Neurosciences and the Human Person: New Perspectives on Human Activities
Pontifical Academy of Sciences, *Scripta Varia* 121, Vatican City 2013
www.casinapioiv.va/content/dam/accademia/pdf/sv121/sv121-frithc.pdf

# False perceptions & false beliefs: Understanding schizophrenia

Chris D. Frith[1] & Karl J. Friston[2]

False perceptions and false beliefs are core symptoms of schizophrenia. If we want to understand these symptoms, we need to explore the interaction between the physical and the mental. Schizophrenia provides a unique opportunity to explore this interaction. To explain how the mental emerges from the physical is the key challenge facing 21st century science.

## 1. The symptoms

Hallucinations (false perceptions) and delusions (false beliefs) are characteristic symptoms of schizophrenia. Typical hallucinations include: *hearing people talking to you or about you, hearing a running commentary on your actions,* and *hearing your thoughts spoken aloud*. Typical delusions include; *believing that other people can hear your thoughts, believing that your actions are being controlled by external forces,* and *believing that people are sending you secret messages* (Mellor 1970). When reporting these symptoms, patients are trying to describe extremely unusual experiences and, as is indicated by the typical verbatim examples given below, the symptom labels listed above do not fully capture these experiences (examples from Kambeitz-Ilankovic *et al.* 2012).

> *If I breathe without other people then they get stuck to me. I get stuck to people and the thoughts come through people. There are things I've learned just before I came in. It was so bad I could hear everybody in my mind. It is like being stuck on the same wavelength as people.*
>
> *I felt myself touched in such a way as if I were hypnotised, electrified, or generally controlled by some sort of medium or some other will.*

## 2. The problem

These strange and frightening experiences lie in the mental domain, and may occur in the absence of any obvious changes in behaviour. However, these subjective experiences are intimately connected with physical events in the

[1] Wellcome Trust centre for Neuroimaging at University College London; Interacting Minds Centre, Aarhus University; All Souls College, Oxford.
[2] All Souls College, Oxford.

brain. It has been known for 50 years that the neurotransmitter dopamine has a role in the generation of the symptoms of schizophrenia. Treatment with drugs that block dopamine receptors reduce the severity of hallucinations and delusions (Johnstone *et al.* 1978). In contrast, drugs, such as amphetamine that increase the levels of dopamine in the brain, can cause symptoms very similar to those associated with schizophrenia (Connell 1958). More recently, advanced imaging methods have demonstrated increased dopamine activity in the mid–brain (striatum) of people with schizophrenia. This excessive activity is probably present in the prodromal phase of the illness before the appearance of florid psychotic symptoms (Frith 2002).

Our approach has been to develop a cognitive account of the particular symptoms associated with schizophrenia. The use of a cognitive framework, that is a computational approach based on cybernetics and information theory, is very useful since terms, such as information and representation, can be applied at the physiological as well as the psychological level of description. A successful cognitive account of particular symptoms should help us, first, to understand, a bit better, what the experience is like, second, to generalize the account to explain the whole range of hallucinations and delusions associated with schizophrenia, and third, to generalize further to how perceptions and beliefs are acquired in the normal case.

## 3. Explaining delusions of control

One of the more striking experiences reported by patients is labelled *delusion of control*. The patient feels that his actions are controlled by external forces (examples from Mellor 1970).

*My fingers pick up the pen, but I don't control them. What they do is nothing to do with me.*

*The force moved my lips. I began to speak.*

In common with a number of other symptoms (such as *hearing one's one thoughts spoken aloud*), this experience seems to spring from confusion between something that I do and something that is happening independently from me in the outside world. It has long been recognised that this distinction creates a problem for the nervous system (Helmholtz 1866). For example, when an image moves across my retina, how do I know whether this is because I am moving my eye, or because an object is moving past me? For the nervous system the difference between these two situations is that, when I move my eye, commands have been sent to the eye muscles to cause the movement to occur. Such commands have not been sent when the object moves past my eye.

Helmholtz's resolution of this problem (and many other aspects of perception) was to regard the brain as an inference machine – generating predictions about the sensory consequences of action (known as corollary discharge). Put simply, if I believe I am moving my eyes, then I will predict and confirm my re-sampling of the visual field. Conversely, if visual input changes in the absence of unintended eye movement, then the best hypothesis – that could explain this sensory evidence – is that the world is moving. Treating perception as hypothesis testing or (unconscious) inference is central to the arguments that follow and is particularly important for attribution of agency: inferring the causes of changing sensory input requires a judicious balance between the precision or confidence I assign my prior beliefs (or hypotheses), relative to sensory evidence. If moving my eyes depends upon the prior belief that visual (and proprioceptive) signals will change, then assigning too much precision to the sensory consequences of moving will provide evidence against any movement, and will subvert the intended action. It is therefore necessary to attenuate the precision of sensory signals when, and only when, they report the consequences of intended movements. This is known as sensory attenuation, whereby sensations associated with voluntary movements are suppressed and ignored. This is why the sensations produced when we tickle ourselves are so much weaker than when someone else tickles us. On the other hand, if the expected feedback is manipulated and distorted then the intensity of the effects produced by our own movements is increased (Blakemore *et al.* 1999).

Several experiments have shown that patients, especially those with delusions of control, are abnormally aware of the sensations associated with voluntary movements; in other words, there is a failure to attenuate the precision of sensations. For example, they find the experience caused by tickling themselves just as intense as that occurring when they are tickled by someone else (Blakemore *et al.* 2000, and Lindner *et al.* 2005 in relation to eye movements, see also Shergill *et al.* 2005 in relation to the sense of pressure). We believe that these observations give us clue about what it feels like to have the experience labelled *delusion of control* (Hohwy & Frith 2004). Because of the failure to attenuate sensory feedback associated with the movement, voluntary movements actually feel like involuntary movements. In other words, it doesn't feel like a movement that has been caused by my intention to move.

## 4. The experience of agency: expectations and outcomes

This account of the delusion of control assumes that an abnormal experience (*failure to attenuate sensory feedback during a voluntary movement*) is

sufficient to create an abnormal belief (*believing that external forces are causing the movements*). But several studies suggest that this is not the case. Patients with delusions do not simply have an abnormal sensory experience; they also have an abnormal experience of agency, the experience that I am the cause of this movement and its consequences. Intensive study of this experience by Patrick Haggard and colleagues has revealed the phenomenon of *intentional binding* (Haggard *et al.* 2002). When we intend to perform an action, that action and its consequences are experienced as being closer together in mental time than they are in physical time. This binding together of actions and their intended consequences has both a predictive and a postdictive component (Haggard & Chambon 2012). The binding effect is greater when the outcome of the action is more strongly expected, occurring 75% of the time rather than 50% of the time. This is the predictive component. On the other hand, the perceived time of making a movement is altered if the expected consequence of making that movement does not subsequently occur. This is a postdictive effect.

Two recent studies of patients with delusions of control have shown that the experience of agency in these patients depend largely on the outcome of the movement (postdictive effect) and very little upon expectations (predictive effect). In the first experiment (Voss *et al.* 2010) these effects were measured directly using the methods developed by Haggard. From this paradigm it appeared that patients showed no predictive component for their awareness of action and an abnormally large retrospective component. In other words their experience of action was almost entirely determined by the outcome of the action. The second experiment (Synofzik *et al.* 2010) used a very different technique in which subjects had to make pointing movements in a virtual reality setup where they were given distorted visual feedback about the position of their hand. For example, if they pointed straight ahead they would consistently see their pointing movement rotated five degrees to the left. With this paradigm it is also possible to distinguish between the role of expectations and outcomes. From their ability to detect the visual feedback rotations it was shown that the patients' motor expectations were less precise than those of the controls. At the same time the patients' pointing behaviour adapted to the false visual feedback better than the controls demonstrating a greater reliance on movement outcomes. The size of both these effects correlated with the severity of the patients' delusions of control. These results suggest that patients with delusions of control have problems combining information from two different sources: that is from prior expectations about motor movements and subsequent outcomes of motor movements. They put too much weight on outcome and too little

on expectations. This result gives us a deeper understanding of why patients are abnormally aware of the consequences of their actions.

## 5. A Bayesian approach

In the normally functioning brain information from different sources is combined in a statistically optimum manner (e.g. Ernst & Banks 2002). The mechanism for achieving this is well captured in a Bayesian framework (Kersten *et al.* 2004, Yuille & Kersten 2006). When we perform an action we predict the immediate outcome of the action on the basis of our prior knowledge. If the outcome is not what we expect (a prediction error) then we modify the knowledge on which our expectation was based and this updated knowledge determines our future expectations. All this happens at a sub-personal level. That is to say, we are not consciously aware of prior expectations, prediction errors, or updating except, perhaps, when the prediction error is large. In terms of neuronal representations, precision can be thought of as amplifying prediction errors associated with a high degree of certainty or reliability. Crucially, we also need to update predictions about the precision of prediction errors. These expectations encode our uncertainty or confidence about predictions – irrespective of their content (the expected precision is sometimes referred to as expected [un]certainty).

More generally, Bayes' theorem (Bayes 1763/1958) provides a measure of the extent to which some new evidence (e.g. the prediction error) requires that we update our beliefs about the world. Within this framework there is no qualitative distinction between perception and belief, since both involve making inferences about the state of the world on the basis of evidence (Fletcher & Frith 2009). In the case of perception, this is the evidence of our senses (Helmholtz 1878). The framework also indicates the statistically optimal procedure for combining evidence from different sources. The different sources of evidence should be weighted by their precision (the inverse of variability), with the more precise evidence being given the greater weight. Likewise, if our belief (prior knowledge) about the world is assigned a greater precision, a much greater quality of evidence will be needed before we up-date it.

Nevertheless, evidence from a very precise source, such as vision, can alter what might be expected to be well-established beliefs that are held with high precision. An example of this is the *rubber hand* illusion (Botvinick & Cohen 1998). To create this illusion the participant sits at a table with one arm out of sight under a shelf. On top of the shelf is placed a prosthetic limb roughly lined up with the real arm. The experimenter then synchronously strokes the real hand and the rubber hand. Within about one minute,

participants have the vivid experience that the rubber hand is now their own hand. Objective evidence for this experience can obtained by threatening the rubber hand which elicits a physiological responses (e.g. Ehrsson *et al.* 2004) and from asking the participant to make aiming movements which indicate that the participant is representing the position of the rubber hand as the starting point for a movement rather than that of the real hand (e.g. Chambon *et al.* 2012).

A Bayesian interpretation of this effect is as follows. Before the development of the illusion, a participant experiences highly synchronised stimuli in vision and touch which seem to come from different spatial locations, the rubber hand and the real hand respectively. To resolve this discrepancy the location of the touch stimulation is "moved" to coincide with the location of the visual stimuli on the rubber hand. In this example, the prior expectation that synchronised stimuli come from the same location and the precision of the visual sense with regard to spatial location over-ride the evidence from the somewhat less precise tactile and kinaesthetic senses.

Using this framework we can model some different ways in which false perceptions and false beliefs might arise (Corlett *et al.* 2009). For example, if too much weight was put on the evidence, i.e. the prediction errors, then people would be constantly up-dating their beliefs about the world, but never fully resolving the problem. At the other extreme, if too much weight was put on prior expectation, then people would see only what they expected to see. In extreme cases, this would lead to perception without any sensory input, resulting in hallucinations. In principle, the different models that can arise in the Bayesian framework might relate to the different forms of hallucinations and delusions associated with different disorders and different pharmacological treatments. Furthermore, the different kinds of illusion to which we are all subject will have different causes in terms of the model.

The rubber hand illusion arises because the discrepancy in the location of tactile and visual sensations is treated as a prediction error, which is eliminated by assuming that the felt real hand is at the same location as the seen rubber hand. Patients with schizophrenia acquire this illusion more rapidly and strongly than control participants (Peled *et al.* 2000), presumably because they put even more weight on the apparent prediction error. The *hollow face* illusion, in contrast, arises because too much weight is put on prior expectations. From our extensive experience with faces, we know that the nose sticks out in front. But if we look at a hollow mask of a face from the back (i.e. a concave face), then this expectation is not fulfilled since the nose is the part of the face that is furthest from us. In this case our expectations override the evidence of our sense and we see a normal convex face. Here

patients with schizophrenia are *less* susceptible to the illusion than control participants (Koethe *et al.* 2006). Thus, in both these examples, as with their perception of agency, patients with schizophrenia put more weight on sensory evidence (prediction errors) and less weight on prior expectations.

## 6. The role of dopamine in the generation of perceptions and beliefs

As we mentioned at the beginning of this essay, it has long been established that the neurotransmitter dopamine is implicated in the generation of hallucinations and delusions. But it is only recently that we are beginning to understand the precise nature of this role (Corlett *et al.* 2009, Kapur 2003, Stephan *et al.* 2009). The breakthrough came with the demonstration by Wolfram Schultz and colleagues that activity in dopamine-containing neurons could be seen as a signal of reward prediction error (Schultz & Dickinson 2000), where later work highlighted the role of dopamine in reporting the certainty or predictability of a reward; namely, the precision of reward prediction errors (Schultz *et al.* 2008).

Prior to this observation, activity in these neurons was seen as a signal of reward, since activity increased immediately after an animal received a reward, for example a drink of juice. Schultz and colleagues used Pavlovian conditioning paradigms in which animals learned that the reward would arrive one second after a visual cue. Before learning had occurred increased neural activity occurred immediately after presentation of the juice. However, after learning had occurred there was no response to the presentation of the juice, but there was a response to the presentation of the cue. These observations fit with the idea that the activity occurs when there is an unexpected signal of reward, i.e. a positive prediction error. When the reward arrives at the expected time after the cue, then there is no prediction error and no activity. In contrast, the animal does not know when the cue is going to arrive. So the cue now creates a positive prediction error. If, after learning, the reward was omitted, there was a reduction of neural activity, consistent with a negative prediction error, since the expected reward did not arrive.

Prediction errors can be used to continuously up-date representations of an ever-changing world. This process can be studied in simple probabilistic learning tasks. For example, the participant has to learn that choice A will be rewarded 80% of the time, while choice B is rewarded 20% of the time. Before learning starts, the two options will have roughly equal value. When a choice is rewarded, this creates a positive prediction error and the value of that option is increased. When a choice is not rewarded the value of that option is decreased. After some experience the participants' internal representations of the value of the options will reflect the reward

probability of these options (Sutton & Barto 1998). The rate at which the subjects learns depends upon the precision of reward prediction errors (Mathys *et al.* 2011) and should therefore depend upon manipulations of expected precision in the brain:

The rate of this kind of learning can be modified by manipulating the dopamine system. For example, Mathias Pessiglione and colleagues (2006) treated human volunteers with L-DOPA or haloperidol, drugs which respectively activate or deactivate the dopamine system, while the participants performed a simple probabilistic learning task. Activation of the dopamine system caused faster learning, while deactivation caused slower learning, although, interestingly, the effect only applied to learning about gains, not losses. This study, along with others, specifies a role for dopamine in probabilistic learning.

There is much evidence that this kind of learning, in which representations about the state of the world (beliefs) are up-dated on the basis of new evidence, is disrupted in schizophrenia. For example, many studies (e.g. Garety *et al.* 1991) have found that patients with schizophrenia "jump to conclusions", in that they base their conclusions on less evidence than controls. There is also evidence for abnormalities in the integration of new evidence into beliefs (Freeman *et al.* 2002) and for a bias against disconfirmatory evidence (Woodward *et al.* 2008).

Traditionally, the delusions, or false beliefs, associated with schizophrenia have been assumed to reflect a defect in reasoning. However, as anyone who has argued with patients about their delusions will have experienced, their logic can be impeccable. The studies listed above suggest that the reasoning problem associated with delusions may be restricted to probabilistic reasoning (see for example Howes *et al.* 2007). When logical reasoning is investigated patients show little abnormality (Kemp *et al.* 1997, Owen *et al.* 2007) or may even perform better than controls (Mellet *et al.* 2006).

At the physiological level there is also evidence for abnormalities in schizophrenia relating more specifically to prediction errors. When performing tasks that elicited *reward* prediction errors (Murray *et al.* 2007) or *causal inference* prediction errors (Corlett *et al.* 2007) schizophrenic patients were observed to show less response to such errors in the dopamine rich areas of the mid-brain.

## 7. How false prediction errors generate false beliefs

Given the evidence that schizophrenia is associated with abnormal probabilistic learning, linked to abnormal modulation of prediction errors and an over-active dopamine system, We shall now speculate on how different

kinds of failure in the prediction error system might lead to false perceptions and false beliefs. As we have seen, delusions of control, in which patients believe that their actions are caused by some external force, are associated with a failure to suppress the sensory consequences of a self-generated movement. This is an example of a falsely attenuated prediction error

What is it that can go wrong, in neural terms, with the Bayesian mechanism that could create false predictions? First, there is the possibility that, through loss of neural connections, prediction error signals are not generated (or selectively enabled by a high precision) when there actually is an error. As a result beliefs are not updated when they should be. This seems to be the case for neurological patients with *anosognosia* (Schultz *et al.* 2008). This disorder is typically associated with damage to the right parietal cortex associated with stroke. Such patients falsely believe that they can and do move their paralysed left limb. There is evidence that this disorder can be explained as follows (Fotopoulou *et al.* 2008). In the normal case, as outlined in section 3 of this essay, the intention to move creates a prediction of the consequences of the movement, both in terms of the future position of the limb and the sensory consequences of the movement. It is predictions that dominate our awareness of the action we are making. These predictions are compared with the actual outcome of the action. If there are discrepancies, prediction errors are generated which alter the representation of the action. In the case of anosognosia, the motor system concerned with the intention to move is intact and predictions are generated. But, due to the right parietal damage, there is no signal concerning the outcome of the action and no prediction error is generated. In consequence patients continue to believe that they have moved their limb.

A second possibility is that, through loss of neural connections, a precise prediction error is generated when it is inappropriate. This seems to be the case for patients with Capgras syndrome (Capgras & Reboul-Lachaux 1923). These patients falsely believe that a familiar person, typically the spouse, has been replaced by a double. The assumption here is that face recognition has a cognitive and an emotional component. Via the cognitive component we discover the identity of the face we are looking at. At the same time and independently, an emotional response is generated if the face belongs to a familiar person. Thus, when a face is identified as familiar an emotional response is expected. In the case of Capgras syndrome, probably through damage to the amygdala or its connections, there is no emotional response to a familiar face. This discrepancy between the identity signal and the emotion signal creates a prediction error. As a result the patient's belief about the identity of the face is inappropriately updated. "*This person looks*

*like my wife, but there is something not quite right about her. It can't be my wife, but someone who looks like her*" (Ellis & Young 1990).

In each of these two examples a circumscribed false belief was associated with circumscribed brain damage. In the case of schizophrenia, the false beliefs typically involve many different domains and often become more widespread with time. If these delusions also result from false predictions, then the abnormality is not likely to result from circumscribed neural disconnections. This expectation is consistent with the evidence that the dopamine system is involved. Abnormalities of this system have an impact on many brain regions.

How might predictions become false in the absence of the kinds of disconnections discussed above? Within a Bayesian framework (see section 5) precision is a very important property of a signal. More weight is given to signals with high precision. Karl Friston and his colleagues have proposed that dopamine controls the precision of prediction errors (Friston *et al.* 2012). If, as a result of excessive dopaminergic activity, prediction errors became abnormally precise, then beliefs would become updated on the basis of signals that would normally have been ignored (see Kapur 2003 for a closely related version of this idea).

A general effect whereby prediction errors became abnormally precise would have an impact on many domains. In addition, the long-term experience of false prediction errors might cause patients to put less and less weight on their prior expectations. This is because these expectations would persistently be signalled as being wrong. This formulation fits nicely with the observations discussed in section 4, showing that the experience of agency in patients with delusions of control depends upon less weight being given to expectations and more weight being given to the outcomes of motor movements.

## 8. A hierarchy of beliefs

There is an obvious problem with this account of the generation of false beliefs in schizophrenia. If they put too much weight on new evidence and too little weight on prior beliefs, then we would expect that they should be constantly changing their beliefs. This is clearly not the case. While the scope of their false beliefs may be slowly modified over time, the striking feature of delusions is that patients will stick with them despite what is perceived by everyone else as very good evidence against their belief. We suggest this problem can be resolved if we recognise that perceptions and beliefs do not exist in isolation, but are developed within a hierarchy. It is the beliefs at the top of this hierarchy that are particularly resistant to change.

In Karl Friston's account of these Helmholtzian ideas (Friston 2005), the brain uses a hierarchy of predictions, where expectations at any level provide prior beliefs for the level below (these are known as empirical priors in statistics). Each level integrates new evidence from the level below and (empirical) prior expectations from the level above to generate a prediction error. This prediction error is fed upwards as the *evidence* for the next level of the hierarchy. Likewise, the prior expectations at the higher levels of the hierarchy (empirical priors) are fed downwards to constrain the possible explanations of the prediction errors coming from the lower levels. Crucially, the weights assigned to bottom-up prediction errors and top-down predictions depend upon the relative precisions (possibly encoded by dopamine) at each level of the hierarchy.

The lowest level of this hierarchy of perceptions and beliefs is the most closely linked to raw sensation, while the higher levels are concerned with more abstract levels of representation. The process of reading provides a useful illustration of the workings of such a hierarchy. At a low level we have the graphic shape components of which the letters are composed, and then we can move up the hierarchy through representations of words and sentences, reaching meaning at the highest level. However, reading is not a linear process, moving steadily upwards from shapes to meanings. The high level of meaning will constrain how signals are interpreted at the low level of shape. Consider, for example, the string of shapes event. The ev in this string is ambiguous and could be seen as "w" or as "ev". How it is seen will depend on the meaning of the sentence in which it occurs: "w" in "Jack and Jill *went* up the hill", "ev" in "the pole vault was the last *event*". The meaning of the sentence has had a top-down effect on our perception at a much lower level of the hierarchy. Presented with these two sentences we will read them easily without ever noticing the ambiguity of the shapes used to write them.

If the prediction errors being generated at the bottom of this hierarchy are treated as being unduly precise, then their effects will gradually work upwards through the hierarchy, and they will never be fully eliminated by changing low-level beliefs about the world. Consider what might happen if something goes wrong with the fancy system in my car that signals problems. In particular, assume that an error warning light is unduly sensitive to fluctuations in the engine's performance from normal levels. This would correspond to a pathologically highly precision at the sensory level, leading to a dashboard warning light that is almost continuously illuminated. I am led to falsely believe that there is indeed something wrong with the engine. I take my car to the garage and they report that nothing is wrong. However,

the light is still on and keeps on signalling an error. So, this leads me to falsely believe that the garage is incompetent. I report them to the "good garage guide" who investigate and conclude that the garage is not incompetent. Now I believe that the "good garage guide" is corrupt.

We suggest that, in the case of schizophrenia, it is the beliefs at the top of the hierarchy that are so resistant to change. This is because, for the patient, they seem to be the only way of explaining away the apparent problems with lower levels of the hierarchy. In the normal case prediction errors at the lowest level of the hierarchy elicit changes in our interpretation of sensory input. This enables us to develop an increasingly accurate account of the causes of our sensations. In other words, we develop a representation of the world that corresponds ever more closely to reality. Falsely precise prediction errors undermine this process and lead us ever further from reality. They are transmitted up our hierarchy of beliefs as we attempt to explain them away (Fletcher & Frith 2009).

This process has been described in a particular striking manner by Peter Chadwick (1993). Chadwick, who has a PhD in psychology, has described in detail his experiences during an episode of paranoid schizophrenia. In my opinion, his descriptions lend themselves well to the account of delusions we have developed in this essay. He says, "*I had to make sense, any sense, out of all these uncanny coincidences. I did it by radically changing my conception of reality*". In our terminology, these uncanny coincidences were false hypotheses engendered by prediction errors with inappropriately high precision or salience. To explain them away Chadwick had to conclude that other people, including radio and television presenters, could see into his mind. This was the radical change he had to make in his conception of reality.

## 9. Conclusions

We suggest that the Bayesian framework, outlined here, for explaining perceptions and beliefs provides a plausible account of the development of hallucinations and delusions in schizophrenia. In addition, the account can be directly linked to physical processes involving the dopamine system of the brain. In principle, such an account can provide a guide for the development of new treatments, whether these are at the cognitive or the biological level. For example, it might be possible to develop a method for reducing the precision of prediction errors.

Explorations of abnormal behaviour and experience will always illuminate our general understanding of the mind. This account of the generation of false beliefs in the case of schizophrenia, makes me realise how fragile this process is and how easily it might go astray in the normal case. Given

the right anomalous sensory experiences each of us could develop some bizarre and erroneous belief system. Why does this not happen more often? We believe that we are usually saved from taking such erroneous paths by the constraints provided by those even higher levels of the belief hierarchy that are external to our brains. These constraints arise from our interactions with our peers and with our culture. Even at the lowest perceptual level of the hierarchy the high level constraints that arise from interactions with others enable us to achieve accounts of the world that are more accurate than those that we can develop on our own (Bahrami *et al.* 2010). It is this submission of our own ideas to the criticism of others that has been formalised in the practice of science.

In the case of schizophrenia, in contrast, these high level external constraints no longer seem to operate. Patients stick to their false beliefs in spite of the objections of others. Is this an inevitable consequence of the process by which prediction errors filter up through the hierarchy, or is this evidence for some additional problem that needs to be identified? Further research is needed.

## Acknowledgements

Our work is supported by the Wellcome Trust, the Danish National Research Foundation and Aarhus University. We are grateful to Uta Frith and Rosalind Ridley for their comments.

## References

Bahrami B, Olsen K, Latham PE, Roepstorff A, Rees G, Frith CD. 2010. Optimally interacting minds. *Science* 329: 1081-5.

Bayes T. 1763/1958. Studies in the History of Probability and Statistics: IX. Thomas Bayes' Essay Towards Solving a Problem in the Doctrine of Chances. *Biometrika* 45: 296-315.

Blakemore SJ, Frith CD, Wolpert DM. 1999. Spatio-temporal prediction modulates the perception of self- produced stimuli. *J. Cogn. Neurosci.* 11: 551-59.

Blakemore SJ, Smith J, Steel R, Johnstone EC, Frith CD. 2000. The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychol. Med.* 30: 1131-39.

Botvinick M, Cohen J. 1998. Rubber hands 'feel' touch that eyes see. *Nature* 391: 756.

Capgras J, Reboul-Lachaux J. 1923. L'illusion de 'sosies' dans un délire systématise chronique. *Bull. Soc. Clin. Med. Ment.* 2: 6-16

Chadwick PK. 1993. The step ladder to the impossible: A first hand phenomenological account of a schizo-affective psychotic crisis. *Journal of Mental Health* 2: 239-50.

Chambon V, Wenke D, Fleming SM, Prinz W, Haggard P. 2012. An Online Neural Substrate for Sense of Agency. *Cereb Cortex*.

Connell P. 1958. *Amphetamine Psychosis*. London: Chapman & Hall.

Corlett PR, Frith CD, Fletcher PC. 2009. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology (Berl)* 206: 515-30.

Corlett PR, Murray GK, Honey GD, Aitken MR, Shanks DR, *et al*. 2007. Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* 130: 2387-400.

Ehrsson HH, Spence C, Passingham RE. 2004. That's My Hand! Activity in Premotor Cortex Reflects Feeling of Ownership of a Limb. *Science*: 1097011.

Ellis HD, Young AW. 1990. Accounting for delusional misidentifications. *British Journal of Psychiatry* 157: 239-48.

Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429-33.

Fletcher PC, Frith CD. 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci* 10: 48-58.

Fotopoulou A, Tsakiris M, Haggard P, Vagopoulou A, Rudd A, Kopelman M. 2008. The role of motor intention in motor awareness: an experimental study on anosognosia for hemiplegia. *Brain* 131: 3432-42.

Freeman D, Garety PA, Kuipers E, Fowler D, Bebbington PE. 2002. A cognitive model of persecutory delusions. *Br J Clin Psychol* 41: 331-47.

Friston K. 2005. A theory of cortical responses. *Philos T Roy Soc B* 360: 815-36.

Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, *et al*. 2012. Dopamine, affordance and active inference. *PLoS computational biology* 8: e1002327.

Frith C. 2002. Attention to action and awareness of other minds. *Conscious Cogn* 11: 481-7.

Garety PA, Hemsley DR, Wessely S. 1991. Reasoning in deluded schizophrenic and paranoid patients. Biases in performance on a probabilistic inference task. *J. Nerv. Ment. Dis.* 179: 194-201.

Haggard P, Chambon V. 2012. Sense of agency. *Curr Biol* 22: R.390-2

Haggard P, Clark S, Kalogeras J. 2002. Voluntary action and conscious awareness. *Nat. Neurosci.* 5: 382-85.

Helmholtz Hv. 1866. *Handbuch der Physiologischen Optik*. Leipzig: Voss.

Helmholtz Hv. 1878. The Facts of Perception In *Selected Writings of Hermann Helmholtz*: Wesleyan University Press.

Hohwy J, Frith C. 2004. Can neuroscience explain consciousness? *J. Conscious. Stud.* 11: 180-98.

Howes OD, Montgomery AJ, Asselin MC, Murray RM, Grasby PM, McGuire PK. 2007. Molecular imaging studies of the striatal dopaminergic system in psychosis and predictions for the prodromal phase of psychosis. *Br J Psychiatry Suppl* 51: s13-8.

Johnstone EC, Crow TJ, Frith CD, Carney MW, Price JS. 1978. Mechanism of the antipsychotic effect in the treatment of acute schizophrenia. *Lancet* 1: 848-51.

Kambeitz-Ilankovic L, Hennig-Fast K, Benetti S, Kambeitz J, Pettersson-Yeo W, *et al*. 2012. Attentional Modulation of Source Attribution in First-Episode Psychosis: A Functional Magnetic Resonance Imaging Study. *Schizophr Bull*.

Kapur S. 2003. Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry* 160: 13-23.

Kemp R, Chua S, McKenna P, David A. 1997. Reasoning and delusions. *Br J Psychiatry* 170: 398-405.

Kersten D, Mamassian P, Yuille A. 2004. Object perception as Bayesian inference. *Annu Rev Psychol* 55: 271-304.

Koethe D, Gerth CW, Neatby MA, Haensel A, Thies M, *et al*. 2006. Disturbances of visual information processing in early states of psychosis and experimental delta-

9-tetrahydrocannabinol altered states of consciousness. *Schizophr Res* 88: 142-50.

Lindner A, Thier P, Kircher TT, Haarmeier T, Leube DT. 2005. Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Curr Biol* 15: 1119-24.

Mathys C, Daunizeau J, Friston KJ, Stephan KE. 2011. A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 5: 39

Mellet E, Houde O, Brazo P, Mazoyer B, Tzourio-Mazoyer N, Dollfus S. 2006. When a schizophrenic deficit becomes a reasoning advantage. *Schizophr Res* 84: 359-64.

Mellor CS. 1970. First rank symptoms of schizophrenia. *British Journal of Psychiatry* 117: 15-23.

Murray GK, Corlett PR, Clark L, Pessiglione M, Blackwell AD, *et al*. 2007. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Molecular Psychiatry* 13(3): 267-76.

Owen GS, Cutting J, David AS. 2007. Are people with schizophrenia more logical than healthy volunteers? *Br J Psychiatry* 191: 453-4.

Peled A, Ritsner M, Hirschmann S, Geva AB, Modai I. 2000. Touch feel illusion in schizophrenic patients. *Biological Psychiatry* 48: 1105-8.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442: 1042-5.

Schultz W, Dickinson A. 2000. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23: 473-500.

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo CD, *et al*. 2008. Explicit neural signals reflecting reward uncertainty. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 363: 3801-11.

Shergill SS, Samson G, Bays PM, Frith CD, Wolpert DM. 2005. Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry* 162: 2384-6.

Stephan KE, Friston KJ, Frith CD. 2009. Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophr Bull* 35: 509-27.

Sutton RS, Barto AG. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Synofzik M, Thier P, Leube DT, Schlotterbeck P, Lindner A. 2010. Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain* 133: 262-71.

Voss M, Moore J, Hauser M, Gallinat J, Heinz A, Haggard P. 2010. Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain* 133: 3104-12.

Woodward TS, Moritz S, Menon M, Klinge R. 2008. Belief inflexibility in schizophrenia. *Cognit Neuropsychiatry* 13: 267-77.

Yuille A, Kersten D. 2006. Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci* 10: 301-8.