

The Salience Theory of Delusions

Theoretical approaches to delusions often seem to have a curiously half-hearted quality. No one has ever bothered to test Maher's theory. Theory of mind abnormality and probabilistic reasoning bias have both run into significant experimental difficulties, and there has been little enthusiasm for addressing them. Important features of delusions such as their impossibility and imperviousness to reason are generally given only token consideration, and in most cases referential delusions are ignored altogether.

Such criticisms do not apply to a further approach to delusions, Kapur's (2003) salience theory. This has been so influential that it recently led to a serious attempt to rename schizophrenia as salience dysregulation disorder (van Os, 2009) – despite the fact that it has only very limited power to explain any other class of symptom besides delusions. Its power derives partly from the fact that it provides an intuitive explanation of what this book collectively refers to as referential delusions. Another source of strength is the central role it accords to dopamine which, despite its many setbacks, is still an important player in schizophrenia research. Nor does it hurt that the principal means of testing the theory involves stepping into the glamorous if not always easily understandable world of functional brain imaging.

Clearly, such an important theory demands detailed and critical consideration, to make sure that its claims hold up theoretically and to examine how far they are supported by evidence. There is also another reason for engaging in such an exercise. This is that the theory only tells half the story. In particular, it will be argued that, while the salience theory's explanation of referential delusions is compelling, what it says about propositional delusions is no more substantial than in any other theory of delusions. Another aim of this chapter, therefore, will be to explore what can be done to repair this weakness. As it turns out, efforts in this direction go back to well before the salience theory appeared on the scene and continue right up to the present time.

Introducing the Salience Theory

The salience theory starts with an assumption. This is that the dopamine hypothesis of schizophrenia is correct, specifically that a functional excess of the neurotransmitter underlies the positive symptoms of the disorder. If this is so, then it follows that these symptoms should be understandable in terms of what is known about the normal function of dopamine. For Kapur (2003), this function was the way in which it acts to assign motivational and reinforcing value to stimuli that are associated with reward (see Chapter 6). Pathologically increased dopamine transmission would then lead to a release of dopamine outside the proper context, which in turn would cause neutral stimuli to inappropriately acquire significance for behaviour, or as Kapur termed it, aberrant salience.

The subjective correlate of saliences being created when there ought not to be any might be that the individual would start to wrongly experience neutral events as important. Such a hypothetical state, Kapur (2003) noted, matched closely with the descriptions that schizophrenic patients gave of the earliest stages of their illness, as recorded in a spate of studies carried out in the 1960s. These included statements such as: 'I developed a greater awareness of ... My senses were sharpened. I became fascinated by the little insignificant things around me,' and 'Sights and sounds possessed a keenness that he had never experienced before' (Bowers & Freedman, 1966); 'It was as if parts of my brain awoke, which had been dormant' (McDonald, 1960); or 'My senses seemed alive ... Things seemed clearcut, I noticed things I had never noticed before' (Bowers, 1968). Related to this there might also be a feeling that the world was changing in a puzzling way that required explanation. This was also evident in the patients' accounts, for example, 'I felt that there was some overwhelming significance in this' (McDonald, 1960), and 'I felt like I was putting a piece of the puzzle together' (Bowers, 1968).

Delusions – by which Kapur (2003) meant propositional delusions in the terminology of this book – were proposed to be the result of the individual's effort to make sense of the experience of aberrant salience as it was repeated over days, months or years:

Delusions in this framework are a 'top-down' cognitive explanation that the individual imposes on these experiences of aberrant salience in an effort to make sense of them. Since delusions are constructed by the individual, they are imbued with the psychodynamic themes relevant to the individual and are embedded in the cultural context of the individual. This explains how the same neurochemical dysregulation leads to variable phenomenological expression: a patient in Africa struggling to make sense of aberrant saliences is much more likely to accord them to the evil ministrations of a shaman, while the one living in Toronto is more likely to see them as the machinations of the Royal Canadian Mounted Police.

Kapur (2003) did not rule out the possibility that additional factors might contribute to the process whereby fully formed delusions developed out of the initially amorphous experience of aberrant salience. These could include a jumping to conclusions cognitive style and poorly developed theory of mind skills, and perhaps aspects of the patient's personality as well.

Kapur (2003) considered that delusions of reference and misinterpretation also arose as part of the attempt at explanation. This drove the patient to search for further confirmatory evidence within the evolving delusional framework, 'in the glances of strangers, in the headlines of newspapers, and in the lapel pins of newscasters'.

This then is the theory. It is not difficult to see why it has become so influential: it provides, perhaps for the first time in the history of schizophrenia research, a simple and intellectually satisfying link between a symptom of the disorder and an underlying biological brain disturbance. If dopamine causes neutral stimuli in the environment to acquire significance – and following the work of Schultz (1998) described in Chapter 6, there seems no doubt that it does – then it seems highly probable that a dopamine excess will give rise to a state which resembles delusional mood. Although not explicitly part of Kapur's theory, there does not seem to be any particular difficulty extending the same concept to encompass all other types of delusion whose central phenomenological feature is an abnormal feeling of significance.

Where the theory fares less well is in its explanation of propositional delusions. The main proposal offered here is that this class of delusions represents an attempt by the individual

to make sense of the experience of aberrant saliency. As such, this part of the theory is not obviously an advance over what Maher (1974) proposed 40 years ago (see Chapter 5). To be sure, the saliency theory avoids one problem Maher ran into, that of having to invoke a 'free-floating feeling of significance' to explain how delusions arise when there are no accompanying perceptual abnormalities. On the other hand, in exactly the same way as Maher's approach, the theory struggles to explain several phenomenological features of propositional delusions, especially the fact that they tend to be bizarre and fantastic.

Finally, and perhaps most importantly, the saliency theory makes the prediction that propositional delusions will always be preceded by delusional mood and/or other referential delusions. This is something that, as Chapters 1 and 3 make clear, is by no means always the case in practice.

Can the Saliency Theory Be Extended to Explain Propositional Delusions?

Before Kapur introduced the saliency theory in 2003, a few other authors had tried to link dopamine to delusions. One of these was Beninger (1983) who, in the course of a review of the role of dopamine in behaviour, suggested that an overstimulation of dopamine receptors might have the consequence that schizophrenic patients would lose their ability to ignore irrelevant stimuli, and that paranoia or delusions of grandeur could represent cognitive elaborations of the apparent meaningfulness of these stimuli. The present author (McKenna, 1987, 1991) proposed something quite similar as one part of an attempt to link dopamine to a wide range of schizophrenic symptoms.

But it was another author who came up with the first concrete proposal for how a hyperdopaminergic state might give rise to propositional delusions. Miller (1984) argued that the associative processes of learning, i.e. the formation of links between stimuli and stimuli (classical or Pavlovian conditioning) and between stimuli and responses (instrumental learning), might also take place at a higher level, leading to the formation of cognitive associations. If so, he speculated, the role of dopamine would in effect be to set the threshold for inductive inference:

For any step of inductive inference there must be a threshold, or set point, comparable in some ways to a criterion of significance in a statistical argument. Below this threshold associational links are rejected as coincidental. Above the threshold they are 'above chance', and, therefore, accepted as real.

A functional increase in dopamine would lower this set point, causing a 'hyperactivity of inductive inference'. This would lead to more cognitive associations than normal being formed, many of which would be spurious. To the extent that these associative links could be equated with conceptual thinking, the result would be propositional delusions.

The idea of dopamine exerting effects on higher cognitive function was controversial enough, and Miller's proposal that it somehow acted to set the threshold for inductive inference was a leap in the dark. But as it happened, his idea resonated with those in a book that had just been published to considerable acclaim (one reviewer compared it to Newton's *Principia Mathematica*), which argued that animals routinely do something very similar to making inductive inferences. This was Gray's (1981) theory of hippocampal (or as he preferred to call it, septo-hippocampal) function, and it was destined to play a significant role in the subsequent evolution of thinking about the role of dopamine in delusions.

Gray's (1981) theory was a highly complicated tour de force that integrated an enormous number of animal behavioural findings on the hippocampus and septal area with almost as much neuroanatomy and neurophysiology. However, at its core the proposal was simple: the hippocampus acts as a comparator, matching, on a moment-to-moment basis, 'actual', i.e. the currently perceived state of the world, with 'expected', or predictions about what ought to be experienced after the animal performs the next step in the sequence of motor acts it is carrying out. Gray noted that the hippocampus was well equipped to receive information about the actual state of the world via its major afferent pathway from the entorhinal cortex; this was known to be a destination for highly analysed sensory information in all modalities. He proposed that the predictive function was accomplished by means of the classical Papez circuit running from the hippocampus to the cingulate cortex (and also the prefrontal cortex in primates) via the mammillary bodies and the thalamus, before projecting back to the entorhinal cortex.

The main way in which the system exerted an effect on behaviour was through what Gray (1981) called behavioural inhibition – a sudden interruption of the sequence of motor responses currently being executed when a mismatch between actual and expected was detected. How the hippocampus managed to gain access to motor systems to produce behavioural inhibition was something of a mystery at the time his book was published. However, a year later an efferent projection from the subiculum (the main output area of the hippocampus) to the ventral striatum was described (Kelley & Domesick, 1982), something that filled the role perfectly.

The septo-hippocampal system could also operate in a 'just checking' mode, when observed matched with expected. In this case the sequence of motor responses being elaborated was allowed to proceed without interruption. When the animal found itself in a new environment, where no predictions could be made, the system fell into yet another, 'exploratory' mode (see Box 8.1).

Box 8.1 Gray's Proposed Modes of Septo-hippocampal Function (Gray, 1981)

Scenario 1: Exposure to a Novel Environment

The animal is in a totally new environment. Under these conditions there can be no predictions for the comparator to match against current experience. It follows that the only task the septo-hippocampal system can perform is gathering information that will make subsequent prediction possible. Information about the novel events is passed on for storage elsewhere.

Scenario 2: Just Checking

There exists a set of expectations which continue to be verified by current sensory input. Under these conditions the system exercises no control over behaviour.

Scenario 3: Mismatch

The comparator detects a mismatch between expected and actual events. In this situation the septo-hippocampal system assumes control over behaviour. Major features of this mode of operation include the active inhibition of motor behaviour and the institution of information-gathering strategies with the aim of resolving the discrepancy. These two together – analysis and exploration – constitute a process analogous to hypothesis generation and testing. Other consequences include tagging the motor programme as 'faulty, needs checking', and executing it more cautiously on future occasions.

Scenario 4: Disengagement

After the discrepancy has been resolved behavioural control passes back to other systems which may, however, now receive updated information as a result of the activities of the septo-hippocampal system.

Two other features of Gray's (1981) theory were also important. One was that information about matches and mismatches was proposed to be passed on to other brain regions where it was used to modify future predictions about what was to be expected in that particular environment (and also to form new predictions in the case of a novel environment). The other was that two modulatory neurotransmitters, noradrenalin and serotonin, which were at the time known to innervate the hippocampus, acted to label stimuli that were novel or associated with aversive events as 'important, check carefully', and to bias the system towards behavioural inhibition. In fact, a large part of the *raison d'être* of Gray's theory was for him to be able to argue that dysfunction in one or both of these transmitters systems would lead to overly frequent behavioural inhibition, which in turn formed the basis of anxiety disorders. He also speculated that the environmental checking that was instituted after behavioural inhibition took place might serve as a model for obsessive-compulsive disorder.

It seemed only a matter of time before the theory would also be applied to schizophrenia, and ten years later Gray and several co-workers (Gray et al., 1991) duly did so. Their main innovation was to add dopamine, which by now was by now also known to innervate the hippocampus, to the model of septo-hippocampal function. Unlike noradrenalin and serotonin, this neurotransmitter was proposed to operate in the system's 'just checking' mode, where it acted to facilitate the transition from one step in a motor programme to the next when no conflict between observed and expected was detected. Excess dopamine, Gray et al. (1991) argued, would result in a special kind of disorder in motor programming whereby one or more responses became inappropriately dominant. (Although the authors said nothing about reduced dopamine in their 1991 article, an interesting aside is that the consequences of this would presumably be something not dissimilar to the akinesia and bradykinesia of Parkinsonism.)

Motor responses becoming inappropriately dominant is a long way from delusions, and Gray et al.'s (1991) main suggestion with respect to these and other psychotic symptoms was that the disturbance caused by a dopamine excess might also extend to the programming of selective attention. More broadly, they also felt that their proposal was consistent with a suggestion for understanding positive psychotic symptoms that had been made a few years earlier by one of the authors of the article (Helmsley, 1987), that they reflected a 'weakening of the influence of stored memories or regularities of previous input on current perception'. If nothing else, this proposal has the dubious distinction of being one of the least testable hypotheses ever formulated in schizophrenia research.

Years later, after the publication of Kapur's salience theory, Gray (2004) wrote a letter claiming that he and his co-authors, in their 1991 article, had themselves proposed that aberrant salience would be a further consequence of a dopamine excess affecting the septo-hippocampal system. As far as the present author can tell, there is no statement to this effect in the article. Gray (1998), however, did note this possibility in a subsequent paper.

Today, Gray's theory languishes in obscurity, eclipsed by a rival theory that he did his best to disparage in his 1981 book, O'Keefe and Nadel's (1978) cognitive map proposal

(which ultimately won one of its authors the Nobel Prize). Nevertheless, the concept of a brain system that compares actual and expected and whose dysfunction gives rise to delusions lives on in the work of a loosely knit group of researchers which includes but is not limited to Corlett, Fletcher, Friston and Frith (e.g. Fletcher & Frith, 2009; Corlett et al., 2009; Corlett et al., 2010b; Adams et al., 2013). For these authors, forming predictions is a general mode of brain function, which is carried out based on Bayesian statistical principles and which underlies not only learning but also perception and in all probability other cognitive processes as well. Equally important is prediction error, to which this process is inextricably linked: predictive models form the basis for the generation of prediction errors, and prediction errors in turn modify the predictive model. At times the theory is almost explicitly Grayian in tone: Corlett et al. (2010b) suggested that when an organism experiences an event that violates predictions, an orienting system is activated which enables the acquisition of new data for a new predictive model. In contrast, when the event matches what is predicted, the current predictive model of the world is strengthened.

With respect to the formation of delusions, Corlett et al. (2010b) agreed with Kapur (2003) that:

during the earliest phases of delusion formation aberrant novelty, salience or prediction error signals drive attention toward redundant or irrelevant environmental cues, the world seems to have changed, it feels strange and sinister...

But now, the occurrence of erroneous prediction errors also leads to a modification of the predictive model of the relevant aspect of the world:

... such signals and experiences provide an impetus for new learning which updates the world model inappropriately, manifest as a delusion.

To which Fletcher and Frith (2009) added that the model of the world can never be successful because it can never eliminate the prediction error. The rogue signal persists however many attempts are made to accommodate it, and so the predictive model deviates more and more from reality.

With this, via a circuitous route involving thresholds for inductive inference and a defunct theory of hippocampal function, the salience theory has arrived at its current state of the art. It now has the benefit not only of an intuitive account of referential delusions, but also something that seems close to a credible explanation of propositional delusions. This qualification 'close to' needs to be appended, because the theory still predicts that propositional delusions will always be preceded by delusional mood and/or referential delusions. It also depends on there being a mechanism whereby dopamine (or possibly some other neurotransmitter) directly influences cognition. As far as the present author is aware, there is as yet no evidence for such a proposal.

Testing the Salience Theory

Increased Dopamine in Schizophrenia

With its simple and intuitive explanation of referential delusions and the strong hints it may be sooner or later also be able to provide an account of propositional delusions, the salience theory certainly talks a good game. But as with any other theory, the only thing that ultimately counts is whether it can gain experimental support. One relevant line of experimental

evidence already exists in the form of the dopamine hypothesis of schizophrenia itself – if this were proved to be correct, it would be a good first step towards the salience theory also being correct, particularly since dopamine appears to have a particular role in positive symptoms.

Unfortunately, whether the dopamine hypothesis is right or wrong has become something of an eternal question, whose definitive proof one way or the other always seems just out of reach. The proposal was first intensively investigated following the discovery, made more or less simultaneously by three different groups of investigators, that post-synaptic dopamine D2 receptor numbers in the basal ganglia were increased in the post-mortem brains of schizophrenic patients (see Seeman, 1987). It was quickly realized that this finding did not in itself constitute proof of anything, because almost all the patients in these studies had been treated with antipsychotic drugs in life, and antipsychotic treatment itself can cause D2 receptor numbers to increase (as a compensatory response to their blockade by these drugs). What was needed were studies examining D2 receptor numbers in never-treated schizophrenic patients. Although challenging, this goal was achieved some years later by combining functional imaging with use of a tracer that attached to D2 receptors (i.e. a radioactively labelled antipsychotic) in living patients who had received little or no previous drug treatment. The first study (Wong et al., 1986), carried out on a group of chronic schizophrenic patients who for one reason or another had never been given drug treatment, found an approximate doubling of basal ganglia D2 receptor numbers compared to healthy controls. The second (Farde et al., 1990), carried out on drug naïve first-episode patients, found no difference. For a time the fate of this version of the dopamine hypothesis hung in the balance, but eventually a series of further studies (Martinot et al., 1990; Hietala et al., 1994; Pilowsky et al., 1994) all supported the negative finding of Farde et al. (1990).

The second wave of studies took a different tack and tested the hypothesis that synaptic release of dopamine, as provoked by amphetamine, is increased in schizophrenic patients. Three studies, two by the same investigators (Laruelle et al., 1996; Abi-Dargham et al., 1998) and one by an independent group (Breier et al., 1997) all had positive findings. These studies were carried out in drug-free patients; however, only a minority of them were drug-naïve. Is it possible that the previous antipsychotic treatment in the majority of patients could have caused an increase in amphetamine-stimulated dopamine release? The answer appears to be yes: the technique used for measuring dopamine release in these studies depended on the displacement of radioactively labelled ligand from post-synaptic D2 receptors. As Laruelle et al. (1999) acknowledged, this meant that the differences found could conceivably have been due to increased dopamine binding to these receptors, caused by the patients' previous treatment, rather than by increased amphetamine-stimulated dopamine release per se. The authors of these studies had forgotten a basic principle of schizophrenia research: in order to convince sceptics (not to mention the many who are constitutionally opposed to any biological theory of the disorder), it is necessary to demonstrate that any alleged brain abnormality is present beyond a shadow of a doubt.

The third and current wave of studies was ushered in by a study that examined the dopamine hypothesis from yet another angle, of whether there is increased production of the neurotransmitter in schizophrenic patients. This study avoided the problem of prior antipsychotic treatment by adopting a strategy of examining patients who had prodromal symptoms of schizophrenia rather than the disorder itself. Howes et al. (2009) compared 24 patients with the so-called at-risk mental state and 12 matched healthy controls. The

patients all showed evidence of attenuated psychotic symptoms and four had previously experienced brief, self-limiting episodes of psychosis. Only one had received treatment with antipsychotics and this was omitted for 24 hours before scanning. All subjects underwent functional imaging using a radioactively labelled form of the dopamine precursor, l-DOPA, and levels of radioactivity in the striatum in the two groups was compared under blind conditions.

The prodromal patients showed a 6.3 per cent increase in l-DOPA uptake compared to the controls in the whole striatal region, a significant difference. When the striatum was divided up into 'motor', 'associative' and 'limbic' subregions (the last corresponding to the ventral striatum), the elevation was found to be restricted to the associative sector. A small group of seven patients with schizophrenia (three drug-free, four treated) also showed a similar increase in l-DOPA uptake.

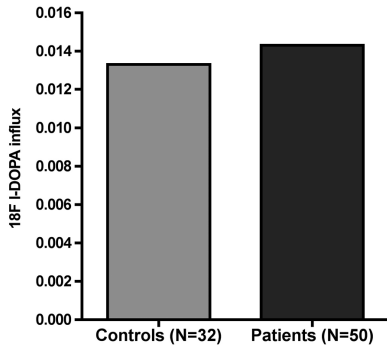
Howes and co-workers' subsequent studies have had mixed fortunes. The original finding was replicated in a second cohort of 26 high-risk subjects and 20 healthy controls (Egerton et al., 2013). The findings for both groups combined are shown in Figure 8.1. In a three-year follow-up of some of the members of both cohorts (Howes et al., 2011a), it was found that the nine who went on to develop full-blown psychosis (schizophrenia in four, schizophreniform psychosis in one and mania with psychotic symptoms in one) had significantly higher baseline levels of striatal dopamine uptake than those who did not. However, this result was only achieved after six high-risk individuals were removed from the analysis on the rather shaky grounds that they also had a diagnosis of schizotypal personality disorder. When Howes et al. (2011b) directly compared l-DOPA uptake before and after the onset of psychosis in eight patients, there was no significant increase in the striatum as a whole, nor in the limbic or associative sectors; however, a significant increase was seen in the sensorimotor sector. A summary of these latter findings is also shown in Figure 8.1.

Reward-Associated Ventral Striatal Activation in Psychosis

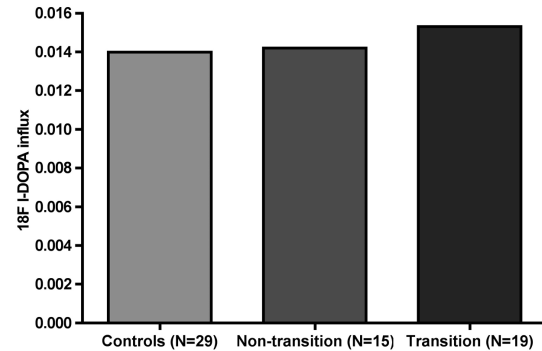
Whether the dopamine hypothesis of schizophrenia can be considered proved as a result of the last two waves of investigation is undecided – attitudes currently range from self-satisfied complacency to world-weary cynicism – but even if it is, this does not automatically mean that the salience theory is also correct. To establish this, and once again convince what will no doubt be a legion of sceptics, some way needs to be found to show that patients with delusions attribute salience abnormally.

Fortunately, such a way exists. By the end of the 1990s, functional imaging studies had demonstrated that the experience of reward, ranging from receiving a small amount of fruit juice and seeing attractive faces at one end of the spectrum, to viewing erotic videos and being administered cocaine at the other, produced a pattern of activation in the brain (McClure et al., 2004). The regions activated were broadly similar to those known to be involved in reward in animals, including the ventral striatum, the amygdala and an area encompassing the orbitofrontal and ventromedial prefrontal cortex. Then Knutson and co-workers (Knutson et al., 2000, 2001a, 2001b) devised a functional magnetic resonance imaging (fMRI) paradigm involving one of the most reliable, powerful and easy to manipulate rewards of all, money.

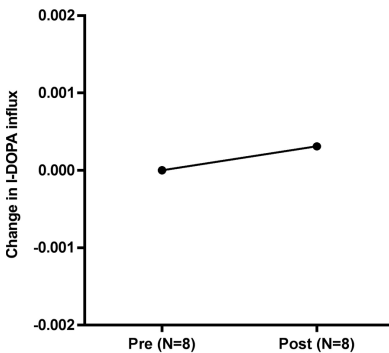
A representation of their paradigm, the monetary incentive delay (MID) task is shown in Figure 8.2. Subjects have to perform a reaction time task (pressing a button when they see a white square before it disappears) whose difficulty is individually adjusted during a



I-DOPA uptake in high-risk patients compared to controls (combined data from Howes et al. 2011 and Egerton et al. 2013).



I-DOPA uptake in controls and high risk patients who progressed and did not progress to psychosis (from Howes et al. 2011a).



Change in I-DOPA uptake in high risk patients who progressed to psychosis (from Howes et al. 2011b)

Figure 8.1 Howes and co-workers' studies of dopamine synthesis in patients with the at-risk mental state.

Source: From Egerton, A., et al., 2013. Presynaptic striatal dopamine dysfunction in people at ultra-high risk for psychosis: findings in a second cohort. *Biological Psychiatry*, 74, 106–112; Howes, O. D., et al., 2011a. Dopamine synthesis capacity before onset of psychosis: a prospective [18F]-DOPA PET imaging study. *American Journal of Psychiatry*, 168, 1311–1317, reproduced with permission; Howes, O., et al., 2011b. Progressive increase in striatal dopamine synthesis capacity as patients develop psychosis: a PET study. *Molecular Psychiatry*, 16, 885–886.

training phase so that they are successful approximately two-thirds of the time. On some trials, the task is preceded by a cue, for example a circle, which signals that they will win a certain amount of money if they perform the reaction time task successfully. Other trials are preceded by a different cue, for example a triangle, which indicates that successful performance will have no monetary consequences. Feedback about whether they have won

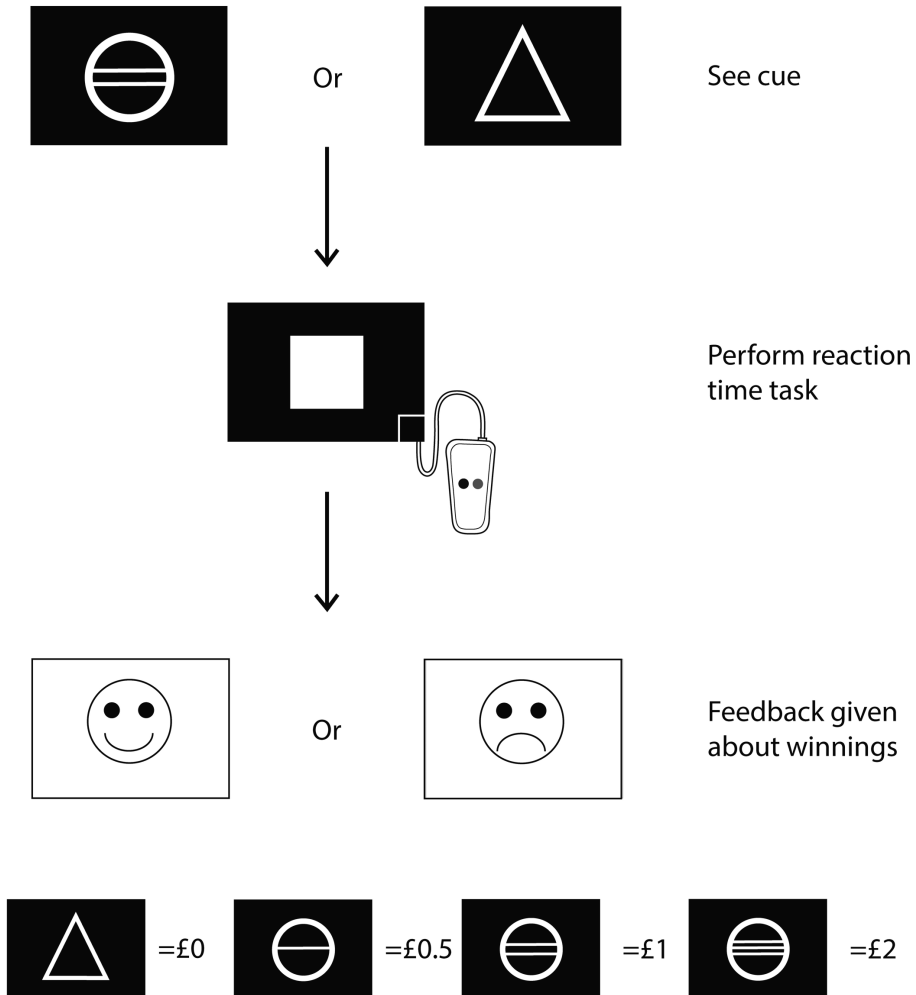


Figure 8.2 The monetary incentive delay (MID) task.

is presented immediately after the response is made. Activation in response to the reward signalling cue compared to the neutral cue provides a measure of the extent to which different brain regions respond to salience.

In many versions of the task the amount of money that can be won on a particular trial varies, and this is indicated, for example by the number of bars superimposed on the cue. There are many other variations of the task – in some, rather than being pretrained, the subjects have to learn the predictive values of the cues by trial and error while being scanned, and in others there is no interpolated reaction time task. These and other modifications make it possible to also measure reward prediction error.

In their first study, Knutson et al. (2000) examined activations in a number of predetermined regions of interest (ROIs): the nucleus accumbens, the caudate nucleus, the putamen, the thalamus, the anterior cingulate cortex and the medial frontal cortex. Twelve healthy



Figure 8.3 Meta-analysis of 25 voxel-based fMRI studies of reward anticipation. Studies included those which used the MID task, as well as other tasks where the reward was monetary, and where the analysis covered the whole brain.

Source: Jauhar, S., Solanes, A., McKenna, P. J., & Radua, J., unpublished data.

subjects were found to show significant cue-related activation in the caudate nucleus and putamen and the medial prefrontal cortex, though not in the nucleus accumbens. In later studies (e.g. Knutson et al., 2001a, 2001b, 2005; Bjork et al., 2004) they replaced ROI analysis with the so-called whole-brain approach which compares the activity of every voxel in the brain (or a proportion of it) in the two conditions, and generates a map of significant differences. These studies additionally documented activation in the ventral striatum.

Jauhar et al. (unpublished) meta-analysed these and other voxel-based fMRI studies of monetary reward anticipation. The findings are shown in Figure 8.3. There was significant activation in large areas of the basal ganglia, including both its dorsal and ventral sectors. This finding tends to support Schultz's (1998) findings in monkeys described in Chapter 6, that dopaminergic neurons coding reward prediction error are distributed throughout the

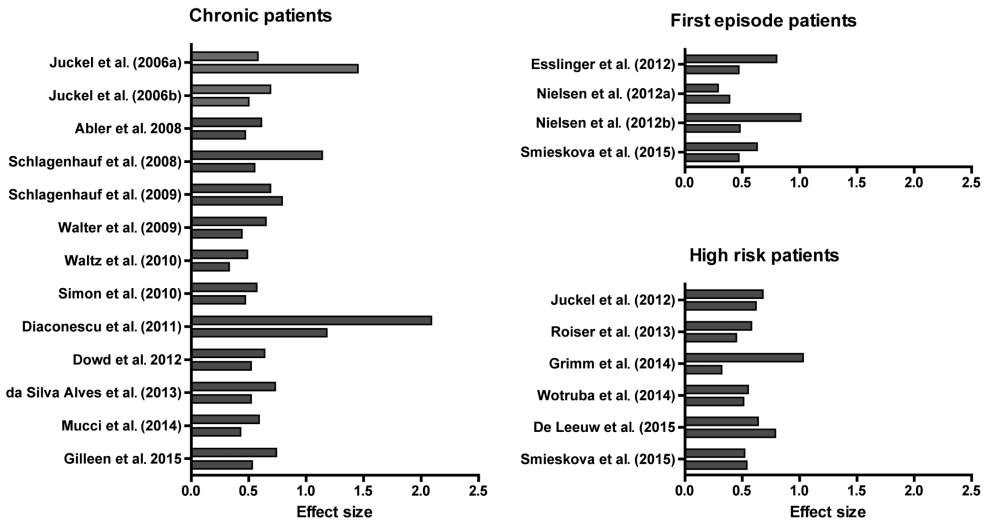


Figure 8.4 Radua et al.'s meta-analysis of fMRI studies examining ventral striatal activation in response to anticipation of monetary reward in schizophrenia, first episode psychosis and those at clinical or genetic high risk. Upper and lower bars in each study are for the left and right ventral striatum.

Source: From Radua, J., et al., 2015. Ventral striatal activation during reward processing in psychosis: a neurofunctional meta-analysis. *JAMA Psychiatry*, 72, 1243–1251; raw data kindly provided by Quim Radua.

striatum, not just its ventral striatal sector. A large and well-defined cortical area encompassing the anterior and middle cingulate cortex and other parts of the medial frontal was also activated, again in line with animal studies. The third main area that was activated was the bilateral insula, a cortical region whose function remains uncertain. Finally, activation was seen in the midbrain, reasonably close to but not actually involving its dopaminergic regions.

The way was now clear to directly test the hypothesis that there is aberrant salience in schizophrenia, and to determine whether it is associated with presence of delusions. Over 20 such studies have been carried out so far. These have examined medicated patients with schizophrenia, as well as samples of first-episode patients, some of whom were drug free or drug naïve, and also high-risk subjects. Radua et al. (2015) meta-analysed 23 such studies which employed an ROI placed in the ventral striatum. The pooled effect size was 0.50 for the left nucleus accumbens (in the medium range) and 0.70 on the right (in the medium to large range). As shown in Figure 8.4, a notable feature is that the effect, although individually variable, is in the same direction in all studies. The bad news is that the direction is the wrong one: patients with schizophrenia, first-episode psychosis and the at-risk mental state show reduced ventral striatal activation in response to reward-predicting stimuli, rather than the increased activation that the salience theory requires.

Eight studies in Radua et al.'s (2015) meta-analysis used measures of reward prediction error rather than just measuring the difference in activation between reward-predicting and neutral cues. The pooled findings were again in the direction of this being lower in patients than controls. Six studies examined the relationship between ventral striatal activation and positive symptoms. No significant association was found, although the authors cautioned that this result might not be reliable due to the small number of studies and also the heterogeneity among them.

Conclusions: Has the Salience Theory Lived Up to Its Promise?

The salience theory can with some justification be regarded as a milestone in the history of delusions research. It is the first theory to link delusions to an underlying brain abnormality. It also provides a highly intuitive link between what is proposed to happen at the neurobiological level and a key aspect of the phenomenology of the symptom, the pervasive feeling that neutral events are significant to the patient. It is not surprising, therefore, that it has captured the imagination of researchers (although why this went as far as trying to rename schizophrenia as salience regulation disorder is something that may leave future historians of psychiatry scratching their heads).

In its original form, as articulated by Kapur (2003), the salience theory had an Achilles heel, in that it offered very little in the way of an explanation for propositional delusions. Since then (and to some extent beforehand) this weakness has been recognized and the work of authors like Corlett, Fletcher, Friston and Frith currently seems to go a considerable way towards remedying it. What they propose, though, comes at the price of having to postulate that dopamine (or possibly some other neurotransmitter) has a direct influence on the cognitive processes that underlie concept formation. Another problem is that the modified theory still on the face of it predicts that propositional delusions will always be preceded by delusional mood.

At the experimental level, is aberrant salience an example of a beautiful theory destroyed by an ugly fact? At first sight it certainly looks that way, with what must be one of the most consistent findings in the history of schizophrenia research indicating that patients with schizophrenia show reduced rather than the predicted increased reward cue-related ventral striatal activity. However, unlike a finding of no change, this leaves the theory with some room for manoeuvre. It could be, for example, that salience attribution tends to be generally reduced in schizophrenia, perhaps related to negative symptoms, and this masks an increase in patients with delusions and other active psychotic symptoms. This is not a particularly strong position to take, given that Radua et al.'s (2015) meta-analysis revealed no hint of a correlation between ventral striatal activation and positive symptoms. Or it could be that simply comparing activations to reward-associated stimuli between patients and controls is the wrong approach to take, and reward prediction error is what needs to be measured. Once again, however, there is little comfort for this view in Radua et al.'s (2015) meta-analysis.

A third, more subtle argument is that reduced activation to reward predicting stimuli in psychosis is actually what would be expected to be seen. If the abnormality underlying delusions is pathologically increased attribution of salience to neutral stimuli, and assuming that attribution of salience to reward associated stimuli continues to occur normally, then subtracting the former from the latter, as is done in fMRI studies, would reveal reduced activation. A hint – no more than this – that something like along these lines may be going on comes from a study by Murray et al. (2008). They compared 13 mostly treated first-episode patients (11 of whom later went on to be given a diagnosis of schizophrenia) and 12 matched healthy controls on a monetary reward task where there was no interpolated reaction time task and in which the participants learnt the cue-reward association while they were being scanned. A whole-brain, voxel based comparison between the patients and controls revealed reduced activation in the patients in the midbrain, the ventral pallidum, the putamen, the hippocampus, the insula, the cingulate cortex, and the medial frontal and orbitofrontal

cortex, among other areas; there were no differences between the groups in the ventral striatum (although these were found in a subsequent ROI analysis). However, the authors also noted that the difference in midbrain activations between the two groups was driven by a combination of attenuated response to reward prediction error in the patients together with an augmented response to neutral prediction error.