations.

Conclusion

This article has proposed a theoretical model explaining the hostile content of auditory verbal hallucinations through a mechanism of fear-mediated misattribution of inner speech. In contrast to approaches that conceptualize fear as a secondary emotional reaction to already-formed voices, the present framework positions fear as an active generative factor shaping the content of hallucinatory experience.

Within this model, impaired recognition of the authorship of inner speech leads to its perception as alien. Given the evolutionarily grounded sensitivity to potential threat, this perception elicits fear, which in turn triggers the subconscious generation of threatening scenarios. These scenarios are then reproduced in the form of voices, creating a self-reinforcing cycle that amplifies both fear and the hostility of the content.

The model enables the integration of findings on the role of traumatic experience, individual variability in voice content, and the sensory vividness of hallucinations into a unified conceptual framework. It also supports a continuum-based perspective on the boundary between normality and pathology, viewing auditory verbal hallucinations as an extreme

manifestation of universal cognitive and affective mechanisms rather than as a qualitatively distinct mental phenomenon.

The proposed approach has both theoretical and clinical significance. It contributes to the reduction of stigmatization, deepens understanding of patients' subjective experience, and opens avenues for therapeutic interventions aimed at reducing fear and restoring a sense of authorship over inner experiences.

Overall, the model of fear-mediated misattribution of inner speech represents a productive direction for future research and may serve as a foundation for the development of empirically testable hypotheses in the psychopathology of perception and self-consciousness.

Statements and Declarations

Funding

This research received no external funding.

Conflict of Interest

The author declares that there are no financial or non-financial competing interests related to the content of this article.

Ethical Approval

This article does not contain any studies with human participants or animals performed by the author.

Data Availability

All data generated or analyzed during this study are included in this published article. No additional data are available.

References

Bentall, R. P. (1990). The illusion of reality: A review and integration of psychological research on hallucinations. Psychological Bulletin, 107(1), 82–95. https://doi.org/10.1037/0033-2909.107.1.82

Chadwick, P., & Birchwood, M. (1994). The omnipotence of voices: A cognitive approach to auditory hallucinations. British Journal of Psychiatry, 164(2), 190–201. https://doi.org/10.1192/bjp.164.2.190

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience, 10(1), 48–58. https://doi.org/10.1038/nrn2536

Friston, K. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Frith, C. D. (1992). The cognitive neuropsychology of schizophrenia. Lawrence Erlbaum Associates.

Hardy, A., Fowler, D., Freeman, D., Smith, B., Steel, C., Evans, J., Garety, P. A., Kuipers, E., Bebbington, P. E., & Dunn, G. (2005). Trauma and hallucinatory experience in psychosis. Journal of Nervous and Mental Disease, 193(8), 501–507. https://doi.org/10.1097/01.nmd.0000172480.56308.21

Johns, L. C., & van Os, J. (2001). The continuity of psychotic experiences in the general population. Clinical Psychology Review, 21(8), 1125–1141. https://doi.org/10.1016/S0272-7358(01)00103-9

Johns, L. C., Nazroo, J. Y., Bebbington, P., & Kuipers, E. (2004). Occurrence of hallucinatory experiences in a community sample and ethnic variations. British Journal of Psychiatry, 184(1), 25–30. https://doi.org/10.1192/bjp.184.1.25

Johns, L. C., Kompus, K., Connell, M., Humpston, C., Lincoln, T. M., Longden, E., Preti, A., Alderson-Day, B., Badcock, J. C., Cella, M., Fernyhough, C., McCarthy-Jones, S., Peters, E., Raballo, A., Scott, J., Siddi, S., Sommer, I. E., & Larøi, F. (2014). Auditory verbal

hallucinations in persons with and without a need for care. Schizophrenia Bulletin, 40(Suppl. 4), S255–S264. https://doi.org/10.1093/schbul/sbu005

Morrison, A. P., Wells, A., & Nothard, S. (2003). Cognitive and emotional predictors of predisposition to hallucinations in non-patients. British Journal of Clinical Psychology, 42(3), 259–270. https://doi.org/10.1348/014466503322278857

Morrison, A. P., Pyle, M., Chapman, N., French, P., Parker, S. K., & Wells, A. (2014). Metacognitive therapy in people with a schizophrenia spectrum diagnosis and medication resistant symptoms: A feasibility study. Journal of Behavior Therapy and Experimental Psychiatry, 45(2), 280–284. https://doi.org/10.1016/j.jbtep.2013.11.003