The Subcortical Confinement Hypothesis: A Neurological Model for Schizotypal Hallucinations

Jan A. Golembiewski

1. Schizophrenia Research Institute, University of Sydney

Corresponding author: Jan A. Golembiewski, greenmanarch@gmail.com

Disclosures can be found in Additional Information at the end of the article

Abstract

Introduction: Different types of hallucinations are symptomatic of different conditions. Schizotypal hallucinations are unique in that they follow existing delusional narrative patterns: they are often bizarre, they are generally multimodal, and they are particularly vivid (the experience of a newsreader abusing you personally over the TV is both visual and aural. Patients who feel and hear silicone chips under their skin suffer from haptic hallucinations as well as aural ones, etc.) Although there are a number of hypotheses for hallucinations, few cogently grapple the sheer bizarreness of the ones experienced in schizotypal psychosis.

Methods: A review-based hypothesis, traversing theory from the molecular level to phenomenological expression as a distinct and recognizable symptomatology.

Conclusion: Hallucinations appear to be caused by a two-fold dysfunction in the mesofrontal dopamine pathway, which is considered here to mediate attention of different types: in the anterior medial frontal lobe, the receptors (largely D1 type) mediate declarative awareness, whereas the receptors in the striatum (largely D2 type) mediate latent awareness of known schemata. In healthy perception, most of the perceptual load is performed by the latter: by the top-down predictive and mimetic engine, with the bottom-up mechanism being used as a secondary tool to bring conscious deliberation to stimuli that fails to match up against expectations. In schizophrenia, the predictive mode is over-stimulated, while the bottom-up feedback mechanism atrophies.

The dysfunctional distribution pattern effectively confines dopamine activity to the striatum, thereby stimulating the structural components of thought and behaviour: well-learned routines, narrative structures, lexica, grammar, schemata, archetypes, and other procedural resources. Meanwhile, the loss of activity in the frontal complex reduces the capacity for declarative awareness and for processing anything that fails to meet expectations.

Keywords: schizophrenia, hallucinations, delusions, cytoarchitecture, dopamine, glutamate, attention, perception, striatum, prefrontal cortex

Introduction And Background

Schizophrenia is a debilitating illness that can strike in any social context, anywhere in the world – although with enormous epidemiological variation [1]. People suffering schizophrenia lose their ability to interact meaningfully with society. They lose their connection with the senses that inform them about pain, hunger, comfort, and other essentials. People with schizophrenia often show bizarre disturbances in thought and speech, and they also lose their abilities to
manage day-to-day tasks. Finally, and most famously, the ability to perceive the outside world is hampered by interference from delusions and hallucinations [2-3]. The aetiology of schizophrenia is unknown, but an enormous body of evidence points to neurological dysfunction, particularly of the mesofrontal dopamine pathway and the N-Methyl-D-aspartic acid (NMDA) receptors. A Medline search brings up 7,623 articles that focus on deregulation of the mesofrontal dopamine pathway and another 1,708 on the N-Methyl-D-aspartic acid (NMDA) receptors.

**Review**

**Developing the aberrant salience hypothesis**

It is difficult to think of perception in terms of neurotransmitters and receptors. For most of us, we simply perceive what’s there to be seen, tasted, touched, smelled, and felt. Even so, we know that perception is somehow mediated by our neurochemistry: this much is demonstrated every time psychotropic drugs distort a person’s perception. Psychotropics change both the way we experience and what we experience – sometimes to the point where experience loses its foundation as a basis for ‘common agreement’. Generally speaking, medications that rein in the psychotic spectra of perception have an antipsychotic action, meaning that in one way or another they limit the supply of dopamine (DA) to the inhibitory D2 class of receptors [4].

Perhaps it is because of this relationship that a number of hypotheses over the last few years have linked D2 receptors to attentional salience [4-5]. These hypotheses find relatively cogent evidence that D2 dopamine receptors compete to draw attention away from one potential subject to another more salient one. The failure of this process is the basis of the ‘aberrant saliency’ hypothesis for the positive symptoms of psychosis: as Kapur, et al. put it, ‘the dopamine system, which under normal conditions is a mediator of context-driven novelty/saliency in the psychotic state, becomes a creator of aberrant novelty and salience... independently of cue and context...Hallucinations in this framework are thought to...reflect the abnormal salience of the internal representations of percepts, language and memories’ [4]. Put simply, the idea is that dopamine dysfunction brings awareness to internal representations without appropriate context. In the aberrant saliency hypothesis, the striatal dopaminergic pathway is used to mediate the structural components of thought and behaviour: narrative structures, lexica, grammar, schemata, meanings, archetypes, routine movements, and other procedural resources. When these are unduly stimulated, attention is drawn to them, triggering hallucinations and well-learned routines, such as the automatic actions that drive us to work in the morning or lift a cup to our lips [4, 6].

The automatic functions of the striatum contrast the functions of the anterior medial lobes of the frontal cortex (especially the highly developed anterior cingulate cortex and prefrontal cortex) further along the same mesofrontal dopamine pathway. The anterior medial frontal lobes do not process automatic and established routines but choice, insight, creativity, declarative awareness, and decisions about situations for which there are no ready responses [7-8]. In humans, the anterior medial frontal lobes are not rich in the inhibitory D2 receptors that are found in the striatum (or in the frontal lobes of many other mammals) – but in the human frontal cortex, we will find an abundance of excitatory D1 receptors [9].

**NMDA dysfunction**

Before discussing the role of dopamine receptors in hallucinations, it is important to note that the aberrant salience hypothesis (while cogent and popular) is disputed by competing theories. The foremost of these is the glutamate hypothesis: advocates of this theory argue that NMDA dysfunction precedes dopaminergic dysfunction, and therefore is likely to be a more primary causal factor in schizophrenia and schizotypal disorders than dopamine [10]. As we shall see, the subcortical confinement hypothesis appears to reconcile the early involvement of NMDA receptors with the dopamine-based aberrant saliency hypothesis.
Dopamine-moderated perception

To understand hallucinations, we must understand the schizotypal dopamine dysfunction with greater nuance. Kapur and colleagues state that the structural components of thought (which are mediated by the dopaminergic activity of the striatum) become aberrantly salient in psychosis. Under normal circumstances, the structural elements of thought only ever require only a minimal 'latent' level of awareness. No more than the subtle-most intention is required to initiate a routine process [11]. We are generally aware of what we want to say (for instance), but the words tumble out quite automatically. We are aware of where we want to drive, and if we know the basics (how to drive and where to go), we do not have to give a great deal of attention to get there. But exceptions can occur when automatic processes receive too much attention. This may cause stuttering or 'performance choking' [12] – thus, aberrant salience may cause mildly 'disorganised' symptoms, but the link between these and the vivid hallucinations experienced by schizophrenic patients and recounted by Kapur and colleagues is unclear.

It is important to understand that the neurons of the mesofrontal dopamine pathway are not maintained in fixed relationships but are interconnected in bistable mosaics. These are subject to switch rapidly from one maintainable state to another [13]. If the D1 receptor in a D1/D2 mosaic fires first, the coupled D2 receptor will drop its dopamine affinity by a factor of about 10 (to a state known as D2Low), and in the presence of certain environmental stimulants, it will turn back to the high affinity state (D2High) in a fraction of a second [14] (Figure 1).

![DIagram of D1/D2 heteromer of the mesofrontal pathway](image)

**FIGURE 1: A simplified model of a D1/D2 heteromer of the mesofrontal pathway**

A simplified model of a D1/D2 heteromer of the mesofrontal pathway. Heteromers are bistable – they modulate two functional outcomes [13]. The D1/D2 heteromer modulates sensitivity to dopamine (that is, the high affinity state (D2High) to low switch: D2Low) the same action also modulates where dopamine is synthesized in the neuron – cortically or subcortically. The transformation is rapidly triggered by external stimulation (either through direct chemical stimulation or indirectly through interneuronal connections) [15]. In this diagram, a voltage gated N-Methyl-D-aspartic acid (NMDA) afferent is shown. This is stimulated by perceptual input prior to the engagement of working memory – and therefore prior to declarative awareness [16]. Here, it is proposed that the D2Low configuration allows greater declarative awareness to be brought to events by engaging the frontal neocortex. This is useful for figuring things out, considering implications, developing normative stances on ideas, and other learning tasks. The D2High, on the other hand, inhibits the same awareness, enabling the conditions for automatic action by engaging subcortical regions, especially the striatum. The activation of automatic processes follows via other neurotransmission systems (largely glutamate, gamma-aminobutyric acid, acetylcholine, and serotonin).
The heteromer and attention

The bistability of the mesofrontal dopamine pathway allows for two types of perceptual attention, which conveniently match those of the perceptual psychology canon: top-down and bottom-up.

Top Down intention is the latent 'feed-forward' mode of attentional awareness, which is given to matches between expectations and experience – that is, the attention that is given to memorised details, likes, dislikes, familiar narrative patterns, and other routines. When I drive, the attention I give to the vehicle’s controls and other traffic is top-down. It is likely that top-down intention is mediated by tonic firing of the D2 receptors of the striatum (Figure 2). This is because top-down intention involves more or less constant engagement with established schemata and the latent level of attention that is given to automatic processes.

The term bottom-up attention is frequently used to refer to the response to any distracting stimuli, but because attentional drift can be captured by familiar interests, in this article, the term bottom-up attention is defined as the attention given to the aberrant, unexpected and unknown – to mismatches. Instead of triggering memories and automatic behaviours, bottom-up attention triggers inquiry: ‘what’s this?’ As such, it serves a very important function in perceptual feed-back from the area of the brain that is associated with conscious attention – the medial anterior frontal lobe where D1 receptors are ubiquitous.

The phasic bursting of the D1 receptors (Figure 2) is ‘critically linked’ to afferent inputs, such as afferents from the glutamatergic cells which, among other tasks appear to moderate the pre-attentive phase of bottom-up perception. These cells function tonically until errors or perceptual aberrance occur, at which point glutamate cells have their own feedback mechanism – the phasic voltage-gated NMDA receptors (Figure 1). These receptors require a significant voltage increase to fire, but when they do, they trigger an event potential. NMDA firing will engage the bottom-up attentional system - a rush of declarative awareness. But the ‘declarative awareness’ feature of an event potential cannot be a product of NMDA, because NMDA will fire even occur during coma, vegetative states and other periods of very low arousal. Dopamine activity, on the other hand, is severely reduced during vegetative states. It is also unlikely that dopamine phasis can occur except during a state of consciousness. Yet when dopamine phasis occurs, it does so with a significant increase in awareness. It is reasonable to speculate then, that the ‘awareness’ feature of attention, is likely to be mediated by dopamine.

**FIGURE 2: The proposed features of dopamine reactions**

The proposed features of dopamine reactions: D2High is proposed as the primary mediator of tonic dopamine – the regular \( \alpha - \beta \) frequency dopamine reaction patterns (cell-specific bursts) associated with specific routine tasks (a) D2Low is proposed as the primary mediator of phasic bursting – irregular, unexpected stimulus induced and high amplitude bursts (b) [5]. Phasic reactions bring attention from a latent state (the grey background) to declarative awareness (white background). Tonic depressions (c) are periods with very low dopamine activity. These follow unmet expectations.
Where the D2High receptors appear to mediate latent attention, the D1Low receptors are well placed to mediate declarative attention because of the D1 receptors are ubiquitous in the anterior medial frontal lobe, an area with established links to declarative awareness [21] and with bottom-up attention [22]. Bottom-up attention needs a lot more stimuli than top-down to ‘cut through.’ The D1 receptors are appropriate for the task: approximately 10 times more dopamine is needed to fire in a phasic action. But dopamine is in limited supply at any given moment, and thus, when the D1 receptors fire before the deeper D2 receptors, the stimuli quite literally takes possession of the mind – at the expense of routine perceptions and tasks [5] and presynaptic dopamine supply. This can be very specific: a selective D1 agonist (SKF38393) severely attenuates D2 moderated tasks such as the instinct to return an eye to its natural direction of gaze [9]. For each D2Low heteromer that is activated, there will be a significant decrease both in D2 availability and of dopamine itself.

It appears that in schizophrenia, the D1 receptors are depleted – thereby disabling the D2Low functional state. With this atrophy, the ability to process aberrance and to experience declarative awareness may also be lost – resulting in many common symptoms, including ipseity disorders (loss of a sense of self) [23-24] and mismatch negativity failures (loss of awareness of aberrant perceptual cues) [16]. Meanwhile, as Kapur and colleagues suggest, the D2High state becomes over-excited [4, 25-26], suggesting that routine and automatic functions may become aberrantly stimulated. Effectively, awareness is increasingly confined to the subcortical latent state and salience (as in declarative awareness) is decreased.

**Confirmation bias**

One of the strands of the subcortical confinement hypothesis for schizotypal hallucinations is a phenomenon known as a confirmation bias and the closely related tendency to jump to conclusions. ‘A confirmation bias is the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand’ [27]. In other words, a confirmation bias occurs because a subject favours the perceptual feedback mechanism that notices positive feedback (what I say/think/do is right/acceptable/likely, etc.) over the mechanism that notices negative feedback (what I say/think/do is wrong/awkward/unlikely, etc.). A person with a confirmation bias checking a lottery ticket against published results would be inclined to notice the positive matches but not that the recognisable digits are in the wrong order. (Note that in this context, positive and negative are not necessarily related to affective or hedonic associations – but to perception of support (positive) or invalidation (negative)).

Despite employing the same perceptual apparatus (eyes, nerves, etc.) and neurotransmitter (dopamine) positive and negative feedback employ different attentional systems. Top-down intention is used to process positive feedback and trigger mimetic perception, while bottom-up attention is used to process negative feedback and trigger deeper consideration and questioning. Top-down intention is fast because it is predictive: it literally ‘jumps to conclusions’ on available evidence [28-29]. These conclusions are resources that are already stored in memory and do not need to be reprocessed, except when experience differs from expectations (a bottom-up feedback process) [30]. Bottom-up attention is usually only engaged when something’s amiss, when events are not as predicted and do not ‘ring true’, and this process is used to resolve discrepancies and to ‘patch’ the epistemological model of the subject.

Either the overstimulation of the top-down perceptual apparatus or the under-engagement of bottom-up attention should be sufficient to cause confirmation biases, but the combination will make the problem worse and may easily lead to delusional ideas. In fact, the ‘jumping to conclusions’ trait is common to the point of being ubiquitous [27], suggesting that it is part of a normal process. But because this trait is significantly more common among paranoid schizophrenic patients than healthy controls (OR 1.54, p=0.004) [31], it seems that this
mechanism is over-regulated in schizophrenia. If the scantest shred of evidence is taken as sufficient to make a judgment, bizarre ideas will be supported by 'evidence', and these can pretty quickly become delusional beliefs. Unsurprisingly, delusions are the single most common (but not compulsory) symptom in schizophrenia [32].

The frontal dopamine neurons in schizophrenia patients are found to be thin or atrophied, but whether this is causal or a corollary effect is unknown [9, 21, 33]. These neurons are active only when the dopamine heteromers are in the D2Low state, and processing bottom-up attention or otherwise providing high-order attention to cognitive or perceptual task. Meanwhile, in the same pathways, mesofrontal dopamine is moderately upregulated subcortically – most notably in the striatum. This imbalance potentially explains some challenging mixed findings of dopamine synthesis levels in schizophrenia: most studies find moderately more dopamine synthesis in schizophrenic subjects, but some find that dopamine synthesis remains the same [34-35]. This appears to depend on methods and the schizophrenic subtype – if researchers were to study dopamine synthesis in the frontal lobes, hypothetically, it would indeed be low, whereas the opposite would be true for striatal dopamine synthesis, particularly in the paranoid schizophrenic sub-classification (DSM-IV type: 295.30), a condition that presents hallucinations and delusions as the dominant symptoms.

The imbalance described above makes the jump-to-conclusions trait understandable, and this in turn may lead to delusions. But how the same neurological imbalance may cause hallucinations is still more complex.

**Hallucinations**

A dysfunctional perceptual feedback system is already implicated in ‘the corollary discharge error hypothesis’ for hallucinations: a popular hypothesis which was first published in 1978 [36] and is still a useful model today. The idea, put simply, is that people must have some kind of feedback mechanism to inform them of their own corollary role in perception. If I watch a moving train, my eyes and head move too, thereby allowing me to fixate my gaze as if the train were stable – but if I were to ‘lose track’ of my body movement, I would experience the train as static, rather than as moving. There’s some evidence that this hypothesis has some validity, because empirical studies have located the mediator for corollary discharge. It is the D2 receptors with prefrontal afferents (i.e. the dopamine heteromers) [9].

The problem with the corollary discharge error hypothesis is that schizophrenic hallucinations are often truly bizarre and are not just illusions (the kind of errors that corollary discharge errors easily explain). In apparent opposition to this hypothesis, paranoid schizophrenic patients are actually less likely to fall prey to simple illusions than controls by at least one standard degree of deviation [37]. It is therefore not logical that the misplacement of corollary effort will make a TV broadcast personal messages or similarly bizarre symptoms in paranoid schizophrenia.

To take the next step in understanding bizarre hallucinations, it is important to first recognise that it is not the presence of, but the nature of hallucinations, that should be a concern. The differences between psychotic experiences and those experienced by healthy people (and also from those with other mental illnesses) are quantitative and qualitative and not absolute. When perceiving, we seem to predict using existing schemata and then recognise matches, much as computer recognition software does [28, 38-39]. We cannot conceivably recognise every object that we perceive afresh, because many objects are unique in one way or another, and the processing power that such a system would require would be phenomenal in any case. Predictive perception is led by the automatic intention to understand, and therefore, by the subcortical systems thought to be responsible for mimetic pattern recognition: the hippocampus/amygdal coupling [40], the striatum, and the dopamine heteromers in the D2High state. But predictive perception needs a backup attentional system to identify and process gross errors and genuine novelty; this requires the D2Low state to engage the frontal neocortex to process information as a
Because the subcortical D2 receptors are inhibitory, the moderate excess of dopamine caused by the failure of the D1 mechanism should mean an increased potential for latent attention. As this is brought to the structural elements of thought and perception, these will be activated. These are predicted to follow the basic patterns and structural resources of thought: ‘voices’, the visuo-spatial sketchpad, and familiar storylines – even in unfamiliar situations. However, because these occupy the latent attention system, yet they are not-intended, they are likely to appear invasive, as if they were intended by someone else: as if someone else was in control.

Hebb’s law of synaptic plasticity will only compound the problem. Synapses that fire together bind together. In normal circumstances, this strengthens the synapses of natural associations, but when there’s a surplus of subcortical dopamine, synapses may fire without intrinsic associations, thereby strengthening lateral and very obscure or personal associations. Ultimately, this will cause the personal ontology of resources to lateralize, so that bizarre linguistic and visuo-spatial definitions and associations will end up becoming as important as primary ones. It’s already known that regular challenges to schemata are required to maintain the differentiation between concepts (which otherwise have a tendency to blur when unchallenged over periods) [41], but the combination of decreased bottom-up attention, and with it, reductions in frontal processing and an increase of D2High firing is predicted to speed up the process and effectively scramble the ontology – even over a relatively short time period.

In keeping with this hypothesis, schizophrenic hallucinations are not particularly creative. They are tediously repetitive, familiar to the point of utter impoverishment and return to the same delusional meanings regardless of circumstances [42]. Thus, the experience of having the TV deliver personal messages is likely to be followed by the same narrative apparently from the mouths of radio DJs and random pedestrians on the street: the messages in all these circumstances are the same because they are predictive representations of the same schemata, the different sources reflect the diminished input of ‘reality’. Hallucinations of this pattern are commonly reported with paranoid psychosis [32, 43].

Bottom up attention triggers creative consideration rather than action. It therefore requires both the inhibition of action and the inverse – the excitation of attention. In contrast, automatic action requires the inhibition of attention to a latent level (because too much awareness prohibits automatic fluency – the choking phenomenon). It is cogent that that the D1 receptors of the frontal cortex are excitatory, where the D2 receptors of the striatum are inhibitory: dopamine neurons do not directly moderate behaviour, they mediate attention. (Actions are largely modulated by the excitatory glutamate receptors.)

A loss of frontal connectivity will have other phenomenological outcomes also: affective flatness and the phenomenal loss of gross awareness. This is because attention is more or less confined to a latent state in the striatum. And other bizarre experience, because of having ones’ attention more or less confined to the striatum, is that occasional flashes of lucidity may reveal that ones’ actions are going on automatically, as if the body were a puppet of an invisible puppeteer or being run by a machine perhaps. All these symptoms are very common and frequently the most terrifying for schizophrenia patients. They are classed as ‘bizarre delusions,’ and are considered to be especially characteristic of schizophrenia [3].

**Schizotypal focus, noise theories, and delusions of control**

Kapur, et al. suggest [4, 26] that the overexcitement of D2 (overactive D2 is linked to D2High functional states [15]) may cause distraction, which in turn causes many positive symptoms. This is unlikely, because it is hypothesised that it is D2Low and NMDA (not D2High) that moderates the attention given to aberrant and distractible stimuli. Furthermore, it is proposed that D2High mediates the attention that is given to normal, intended and expected routines, so if anything,
the opposite should be true: In schizophrenia, patients are predicted to a) experience diminished declarative awareness, b) have a diminished ability to notice aberrant distractions (due to a reduction of bottom-up attention), and c) feel that expected routines seem to function more autonomously (due to an increase of top-down intention).

This prediction is in keeping with the known phenomenology of schizophrenia: patients frequently report that during entrenched schizophrenia they experience an erosion of awareness [23]; while delusions of external control are very common: patients are able to do routine tasks, but feel as though they are being controlled autonomously – as if they were a robot [42, 44]. Many bizarre hallucinations can be traced back to these delusions: the feeling that ‘I am not in control’ of my own actions and thoughts. Symptoms like this are distinctly bizarre, and may also be classed as somatic delusions.

It is not only Kapur, et al. [4, 26] who assert the common heuristic belief that schizophrenic symptoms may be caused by distraction – whether it be the distraction of neural ‘noise’ or of internal stimuli (a cursory Medline review turned up 2,000 papers referring to similar ideas, including the 1994 edition of the DSM). However, under close examination, it seems that this is possibly not a symptom of schizophrenia: When tasks are a) well-defined, b) contain no unexpected challenges, and c) do not require insightful solutions or otherwise engage the bottom-up attentional system, tasks should be and are easier for paranoid schizophrenic patients. The relative increase in D2High should mean improved directed, top-down intention than for controls, and this means that paranoid schizophrenics should be less distractible than controls. In support, there is a growing body of reports of better performance by the paranoid schizophrenic patients than controls for simple, well-defined perceptual tasks [37, 45-46].

**Conclusions**

Hallucinations vary from condition to condition. Even in healthy populations, hallucinatory experience is relatively common. Healthy hallucinations tend to be non-invasive, single-modal and generally affectively positive [47]. Hallucinations in dementia have similar profile but are generally less ephemeral and more complex [48]. Manic hallucinations appear to be triggered by events and opportunities that have a real basis [6] and schizotypal hallucinations are distinguished by their intensity, multimodality and by the repetitive and delusional nature of their themes and narratives. This last type hallucination may be explained by the relative overstimulation of the striatal D2 vs. the frontal D1 receptors within in the D1/D2 heteromers of the mesofrontal dopamine pathway.

This kind of hallucination is the product of an increased tendency to accept erroneous perceptions combined with a decreased ability to spot and process errors in perception, poor fact checking processes, and a loss of definition in the patient’s ontology. This pattern causes a range of bizarre experiences, all of which are very familiar symptoms of schizophrenia. Among these is a tendency to perceive what a patient most expects. These expectations only become more distorted as the ontology loses definition.

The fact-checking mechanism (bottom-up attention) is one of the functions of the frontal region of the dopamine pathway. This mechanism involves the activation of the low-affinity mode of the D1/D2 heteromers, which draw approximately 10 times the amount of available dopamine. When this system shuts down or slows, the heteromer is left in a stable high affinity state, and with a surplus of dopamine to synthesize. This causes top-down attention to become upregulated, resulting in the over-stimulation of familiar and expected schemata (phonological and visuo-spatial resources). This causes primary perceptual errors. With the loss of the fact-checking feedback system, perceptual errors are met with an increased sense of surety, even though they are false. The result is that experience genuinely becomes at once familiar, repetitive and bizarre.
Additional Information

Disclosures

Conflicts of interest: The authors have declared that no conflicts of interest exist.

Acknowledgements

This work was supported by the Schizophrenia Research Institute, utilizing infrastructure funding from NSW Department of Health.

References