Published in final edited form as:
Annu Rev Clin Psychol. 2017 May 08; 13: 265–289. doi:10.1146/annurev-clinpsy-032816-045145.

Predictive Processing, Source Monitoring, and Psychosis

Juliet D. Griffin and **Paul C. Fletcher**
Dept. of Psychiatry, University of Cambridge

Abstract

A comprehensive and useful understanding of psychosis will require models that link multiple levels of explanation: the neurobiological, the cognitive, the subjective, the social. Until we can bridge several explanatory gaps, it is difficult to envisage understanding how neurobiological perturbations could manifest in bizarre beliefs or hallucinations, nor how trauma or social adversity could perturb lower-level brain processes.

We propose that the predictive processing framework has much to offer in this respect. We show how it may underpin and complement source monitoring theories of delusions and hallucinations and how it easily extends, when considered in terms of a dynamic and hierarchical system, to provide a compelling model of several key clinical features of psychosis. Crucially, we see little conflict between source monitoring theories and predictive coding. The former acts as a higher level description of a set of capacities, and the latter aims to provide a deeper account of how these and other capacities may emerge.

Keywords

source monitoring; predictive coding; schizophrenia; psychosis

Predictive processing, prediction error and the emergence of psychosis

In recent years, the growing idea of the brain as an organ of predictive inference (Clark 2013, Friston 2010) has been central to establishing Computational Psychiatry as a framework for understanding how alterations in brain processes could drive the emergence of high-level psychiatric symptoms (Huys et al 2016, Adams et al 2015, Friston et al 2014, Corlett & Fletcher 2014). Collectively known as “predictive coding” or “predictive processing”, such theoretical approaches have been applied to perception, belief, attention and action and have even been put forward (notably in the context of the “Free Energy Principle” (Friston 2005, Friston et al 2006, Friston & Stephan 2007, Friston 2009, Friston 2010) as the basis for a unified theory of brain function. In practice, they comprise a loosely grouped set of models and ideas embodying various assumptions, implicating differing brain processes, and expressing themselves at different levels of explanation – the computational, the algorithmic, the mechanistic (see Teufel & Fletcher, *Brain*, in press, for discussion).

Please address all correspondence to Prof. Paul Fletcher (pcf22@cam.ac.uk), Department of Psychiatry, Herchel Smith Buidling for Brain & Mind Sciences, Forvie Site, Robinson Way, Cambridge CB2 0SZ, UK.

Europe PMC Funders Author Manuscripts

In the current paper, we invoke predictive processing within a fairly narrow range, using it as a cognitive framework within which to consider, in simple terms, the updating and inferencing processes that optimise the brain's capacity to model the statistical regularities of its environment. On occasion, we attempt, tentatively, to relate these processes to underlying brain systems and structures (for example, prediction error signal in the mesocorticolimbic dopamine system). For the most part, however, we are aiming to express the ideas primarily at the cognitive level, but in a way that capitalises on one of computational psychiatry's key advantages: the opportunity that it affords to unite levels of description. Moreover, we propose that ideas of the brain as a predictive device making inferences about its environment allow us to make links not only from cognitive processes to brain processes, but also to higher level, subjective and social phenomena (Hohwy 2013, Clark 2013). In doing so, we are presented with the exciting opportunity of developing ways of discussing some of the key symptoms of schizophrenia in terms that have currency at the cognitive and systems neuroscience levels. In particular, we argue that the aberrant and apparently irrational perceptions and beliefs that characterise psychosis may fruitfully be considered within this framework. We do not see this perspective as necessarily replacing, or contradictory of, other influential, higher-level cognitive accounts such as those invoking source monitoring deficits, but rather as a means of complementing these descriptions and a route towards relating them to underlying neurobiology. Indeed, a large part of our discussion attempts to show how complementary the predictive processing and source monitoring models may be.

Our overall aim is to show how the predictive processing (PP) framework offers new conceptualisations of psychotic illness, and how such conceptualisations have remarkable power to provide credible explanations for the emergence, persistence and co-occurrence of key symptoms and difficulties.

A predictive processing model of psychosis: early ideas

An initial version of these ideas lay in the simple perspective that the early emergence of delusions and hallucinations could be conceptualised in terms of associative learning models and, more particularly, in terms of a disruption to prediction error-dependent updating (Corlett et al 2006, Corlett et al 2007, Corlett et al 2009, Fletcher & Frith 2009). This perspective draws on an idea central to PP models: that the brain is foremost attempting to model the world (Conant & Ashby 1970) and that it does so by updating inferences through iteratively reducing 'prediction errors': signals connoting a mismatch between the current model's predicted inputs and the actual inputs it receives from the world. The link between prediction error signalling and dopaminergic function (Schultz et al 1997) offered indirect but enticing support for supposing that a dopamine-related alteration in prediction error may relate to psychotic illness. It should be noted that this view may be considered a modification of earlier views about how dopaminergic dysfunction may shape the emergence of delusions (Kapur 2003) and longer-standing hypotheses relating the neurobiology of associative learning to schizophrenia (Miller 1976, Helmsley 2005a, Helmsley 2005b, Gray et al 1991). The idea was framed simply: in the context of aberrant prediction error signalling, the capacity to make accurate inferences about the world would be severely compromised: the world model would be inaccurate and sub-optimal and, ultimately,

abnormal perceptions and beliefs would arise. Ensuing studies, which will not be reviewed here (see Heinz and Schlagenhauf 2010 and Corlett and Fletcher 2015 for overviews), have provided support for the model but it is important to acknowledge that its earlier, rudimentary form left a series of gaps. First, aside from a general idea of "aberrant" prediction error, early formulations were largely silent on the precise nature of any deficit (whether the prediction error signal was abnormally large, strong, inconsistent or noisy was not fully addressed (though see Fletcher & Frith 2009 who speculated that perhaps the aberration lay in the abnormal "gain"). Second, there were a number of features of psychosis that were unaccounted for (notably, the strong (largely negative) emotional component, the striking resilience of the beliefs and their imperviousness to seemingly contradictory evidence, the almost ubiquitously social nature of the symptoms, and the accompanying cognitive deficits and negative symptoms that may occur). Third, there was little formal attempt to integrate the prediction error model with other valuable, and empirically supported, perspectives expressed at different levels (notably, the source monitoring explanation (Keefe et al 1999, Woodward & Menon 2012, Arguedas et al 2012, Cannon 2015)).

Here, we aim to at least partially fill these gaps. We speculate on the specific ways in which prediction error signalling, within a PP framework, might go wrong, arguing that a hierarchical and dynamic perspective has much to offer. By this we mean that the integration of predictions with evidence -an informational conversation that is central to PP ideas - occurs at multiple levels with varying timescales and degrees of abstraction (i.e. is hierarchical) and that it reflects an ongoing and evolving attempt to optimise one's world-model in response to new evidence (i.e. is dynamic). We consider how PP both complements and extends ideas on source monitoring and how, indeed, source monitoring and the attribution of agency (itself an important consideration in psychosis) fall naturally within the ambit of PP and prediction error minimisation. Finally, we apply these perspectives to some of the key characteristics of psychosis beyond those considered within the original models.

Extending the model: newer perspectives on predictive processing, prediction error and psychosis

We need to consider more precisely how prediction error (PE) signalling is altered in psychosis, and how such a perturbation may evolve over time. This evolution, we suggest, takes an interesting form and can encompass the gradual change from elasticity to fixity of beliefs. Moreover, by considering how the brain's model - its explanations about and predictions of the world -takes a hierarchical form, it becomes possible to enrich the explanatory framework appreciably.

The successful brain must model its world (Conant & Ashby 1970). It needs to be able to generate accurate predictions about upcoming events, so that it can marshal preparatory responses. To do this, it must make associations. Theoretical models of associative learning invoke simple error minimisation to achieve this: the model is updated as a function of its failure to predict accurately (Rescorla & Wagner 1972). However, in a world that is often probabilistic, it is fruitless to seek unerringly accurate predictions, meaning that there are

many conditions under which a model should not be updated even when there is appreciable error in a given instance. But in tolerating inaccuracy, sensitivity to change may be sacrificed: a potential problem in making efficient adjustments to environmental volatility. Added to this, the associative regularities of the world, in addition to being noisy and volatile, may be highly context-dependent. There are second, third and even higher order associations in our world and these pose a real challenge to simple error-dependent updating. Imagine finding a bush that yields delicious berries. The 'bush-berries' association that must be formed in order to exploit this contingency in the future could be set up very easily through reinforcement learning, based on occasions on which the bushes were paired with rewarding berries. But should one update this association upon discovering one day that there are no berries on one of these bushes? Perhaps a blight has struck the whole species and a new foraging strategy will be required. Perhaps this particular bush has been exhausted but the general association should remain strong. Perhaps the berry season has ended and one must learn a context dependence between berry yield and time of year. When to maintain expectations, when to update them, when to generalise them, and when to subordinate them to higher-order context-dependent expectations are complex problems. They can only be solved by a system capable of forming associations which are flexible enough to be modified when they no longer pertain, yet robust enough not to extinguish in the face of occasional PEs. And the system must ideally be capable of identifying and flexibly incorporating representations of higher order contingencies, such as those produced by the changing seasons.

Put simply, inferences must be flexible in the face of change but robust in the face of noise. As an aside, it is noteworthy that dopamine, in addition to signalling the magnitude of reward PEs (Schultz et al 1997) has been suggested to be critical in signalling the precision (Fiorillo et al 2003), or epistemic value (Schwartenbeck et al 2016) of incoming information, setting the gain on cortical PE signals. This could provide a powerful means of attuning the brain's model as sensitively as possible to the statistical regularities of its environment, while minimising the potential for premature, unnecessary and misleading updating. More recent formulations of psychosis (Adams et al 2013) have shown the potential power of this perspective in understanding delusions and hallucinations. Considering this fundamental problem more closely, Preuschoff and Bossaerts (2007) consider, in computational terms, how the persistent experience of a noisy, unreliable (and hence informationally-devalued) PE signal should cause an information-processing system to adapt by down-regulating the updating that occurs in response to a given magnitude of PE signal. This capacity to adapt learning rate to the noisiness of the PE (specifically, to the standard deviation of a reward value) has been demonstrated (Nassar et al 2010) and related to the ability to learn optimally (Diederen and Schultz 2015), with ensuing fMRI work corroborating the idea that the adaptation occurs within the dopaminergic system (Diederen et al 2016). Moreover, there is recent evidence that such context-dependent adaptive coding of rewards is disrupted in schizophrenia (Kirschner et al 2016). One can envisage how a small but persistent disruption in the ability to optimally weight each PE could create sub-optimal adaptation, leading to evolving consequences such as initially high levels of PE-driven updating being followed by later downregulation of learning.

Clearly, in speculating on PEs in psychosis, we must consider not just how they drive inferences, but how inferences about errors themselves can modulate this drive over time. Indeed, recent theoretical work suggests that psychosis could be characterised by fundamental disturbances in adapting learning responses to both PE variability and environmental volatility (Adams et al 2013). Related work using ketamine as a pharmacological model of early-stage ('prodromal') psychosis shows its infusion to be associated with reduced tendency to downregulate PE-driven learning in order to capitalise on existing statistical regularities within the environment (Vinckier et al 2015)).

In addition, evidence from patient studies is suggestive of an impairment in the flexible modulation of PE-driven learning: an impairment whose nature and direction may change systematically over time. People with chronic schizophrenia show impairments in the flexible generalisation of (correctly learned and accurately remembered) associations to novel contexts (Shohamy et al 2010), and they classically have a perseverative, inflexible response pattern (consistent with a tendency to underutilise prediction errors) on probabilistic reversal learning (PRL) tasks (Elliott et al 1995, Kerns & Berenbaum 2002, Pantellis 1997 – although see also Waltz et al 2013). In contrast, Schlagenhauf et al (2010) demonstrated an increased tendency to update in response to prediction error in unmedicated people in their very first episode of psychosis. Consistent with this, Culbreth et al (2016) demonstrated a similar tendency in schizophrenia, which they attributed to 'unstable' representations of value (but which they might equally have been attributed to 'noisy' or 'imprecise' prior representations of value) and which was mediated entirely by impaired activation in cingulate, parietal and frontal regions. This 'cognitive control network' has been implicated in predicting and preventing errors (Brown & Braver 2005, Botvinick 2007), in representing uncertainty, volatility and the value of information (Rushworth & Behrens 2007, Behrens et al 2008) and in implementing predictive, context- and goal-dependent control over processing and behaviour, to accord with anticipated task requirements (Cole & Schneider 2007). The network's role in implementing such 'proactive control' includes flexibly modulating low-level reinforcement learning parameters via its connections with the striatum (Ceaser & Barch 2015), and it overlaps significantly with the ketamine-susceptible network identified in Vinckier and colleagues' pharmacological study (2015) as involved in updating a representation of meta-level confidence in one's prior belief which modulates lower-level updating of value representations during reinforcement learning. This proactive control system's ability to facilitate flexible, rapid updating in response to relevant information, while preventing interference from irrelevant information, has been suggested to depend on temporally precise dopaminergic neuromodulation of this network (Braver 2012). It may be, therefore, that in the early stages of the disorder, dopaminergic dysregulation causes uncontrolled interference from unreliable prediction errors, leading directly to the 'aberrant salience' experiences characteristic of the prodrome (Ceaser & Barch 2015, Kapur 2003), while in later stages the dysregulation's effect on the proactive control network is to impede the flexible updating of beliefs, goals and behavioural policies in response to useful prediction errors.

This is all speculative, but the key point is that a persistent underlying disturbance could, as a result of ongoing adaptive processes, manifest in different ways across time. From a clinical perspective, persistence in the face of contradictory evidence is a defining

characteristic of delusions, but the character of this persistence is rather different depending on the stage of clinical presentation. People in the early, prodromal stages creatively embellish their delusions to flexibly explain away counterevidence, whereas ‘delusional persistence’ as it applies to first episode and especially chronic schizophrenia is often characterised by inflexibility. Over time, delusions become epistemically insulated and inert: people who once would have strenuously challenged counterevidence may ignore it altogether, and in some cases even fail to acknowledge it as relevant to the content of their delusion (Roessler 2013). We concur with a recent review stressing the need for the field in general to adopt this kind of temporally extended perspective (Cannon 2015) and suggest, given the perspective outlined above, that PP, by virtue of being instantiated in computational terms, provides a remarkably rich framework within which to consider the systematic, time-dependent evolution of symptoms and experiences within an information-processing system.

How does Predictive Processing relate to Source Monitoring?

We suggest that PP-based explanations complement rather than compete with source monitoring accounts of psychosis, by reframing such accounts in terms of the mechanisms that underlie how one makes a decision about the source of an input, including decisions about whether the input was caused by an external source or is the consequence of one’s own actions. Specifically, the organism is repeatedly faced with the challenge of integrating an array of (sometimes competing) cues to produce the best explanation for an input (Lindsay and Johnson 2000). Only through such integration is it possible to make key source decisions (e.g. did I cause this or was it some external agent?). Clearly, it is critical to be able to do this, and even a subtle impairment in source monitoring ability could profoundly change one’s world model. Given this importance, source monitoring accounts may provide powerful explanations for an array of psychotic symptoms, particularly given that source monitoring must extend to representations brought to mind by recall as well as by both interoceptive and exteroceptive signals. Thus, while the judgement that a rustling sound in the forest was more likely to have been caused by the wind than by a bear is an example of source monitoring, so too is the unbidden recollection of who it was that reassured you that there are no bears in this particular forest. As we shall discuss, this source information may be critical especially insofar as it could provide a marker for the reliability of this information.

A specific case of this form of processing is ‘reality monitoring’, in which the subject must discriminate internally-generated from externally-derived stimuli or information. According to the source monitoring account, delusions and hallucinations result from incorrect attribution of imagined or self-generated representations to external causal sources. Thus, auditory hallucinations may result from inner or subvocal speech being misidentified as externally caused (Moseley et al 2013, Frith & Done 1989), and delusions of threat may result from intrusive memories of past trauma being misidentified as (for example) psychic premonitions rather than memories (Holmes & Steel 2004, Steel et al 2005).

The source monitoring account of psychosis has been especially compelling in relation to ‘passivity phenomena’, such as delusions of control and ‘thought insertion’ (the experience

that another agent's thoughts are being inserted into one's own subjective consciousness) (Nelson et al 2014a). Moreover, since it has been suggested that a failure to correctly identify the origin of a movement, thought or experience may capture the essential problem that lies at the heart of other features of schizophrenia beyond those formally described as passivity phenomena, the source monitoring account could offer a truly fundamental description of the condition (Nelson et al 2014b). Indeed, source monitoring problems are relatively specific to schizophrenia spectrum disorders (Nordgaard and Parnas 2014, Parnas et al 2003, Haug et al 2012a), predict transition in nonpsychotic clinical patients and high-risk groups (Nelson et al 2012, Parnas et al 2011), and are associated with the most debilitating functional impairments: suicidality, lack of insight, and social dysfunction (Haug et al 2012b, Parnas et al 2013, Haug et al 2013).

The functional correlates of source monitoring, and of its deficits in schizophrenia, are well characterised (Johnson et al 2009) and the account complements research into other central cognitive features of schizophrenia, notably the deficits in episodic memory (Leavitt & Goldberg 2009) and in metacognition (Lysaker et al 2015). Source monitoring and episodic memory are not fundamentally different psychological constructs (Johnson 2005, Johnson 2006), and reality monitoring ('was this real, or did I imagine it?') relies heavily on discriminating which cognitive operations were involved in causing the representation in question (Johnson et al 1979, Johnson et al 1981) – itself a form of metacognition.

In short, source monitoring failure provides a compelling descriptive account of what is intuitively a core problem in psychosis: a failure to discriminate internal from external, reality from imagination. The source monitoring framework has great appeal as a clear and compelling description of the nature of psychotic symptoms. This in itself is important in engaging with those who suffer from these experiences, since explanations must work for those whose questions they are intended to answer (Wilkinson 2014). Kapur (2003) rightly points out the inadequacy of offering a psychiatric patient, by way of 'explanation' for his delusional symptoms: 'you believe your neighbours are trying to kill you because of dysregulated dopamine'. In this respect, the source monitoring account does seem to offer more in the way of direct explanation.

However, one shortcoming of the source monitoring account, with respect to its potential for clinical translation, is that it is not well specified at a computational level (Huys et al 2016). Indeed, it was never intended to be a process theory: Marcia Johnson has been explicit about the need to investigate the specific details of the cognitive processes that *produce* a source monitoring decision as their output (Johnson & Hirst 1993, Johnson et al 1993, Mitchell & Johnson 2009). Source monitoring is a high level description of a cognitive capability, one that encompasses other processes, rather than being a specific cognitive process itself. There is a need for a formal computational account that could bridge the explanatory gaps between brain, cognition and symptoms (Frith 2015). The PP framework points towards this bridge. It does not replace source monitoring ideas but rather offers them a deeper, computational foundation. We argue that source monitoring deficits are an emergent feature of PP imbalance. This assertion follows from the idea that we can think of source monitoring attributions as inferences about what (or who) was the most likely cause of an event, made

under conditions of noise, ambiguity and uncertainty. These are precisely the sorts of inferences with which PP is concerned.

We begin with a PP account of one important example of source monitoring - sense of agency - and of how it may be disturbed in psychosis (Moore & Fletcher 2012). Imagine I am walking in a forest and hear a rustling noise. This mere sensory experience does not uniquely identify any single underlying cause: it might have been caused by a bear, a squirrel, a falling pine cone or simply my own movements through the foliage. There is inferential work to be done to determine whether I caused the sound or something else did (i.e. whether its source was internal or external), and further work to determine whether, if external, it was caused by agent or accident - all of which may, of course, have important implications for ensuing decisions and actions. Since the sensory evidence itself is ambiguous, the inference must take into account which of these hypotheses is a priori most likely. That is, in the face of evidential ambiguity, I must make a best guess. Here, valuable information comes from how well-predicted the noise was. A noise, for example, that accords precisely with the timing of one's foot hitting the grass can be confidently identified as self-caused. But if the noise cannot be predicted on the basis of the antecedent state of the system, the ensuing prediction error will be potentially informative that there is an external source. In short, sensory events for which I am the source are predictable in advance (Frith et al 2000, Blakemore et al 2002, Wolpert & Flanagan 2001) while sensory events arising from external sources are not. Prediction error therefore offers itself as a crucial signal for 'internal source monitoring' (inferring whether the cause of an event was me or something else).

But this is rather simplistic and it is important to note that not all self-generated experiences will be fully predicted. Moreover, cues – visual, auditory, sensory, proprioceptive – will not always be compatible with each other: a visual cue may locate the origin of a sound to one place, while an auditory cue may conflict and locate it elsewhere (Knill & Pouget 2004). The question arises as to how cues must be combined optimally to make the best inference. A useful perspective here – based on a simple Bayesian formulation – is that cues are combined and weighted according to their reliability or precision (Moore & Fletcher 2012, Knill & Pouget 2004). In bright daylight, a visual cue may offer the most precise information about the source of a noise, while in darkness it may be superseded by the estimated location of the noise, based solely on auditory cues such as the inter-aural time and intensity differences (Moore 2004). In light shoes, the sensory input from one's own footfalls will have a higher precision than when wearing heavy boots or when numb from the cold. The ultimate inference (self or external? Agent or accident?) may also be shaped by higher level prior expectations. Perhaps the strong conviction that one is not alone, or the ready accessibililty of an unpleasant memory from a horror film, will shape or determine how the evidence is composed into an inference. Pezzulo provides an elegant and comprehensive discussion of these ideas (Pezzulo 2014).

Ultimately, we see source monitoring as a conclusion or inference to best account for the overall data. That is, it is an abductive inference: an inference from the data back to the most likely cause (see Coltheart 2010). Given that the need to make abductive inferences is at the

heart of PP frameworks, it becomes clear that these frameworks must concern themselves directly with source monitoring and its disruptions.

PP accounts of passivity phenomena in psychosis have been mainly concerned with what the source monitoring literature terms ‘internal source monitoring’ errors – that is, they have endeavoured to understand what underlying computational factors could produce the erroneous yet compelling experience that one’s own internally-willed movement was caused by an external source (Blakemore et al 2002, Frith et al 2000, Frith 2005, Shergill et al 2005, Shergill et al 2014). While these investigations into internal source monitoring and its errors have been remarkably successful, the PP account can go further. It can be naturally extended to encompass source attributions more broadly, and in so doing, it has the potential to offer a computational explanation for a hitherto poorly understood characteristic of passivitiy experiences in schizophrenia: they involve not just the failure to recognise oneself as the agent of an action, but also its attribution to an external agent. Recalling the toy example of a noise in the forest above, suppose that I infer that the noise was caused by an external source. There remains the key question: is the source agentic, or not. Patients with psychosis do not merely attribute their movements to some accidental external force – rather, the movement is inferred to have been caused by an external agent (Pacherie et al 2006).

Events caused by agentic sources usually have features that distinguish them from those caused by external physical forces: they start and stop, they speed up and slow down, they take non-linear trajectories. They are, in short, relatively unpredictable without recourse to some higher-level model possibly involving goals and intentions. This is in contrast to the regularity and predictability of non-agentic movements (Biro et al 2007, Premack 1999). Once again the prediction error signal, within a model that encompasses these higher order expectations, is very useful here. We suggest that, in psychosis, the fine-grained predictive model of the moment-to-moment changes in sensory input that are expected on the basis of one’s own planned movement is relatively imprecise (Synofzik et al 2010, Voss et al 2010). Thus, the sensory consequences of one’s own actions are associated with an unusually high PE at this level, suggesting that this was not one’s own agentic movement. However, those same sensory inputs nevertheless retain the perceptual qualities that are typical of self-propelled, goal-directed action and therefore they accord with our model of the general characteristics of agentic action (Cicchino et al 2011). The abductive inference to the best explanation – that is, the one that minimises prediction error at all levels of the hierarchy – therefore entails not merely that the movement was caused by something other than myself, but also that it was caused by an agent other than myself. Indeed, one might tentatively suggest that the more unpredictable an outcome, up to a point, the more likely it is to reflect a cause that is both external and agentic. We might speculate further that simple unpredictability may act as a marker for an externally-generated event while more sustained unpredictability might indicate a source that defies description in terms of simple physical laws and might therefore be agentic. We are therefore offered a parsimonious account for both the misattribution (external rather than internal) and for the accompanying attribution of agency to the supposed external source.

In short, PP offers a framework for understanding source monitoring capabilities as well as a parsimonious explanation for how specific perturbations might generate characteristic

oddities of both source and agency attribution. Given that the self-same perturbations can give an account of several other characteristics of psychosis, we find this attractive and useful in generating a deeper level of understanding. But, more than that, it provides opportunities to develop a richer perspective on how source monitoring problems interact causally with these other features of the illness, both driving them and, in turn, being influenced by them. The essence of this admittedly speculative argument is that not only would problems in uncertainty-weighted PP lead, as we have shown, to source monitoring and agency-attribution disturbances, but these emergent disturbances would themselves lead to fundamental shifts in the optimal weighting of evidence, with serious ensuing consequences for accurate modelling of the world. This arises, we argue, because source monitoring itself has an extremely important and far-reaching role to play in allowing us to make sense of the world and to ensure that we update our model only when necessary. We suggest this because a notion central to PP frameworks is that sensory evidence has the capacity to drive inference only insofar as it has sufficient precision or reliability to usefully update inference in the context of existing prior expectations, and because an accurate representation of the source of some evidence will very often be an important clue to estimating its reliability.

Different sources of a piece of evidence will influence how it drives updating in complex ways, being relevant both in making inferences about the meaning of the evidence, and about its significance. If you are at the pond in the park, a shout of ‘duck!’ will have importantly different meaning if it comes from your toddler or from the group of people behind you playing Frisbee. The source becomes key to the meaning of the evidence in this instance, while in other cases it is key to the weighting (Knill & Pouget 2004). A prediction error from a known reliable source should be weighted more strongly than one from a new source or a source that has proven itself unreliable. Again, this gets back to simple ideas of how cues are integrated depending on their reliability (Knill & Pouget 2004, Moore & Fletcher 2012) and a key point is not just what the evidence is, but the optimising meta-level judgments about how reliable that evidence is expected to be (Mathys 2011) and, by extension, the degree to which it should be weighted.

To summarise, a consideration of impairments in source monitoring capacities seems critical to understanding psychosis, not merely because such impairments can account descriptively for certain features of agency disruption that appear to be central to psychosis, but also because they would naturally feed into many of the other disturbances in inference and updating that are characteristic of the condition. We suggest that the PP framework provides a deeper understanding of source monitoring and offers a compelling account of how the hypothesised impairment in the processing of prediction errors could lead to agency disturbances and, via subtle changes in how sources of evidence are weighted, a profound change in how a person models and interprets their world. Note that, while the source monitoring literature has typically focussed on memory rather than on current perception, and on reality monitoring rather than on discrimination between possible external sources, the PP framework encompasses all of these within the same mechanisms of updating and inference. In the following sections, we explore how the PP framework may account for other aspects of psychosis, particularly those associated with the schizophrenia syndrome. We propose that PP has yet more to offer as a framework for thinking about key symptoms

and we begin by considering the often-neglected, yet clinically debilitating cognitive and negative symptoms. We then go on to discuss a central problem with earlier ideas of how a prediction error model could account for delusions: specifically, why, given that we are positing a very domain-general problem with reliability-weighting of information in Bayesian inference, delusions tend to be domain-specific in their content, which usually concerns 'the patient's place in the social world' (Bentall et al 1991). Third, we ask why delusions tend to have a negative emotional colouring. Finally, we consider how the proposed subtle disturbance to an essentially optimal system for performing rational inferences could ultimately lead to delusions that are so bizarre to the observer.

Episodic memory, Source monitoring and predictive processing: an account of cognitive deficits in schizophrenia

Ongoing assessments of the source of sensory evidence, and higher level estimates of that source's reliability, may have clear advantages for immediate decision-making. It is noteworthy too that long-term memories frequently incorporate such source information. A key component of episodic memory in particular is a broad range of accompanying source information. Indeed, source monitoring tasks draw on the same pool of processes and representations as do episodic memory tasks (Johnson 2005, Johnson 2006). It is relevant too that a meta-analysis by Achim and Weiss (2008) found that source-monitoring deficits in schizophrenia are not specific: rather, they are just one facet of a more general problem with associative memory. An important function of episodic memory is the facilitation of flexible, context-dependent behaviour (Suddendorf & Corballis 2007). The source of a memory – the what-where-when – may be of little value in learning and acting upon simple, non-stochastic associations and factual knowledge (e.g. that bushes yield berries), but it becomes key for dealing with higher order statistical regularities in the world (e.g. how the seasons impact the reliability of those yields). Consequently, source monitoring is crucial for flexibly adjusting the degree to which simpler, lower-order associations can dictate behaviour, or can be updated, in any particular context. It underpins the kind of context-dependent learning necessary to deal with complex, higher-order worldly contingencies, and with cues that have differential reliability and predictive power. Even if currently irrelevant for action choice, information about the source of an event may be useful in guiding decision-making when the event is later recalled – and importantly, it may be useful in ways that are not always predictable in advance: hence the importance of source memory for decision flexibility, in particular.

Given the importance of associative memory (for source and spatiotemporal context) in facilitating flexible behaviour, we advance the possibility that the PP explanation of a reduction in this capacity could have further explanatory power, accounting for difficulty in using source and context information to allow flexibility of thinking and decision-making. This could affect domains of belief (Woodward et al 2008, So et al 2012), cognition (Waltz 2016), thought (Lysaker & Lysaker 2002), speech (Barr et al 1989), mood (Lincoln et al 2015) and behaviour (Kraeplin1971, Ridley 1994.). We suggest that exploring these areas within the PP framework is necessary to fully realising its potential for explanations that go

beyond the positive features of psychosis: a critical need in future research (Insel 2010, Keefe and Fenton 2007, Kaneko and Keshevan 2012, Remington et al 2016).

A consideration of PP and source monitoring capacity is useful in considering the cognitive deficits and negative symptoms accompanying schizophrenia. Inaccuracies in internal source monitoring ('did I say that aloud, or did I just think it?') relate specifically to thought disorder (Nienow et al 2004), and to a specific type of communication disturbance - the 'missing information reference', whereby the subject refers to something which they have not mentioned before and which their interlocutor therefore cannot possibly know (Nienow et al 2005). Curiously, Brébion et al (2002) discovered that certain source monitoring errors that were *positively* associated with positive symptoms were *negatively* associated with negative symptoms: suggesting that 'positive and negative symptomatology appear to have opposite links to the source monitoring errors observed in patients with schizophrenia'. This result accords with the idea that negative symptoms arise as a result of over-compensative adaptation to a pattern of dysregulation that initially led to the development of positive symptoms.

Predictive processing, source monitoring and the social nature of delusions

The ability to identify the causal sources of current inputs is a key concern of PP, and may be crucial for flexible learning and behaviour. Moreover, storing and retrieving a representation of the source of a memory trace may be strongly relevant for learning, and (especially) generalisation. By identifying the source of some ostensibly predictive information, and retrieving that source memory at the time the outcome is expected, the system can make added use of any prediction error that then ensues. PE is useful not only in updating one's prior belief about the reliability of the information itself ('how reliably does this signal produce the outcome?') but also in updating a meta-level prior as to the general reliability of its source ('how reliable are signals from that source?).

This capacity may be particularly important and challenging in the social domain, and consequently any perturbations in source memory might have especially severe manifestations in social functioning. The advice and information people share with you now can turn out later to have been much more or less useful than you anticipated, and remembering who said what is important for prediction-error driven learning about who (not) to rely on, and who (not) to trust. Retrospectively inferring 'who said what' is likely to be a highly challenging task: a great deal of potentially-identifying perceptual information in the speech signal is filtered out during encoding in order to prioritise semantic processing (Johnson 2005), and there may be a very long delay before a piece of information is re-evoked and updated. Given that people with psychosis have difficulty discriminating the source of verbal memory traces (Brébion et al 2005), it follows that they would be compromised in their ability to identify an optimal balance between trusting and distrusting socially communicated information, depending on the reliability of the source (see Fonagy and Allison 2014). Just as reduced discriminability in PE signalling could lead to a consistent sense of unease or surprise, so too could reduced discriminability between social

Europe PMC Funders Author Manuscripts

sources make everything (and everyone) seem uniformly unreliable - even suspicious. This could account not only for the peculiar social content of delusions - widespread conspiracies involving a growing list of suspicious people - but also for their social form: delusions involve a kind of 'Global social estrangement…[and] a failure to treat others as reliable sources of information' (Ratcliffe 2015). Such a failure could partially explain why delusions persist, and may even strengthen and grow, in the face of friends' and doctors' efforts to refute them.

Related to this, it is worth noting that social cues may be inherently more uncertain than non-social ones, since they rely on inferring intentions from ambiguous physical acts. Consequently, representations of the social world could be the first to 'break' when the system encounters a relatively minor impairment in uncertainty-weighting inference. Indeed, because social evidence is so ambiguous and uncertain, high-level priors may be particularly influential in socially-based inferences, compared with physically-based inferences for which the evidence can be more precise and less susceptible to such re-shaping. To give a simple example, if I see someone waving his arm, I can update a social inference ('he is threatening rather than greeting me') quite easily without having to challenge the sensory evidence itself. The social world is highly volatile, and our access to it extremely noisy. Inferences about other people's intentions are inferences about deeply hidden states. To understand what another person is saying (let alone what they might be thinking but not saying) we need to make use of priors at multiple levels, and cues from facial expression, voice, body language and conversational context – all of which must be combined in a manner that is sensitive to higher-level expectations about their relative uncertainties. In short, a dysregulation in a very domain-general mechanism, such as we are proposing here, may well show a rather more specific manifestation, since its primary effects may prove most striking in those domains that pose the greatest challenge to the capacities supported by that mechanism's proper functioning.

This assertion echoes recent discoveries in the field of autism, based on longitudinal studies of children at high genetic risk. Elsabbagh and Johnson (2016) suggested that a 'domain general and diffuse' abnormality in synaptic connectivity and transmission has 'partial and mild effects on multiple systems…but [these] effects are only clearly revealed when more complex multisensory integration and high temporal resolution processing becomes important…[in] the increasingly complex social environment'. In psychosis, just as has been suggested for autism (Elsabbagh and Johnson 2016), the sometimes- profound loss of capacity (or confidence) in the social domain could lead to highly selective sampling of information, avoidance of especially uncertain sources of evidence, and a tendency to cut oneself off from a very valuable source of uncertainty resolution: the information and beliefs passed on to us by other people (Ratcliffe 2015). This kind of potentially maladaptive orientation away from the shared social world may underlie the development of negative symptoms such as asociality and alogia. Moreover, by isolating the individual from the usual interpersonal 'checks' on epistemic reasoning, it may also have the indirect effect of allowing idiosyncratic beliefs to develop and become entrenched.

Why is psychosis associated with negative affect and experience?

The PP model of psychosis, in proffering a general framework for understanding how such experiences and beliefs emerge, does not obviously account for one of the very striking features with which all clinicians will be familiar: people often suffer from these experiences. They are negative, frightening, unpleasant and depressing. Why should this be the case? After all, an imbalance within a PP system such that a person struggles to model the world and to update inferences does not necessarily mean that the ensuing models, although false, should also be unpleasant.

To begin, it is important to consider that the largely negative emotional tone of clinical psychosis may, in part at least, reflect the fact that such clinical characteristics are largely derived from studies of people who are help-seeking. That is, the psychiatrist tends not to see the person who remains unperturbed by their beliefs. Even in the instance of grandiose delusions, when the content of the belief can be highly positive, it is usually when there is some suffering being caused that the psychiatrist is called upon. In short, we should remind ourselves that a psychotic experience that comes to clinical attention may be highly selected.

However, there is a further consideration when trying to account for the content and the emotional hue of psychotic experiences. Notably, as pointed out elsewhere (Kapur 2003), the content must surely reflect the knowledge, experiences and preoccupations of the individual. This is an integral part of the PP framework: that the current model of the world, and the ways in which it is updated in response to sensory evidence, fundamentally depend on priors formed from past experiences, which shape current expectations at multiple levels.

We suggest that there is something to be learned from examining the manner in which personally meaningful prior experience comes to powerfully and inappropriately dominate current perception in post-traumatic stress disorder (PTSD). PTSD ‘flashbacks’ involve a transient psychosis-like state in which the person re-experiences a traumatic memory as though it were real, happening in the here and now. In essence this is a profound, state-dependent reality monitoring deficit in which the sufferer is unable, momentarily, to reliably distinguish memory from current reality, or mental images from real sensory evidence.

To some extent, the symptoms of psychosis may reflect a more subtle, but persistent, version of this same disturbance: a reality monitoring deficit conferring undue weight on memories in shaping the current world model. But the question remains as to why a general reality monitoring deficit, of the kind posited in psychosis, should disproportionately affect memories with a negative emotional colouration. One clue comes from a consideration of which cues are most important for discriminating imagined from externally-caused representations, and an examination of whether there are any specific characteristics of emotionally negative memories that might render these cues less reliable - thus accounting for why negative memories might prove especially vulnerable to external misattribution.

One of the most informative cues available for reality monitoring is a representation’s vividness. Mental images ‘seem to behave like weak versions of externally triggered perceptual representations’ (Pearson et al 2015). From a computational standpoint, the relatively low precision of mental imagery could explain its low probability of contributing

to updating priors. By analogy with self-generated movement, the precision of self-generated images may be 'dampened down' because they contain nothing newsworthy -their content is already highly predictable given the system's antecedent state – and their consequently reduced salience or vividness will be a powerful cue that they are indeed self-generated rather than externally caused (Frith 2005). Extending the analogy with the agency case, mentalmages whose content is for some reason poorly predicted by the current cognitive context will be experienced as more vivid, and thus particularly vulnerable to external misattribution. The idea that hallucinations result from vivid mental imagery being misattributed to an external cause is a long-standing one (Mintz & Alpert 1972), and indeed, increased vividness of imagery does seem to be a trait-like feature of schizophrenia (Sack et al 2005, Oertel et al 2009).

Traumatic memories may (perhaps because they are processed via an amygdalar rather than hippocampal-led route (LeDoux et al 1988) be particularly susceptible to undergoing arousal-dependent or cue-dependent intrusions into consciousness, in a format that is both phenomenologically vivid, and unpredictable given the context (Brewin 2001, Holmes & Steel 2004, Steel et al 2008). Given the centrality of unpredictability and vividness cues for successful reality monitoring, this will make memories of traumatic events particularly susceptible to external misattribution. Moreover, people with higher trait schizotypy may have a lower stress threshold beyond which memories are encoded via the amygdala-led route (perhaps due to problems with hippocampal-mediated associative memory), rendering these individuals particularly prone to frequently experience unpredictable, vivid intrusions of moderately threatening memory fragments into conscious awareness (Steel et al 2005, Helmsley et al 1994). Thus, a PP exploration of the phenomenological characteristics of psychosis (e.g. the experiences' unpleasantness) also has the potential to make explanatory links with its specific aetiological risk factors.

It is also noteworthy in this regard that the risk of psychotic disorder is elevated in people with a history of childhood trauma (Schafer & Fisher 2011, Varese et al 2012). Indeed, many sufferers see their current experiences as a re-enactment of previous traumatic events (Hardy et al 2005, Steel 2015). The PP framework offers a mechanism, via perturbed source and reality monitoring, by which these traumatic past events could come to influence the current world model so profoundly and distressingly.

Moreover, in relation to the discussion above about the social nature of delusions, Fonagy and Allison (2014) have offered a mechanistic account for how childhood trauma could lead to an inability to appropriately assign epistemic trust. Since neglect and abuse prevent the formation of specific attachment relations that differentially pick out particular adults who can be relied upon as sources of valuable, generalizable information, people who suffered neglect and abuse may have thereby formed a persisting meta-level prior that 'who said it' makes little difference: in other words, that social source is simply not a very useful guide to epistemic reliability, and therefore not worth carefully monitoring. This is admittedly speculative, but it highlights the potential of the PP account to offer a unified explanation for a remarkably wide array of the disorder's symptoms, features and risk factors.

Why are Delusions Bizarre?

Theories of psychosis can be criticised for failing to account for the content of delusional beliefs. In particular, when a delusion is theorised to be an essentially rational response to strange experiences (Kapur 2003, Maher 1974), it is hard to comprehend the highly improbable and, prima facie, irrational content of the belief itself ('I am being followed by spies') compared to other equally possible, and much more probable, interpretations ('I am suffering from a mental illness which affects my experience of the world').

This objection has been eloquently outlined by Coltheart (2007) who make a powerful case that a separate factor, in addition to that which generates the strange experiences,must be invoked to account for why such an unlikely belief is not rejected. McKay (2012) takes this further in relation to a Bayesian account of Capgras delusion (the belief that a loved one has been replaced by an imposter) proposed by Coltheart et al (2010). The idea that an altered physiological response to a familiar person confers a strange sense of unfamiliarity, in the context of preserved conscious recognition of the person, has been argued to provide the germ of the ultimate belief that the person has been replaced by an impostor (Ellis et al 1997). McKay questions whether the delusional inference ('this person is not my wife, she is an identical imposter') provides a reasonable explanation of this strange experience (McKay 2012). From a Bayesian perspective, the imposter hypothesis has such a low prior probability that it is hard to understand how it could be selected as the most likely cause of the experience. This is an important objection, and one that we must consider seriously in relation to the PP account of delusional beliefs.

We suggest that the incorporation of a more dynamic perspective on how perceptions and beliefs unfold over time, together with a greater appreciation of how predictions (particularly those experienced as being precise and reliable) can repeatedly shape sensory evidence, allows the PP framework to offer powerful explanatory insights into the question of why delusions are often so bizarre.

A key point, one that is made explicit in PP models, is that beliefs are not just 'caused' by evidence but themselves play a powerful subsequent role in determining how evidence is sampled, weighted and interpreted. High-level, abstract ideas can permeate down the representational hierarchy to disambiguate uncertainty at lower, more 'experiential' levels (Teufel & Nanay 2016). A delusion, once in place, may therefore become self-reinforcing as evidence is sampled and interpreted in accordance with expectations. Indeed, Schmack and colleagues (2013) demonstrated that the effect of high-level knowledge on perceptual processing, in healthy subjects, increases with their tendency towards delusional ideation – suggesting that highly schizotypal individuals are particularly prone to this self-reinforcing effect, which perhaps contributes to their elevated risk of psychotic disorder.

In acknowledging the dynamic nature of the updating process, we suggest that the very low prior probability of the imposter hypothesis is not ultimately a bar to its eventual acceptance. This is because the hypothesis doesn't have to be accepted wholesale on the basis of just the initial experience, nor even on the basis of a cumulative succession of identical experiences. The first experience may consist of nothing more strange or specific than a mild sense of

jamais vu that leads the man to fleetingly imagine the imposter idea without taking it seriously.

However, on the second occasion that he encounters his wife in the absence of the expected physiological response, PP provides a mechanism by which merely having entertained the imposter idea at a high level could be enough for the idea to imbue the content of the experience itself with some of its semantic content. Perhaps, this time, the man's wife seems not merely unfamiliar, but also suspicious (her smile now perceived perhaps as having a mocking or sinister quality) or slightly altered in physical appearance (e.g. she looks fractionally thinner or fatter than normal). In Bayesian terms, the imposter hypothesis' likelihood is higher, given this second experience, than it was given the first experience.

In addition, there is evidence that the mere act of imaginatively entertaining a hypothesis raises its subjective probability next time it is entertained (Arkes 1991, Garry et al 1996, Roediger & McDermott 1978). This 'imagination inflation' effect is stronger the more vividly the hypothesis was imagined (Sherman et al 1985) - suggesting that people with, or at risk for, schizophrenia may be particularly susceptible to it, due to their trait-like tendency to experience particularly vivid mental imagery (Oertel et al 2009). Interestingly, the 'imagination inflation' effect is greater for emotional interpersonal events than it is for neutral ones (Szpunar & Schachter 2013), perhaps contributing to the tendency for delusions to have a negative emotional tone. Schwartz (1998) frames these overall ideas eloquently: "belief is not an all-or-nothing phenomenon, but, rather, varies along a continuum from more to less…If we implicitly interpret our thoughts as evidence indicative of the presence in the world of that to which the thought corresponds, then it seems reasonable to assume that the stronger the presence of some information in our minds, i.e., the clearer, more distinct, and more familiar the thought, the more likely we are to accept the validity of the situation posited by the thought".

Note that the temporally extended PP account we have adumbrated in this section relates to a cognitive model proposed by Fleminger (1992), who refers to 'a morbid cycle of misperception and delusion reinforcing one another' occurring iteratively across time. Consistent with this overall picture, although there are reports of "autochthonous" delusions which appear to arise suddenly and fully-formed, the more characteristic pattern is that delusion formation is preceded by a prodromal period of 'delusional mood', involving a vague sense that the world has subtly changed somehow to become more sinister or foreboding. It is only over the course of time that these beliefs take shape.

One possible objection to this account might be that the postulated iterative process of self-reinforcement simply would not occur in a rational (perhaps Bayesian) brain. After all, a central tenet of PP is that the extent to which a high-level prior can influence perception depends crucially on its informational value, or epistemic reliability. However, in a system that is already noisy and uncertain, we may have a system that is "hungry for priors" (Dakin S, Personal communication) leading to excessive top down influences (Teufel et al, 2015).

A further relevant consideration relates to the links we have drawn between PP disturbances and source monitoring deficits. We have seen how a relatively subtle disturbance could give

rise to the deficits that people with psychosis demonstrate onsource monitoring tasks in which they have to inferentially discriminate whether a representation was externally or internally caused. It is likely that similar computations are involved in the important task of metacognitively discriminating between classes of internally-caused representations - in particular, between beliefs and imaginings (Johnson et al 1979, Johnson et al 1981). People with psychosis, therefore, may well have trouble making such discriminations accurately. Indeed, Currie (2000) has already proposed a 'metarepresentational' account of delusions as being due to the individual *imagining* some state of affairs (e.g., that his wife has been replaced by an imposter) and misidentifying this imaging *as a belief*. Such a tendency could arise from a relatively subtle disturbance in PP, of the same kind as we have suggested gives rise to the external misattribution bias on source monitoring tasks. Misattributions of this nature could have severe consequences since imaginings could become able to influence perception which would in turn support developing beliefs, via the process of gradual self-reinforcement that we have proposed.

Cautionary thoughts

It is important to stress that the PP framework is not proposed here as a replacement for descriptive theories of psychosis. This would be a category error. The source monitoring description has value when thinking about experiences and capacities that are of clear importance, in terms that are understandable to patients themselves. The observation that a deeper level of description may explain source monitoring more generally does not demonstrate that the latter is redundant. The predictive processing framework is geared towards understanding lower level processes that could account for source monitoring capacities and the specific nature of their derangement in psychotic illness. As we have tried to demonstrate in this paper, in seeking this level of understanding there is the possibility of attaining a more generalised explanation, one that goes beyond source deficits to account more comprehensively for other symptoms in a way that offers direct links to brain processes while remaining closely in touch with cognitive and social/environmental factors and occurrences.

But, in a field that has so many models, a number of questions present themselves: how do we know when to discard a particular model? To what extent are our models falsifiable? How do we decide which models most suit our current purposes? These are related questions and there are large and complex literatures on the nature and purposes of models, and onhow they may be compared (Bender 1978). As above, it would be misguided to compare directly the predictive processing and source monitoring accounts, because they are pitched at different levels. The circumstances and the questions being posed should indicate which level is currently appropriate, and the model can be selected accordingly. As an analogy, if we wish to use a model plane to characterise the aerodynamic properties of a particular shape of fuselage, we need not build a model of the internal structure, including seats, passengers and drinks trolleys. If, on the other hand, we wish to explore how weight distribution affects these properties, we would need to consider such details. The question of which aspects of reality a model should represent, and which it should it leave out, is unanswerable without a consideration of the model's purpose and intended level of explanation or description (see Teufel and Fletcher, Brain 2016 for full discussion).

Europe PMC Funders Author Manuscripts

To take a more partisan approach, we suggest that the predictive processing framework capitalises on at least three advantageous characteristics of formal computational modeling (Teufel and Fletcher 2016). First, in translating conceptualisations to a mathematical framework, it demands an explicitness and precision that conceptual/descriptive models do not. Second, it can stimulate new conceptualisations of how seemingly disparate phenomena (such as perception and belief) may be less distinct at a lower level of description. Third, it offers opportunities for bridging levels of understanding that extend beyond merely tabulating the neural correlates of a capacity or behaviour. The computational parameters invoked by predictive processing models appear, prima facie at least, to lend themselves readily to formulations in terms of brain processes. It is possible to consider information processing in terms of physical systems as well as cognitive ones (Frith 2008) and, in this respect, predictive processing relates more readily and directly to the brain.

Limitations and Future directions

For all of the advantages and attractions that we see in applying PP to understanding psychosis and its accompaniments, a full appraisal must acknowledge the incompleteness and potential pitfalls of this approach.

- Whatever level and vocabulary a model uses, it must be expressed with sufficient precision that one can envisage a set of observations that could lead one to alter or abandon it. Models must be breakable (Fletcher and Teufel, 2016). Indeed, it strikes us that the state of psychosis may also be conceived as incorporating a model of the world that does not break in response to contradictory evidence. Clearly, a model as general and, as yet, imprecise as predictive processing must beware of this. As yet, the application of PP models has been largely descriptive.While this research has established, as we hope has been shown here, an impressive degree of explanatory power, PP has yet to be fully and formally tested by examining novel predictions that it may make.
- While we have emphasised the potential for PP to link multiple levels of explanation in psychosis, there is a huge amount to be done in adding flesh to the basic framework. Moreover, as has been discussed elsewhere (Teufel and Fletcher, 2016), there are internal inconsistencies with respect to neurobiological implementations of PP that demand careful scrutiny. It is likely that opportunities for falsifying predictions may by PP models will naturally present themselves in the course of its maturation, as the boundaries of its explanatory scope become clearer and its predictions become more precise. An important part of the development of a model lies in its comparison with other models at the same level. In this regard, it should be noted that emerging work on circular belief propagation is producing elegant and impressive insight (). Theoretical developments should therefore include a deeper consideration of the links, and points of divergence, between these two sets of models.

Europe PMC Funders Author Manuscripts

Concluding thoughts

In science, all models are necessarily simplifications. Their usefulness may be judged by how well they account for and predict that particular aspect of reality with which they are concerned. The brain is no different. It must create, judge, compare and develop models of external reality and must be clear too about the aims and, importantly, the scope of each model. One ultimate consequence of a failure to do this is psychosis. Psychiatric research has produced a great many theoretical accounts of psychosis and of mental illness more generally. Different accounts can often seem to be in competition or conflict, when in fact they may show themselves to be complementary when we identify the particular level(s) at which their claims are intended to apply, and the particular characteristic(s) of symptoms with which they concern themselves. We have attempted to show that an account of psychosis as resulting from an imbalance in PP strongly accords with, and extends, a source monitoring impairment theory expressed at a higher level. The brain basis of source monitoring has been explored (Mitchell & Johnson 2009) but, without intervening levels of understanding, correlative links between brain activity and source monitoring capacities will remain frustratingly vague. Predictive processing, we suggest, provides a starting point for a computational understanding of these links, and of how they may be changed in psychosis. Moreover, the key components that are the concern of a hierarchical and dynamic PP framework – prediction error, reliability, uncertainty, adaptivity – readily translate to the subjective experiences that are key to source monitoring ideas: self versus other; real versus imagined; agentic versus non-agentic.

Put more simply, PP essays an account of how the brain optimally infers the causes of its noisy, unreliable and ambiguous inputs. The type of tasks explored in the source monitoring framework are important examples of this causal inference problem. Moreover, by scrutinising source monitoring afresh within the PP context, we are encouraged to appreciate more fully the value of source information in optimally modelling the world, and the implications of a source monitoring impairment in profoundly and pervasively affecting the modelling process itself, leading to models that are suboptimal and harmful in very specific ways.

Bibliography

Achim AM, Weiss AP. No evidence for a differential deficit of reality monitoring in schizophrenia: a meta-analysis of the associative memory literature. Cogn Neuropsychiatry. 2008; 13(5):369–84. [PubMed: 18781492]

Adams RA, Huys QJM, Roisier JP. Computational Psychiatry: towards a mathematically informed understanding of mental illness. J Neurol Neurosurg Psychiatry. 2015; 87:53–63. [PubMed: 26157034]

Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Front Psychiatry. 2013; 4(47)

Andreasen NC. Thought, language, and communication disorders: I. Clinical assessment, definition of terms, and evaluation of their reliability. Arch Gen Psychiatry. 1979; 36:1315–21. [PubMed: 496551]

Arguedas D, Stevenson RJ, Langdon R. Source monitoring and olfactory hallucinations in schizophrenia. J Abnorm Psychol. 2012; 121(4):936–43. [PubMed: 22428787]

Arkes HR, Boehm LE, Xu G. Determinants of judged validity. J Exp Soc Psych. 1991; 27(6):576–605.

Barch DM, Ceaser A. Cognition in Schizophrenia: Core Psychological and Neural Mechanisms. Trends Cogn Sci. 16(1):27–34.

Barr WB, Bilder RM, Goldberg E, Kaplan E, Mukherjee S. The neuropsychology of schizophrenic speech. J Communication Disorders. 1989; 22:327–349.

Behrens TE, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. Nat Neurosci. 2007; 10(9):1214–21. [PubMed: 17676057]

Bender, EA. An introduction to mathematical modelling. New York, NY: John Wiley & Sons; 1978.

Bentall RP, Kaney S, Dewey ME. Paranoia and social reasoning – an attribution theory analysis. Br J Clin Psychol. 1991; 30:13–23. [PubMed: 2021784]

Biro, S., Csibra, G., Gergely, G. The role of behavioural cues in understanding goal-directed actions in infancy. Progress in Brain Research. con Hofsten, C., Rosander, K., editors. Vol. 164. Amsterdam: Elsevier; 2007. p. 303-322.

Blakemore SJ, Wolpert DM, Frith CD. Abnormalities in the awareness of action. Trends Cogn Sci. 2002; 6(6):237–42. [PubMed: 12039604]

Botvinick M. Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. Cogn Affect Behav Neurosci. 2007; 7:356–66. [PubMed: 18189009]

Braver TS. The variable nature of cognitive control: a dual mechanisms framework. Trends Cogn Sci. 2012; 16:106–13. [PubMed: 22245618]

Brébion G, Gorman Jack M, Dolores Malaspina D, Amador X. A model of verbal memory impairments in schizophrenia: two systems and their associations with underlying cognitive processes and clinical symptoms. Psychol Med. 2005; 35(1):133–42. [PubMed: 15842036]

Brébion G, Gorman JM, Amador X, Malaspina D, Sharif Z. Source monitoring impairments in schizophrenia: characterisation and associations with positive and negative symptomatology. Psychiatry Res. 2002; 112(1):27–39. [PubMed: 12379448]

Brewin CR. A cognitive neuroscience account of posttraumatic stress disorder and its treatment. Behav Res Ther. 2001; 38:373–93.

Brown JW, Braver T. Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. Science. 2005; 307(5712):1118–21. [PubMed: 15718473]

Canon T. How Schizophrenia Develops: Cognitive and Brain Mechanisms Underlying Onset of Psychosis. Trends Cogn Sci. 2015; 19(12):744–56. [PubMed: 26493362]

Ceaser AE, Barch DM. Striatal Activity is Associated with Deficits of Cognitive Control and Aberrant Salience for Patients with Schizophrenia. Front Hum Neurosci. 2015; 9:687. [PubMed: 26869912]

Cicchino JB, Aslin RN, Rakison DH. Correspondences between what infants see and know about causal and self-propelled motion. Cognition. 2005; 118(2):171–92.

Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 2013; 36(3):181–204. [PubMed: 23663408]

Cole MW, Schneider W. The cognitive control network: Integrated cortical regions with dissociable functions. Neuroimage. 2007; 37:343–60. [PubMed: 17553704]

Coltheart M. The neuropsychology of delusions. Ann N Y Acad Sci. 2010; 1191:16–26. [PubMed: 20392273]

Coltheart M, Menzies P, Sutton J. Abductive inference and delusional belief. Cogn Neuropsychiatry. 2010; 15:261–87. [PubMed: 20017038]

Conant RC, Ashby WR. Every Good Regulator of a System Must be a Model of that System. Int J Syst Sci. 1970; 1(2):89–97.

Corlett PR, Fletcher PC. Computational psychiatry: A Rosetta stone linking the brain to mental illness. Lancet Psychiatry. 2014; 1(5):399–402. [PubMed: 26361002]

Corlett PR, Fletcher PC. Delusions and prediction error: clarifying the roles of behavioural and brain responses. Cogn Neuropsychiatry. 2015; 20(2):95–105. [PubMed: 25559871]

Corlett PR, Frith CD, Fletcher PC. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacol (Berl). 2009; 206(4):515–30.

Corlett PR, Honey GD, Aitken MR, Dickinson A, Shanks DR, et al. Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis. Arch Gen Psychiatry. 2006; 63(6):611–21. [PubMed: 16754834]

Corlett PR, Honey GD, Fletcher PC. Prediction error, ketamine and psychosis: An updated model. J Psychopharmacol. 2016

Corlett PR, Murray GK, Honey GD, Aitken MRF, Shanks DR, et al. Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. Brain. 2007; 130:2387–400. [PubMed: 17690132]

Culbreth AJ, Gold JM, Cools R, Barch DM. Impaired Activation in Cognitive Control Regions Predicts Reversal Learning in Schizophrenia. Schizophr Bull. 2016; 42(2):484–93. [PubMed: 26049083]

Currie G. Imagination, delusion and hallucinations. Mind Lang. 2000; 15:168–83.

Denève S, Jardri R. Circular inference: mistaken belief, misplaced trust. Curr Opin Behav Sci. 2016; 11:40–48.

Diederen KMJ, Spencer T, Vestergaard MD, Fletcher PC, Schultz W. Adaptive Prediction Error Coding in the Human Midbrain and Striatum Facilitates Behavioral Adaptation and Learning Efficiency. Neuron. 2016; 90(5):1127–38. [PubMed: 27181060]

Diederen KM, Schultz W. Scaling prediction errors to reward variability benefits error-driven learning in humans. J Neurophysiol. 2015; 114:1628–1640. [PubMed: 26180123]

Elliott R, McKenna PJ, Robbins TW, Sahakian BJ. Neuropsychological evidence for frontostriatal dysfunction in schizophrenia. Psychol Med. 1995; 25:619–630. [PubMed: 7480441]

Ellis HD, Young AW, Quayle AH, de Pauw KW. Reduced autonomic responses to faces in Capgras delusion. Proc R Soc Biol. 1997; 264:1085–92.

Elsabbagh M, Johnson MH. Autism and the social brain: the first-year puzzle. Biol Psychiatry. 2016; 80(2):94–99. [PubMed: 27113503]

Fiorillo CD, Tobler PN, Schultz W. Discrete coding of reward probability and uncertainty by dopamine neurons. Science. 2003; 299(5614):1898–902. [PubMed: 12649484]

Fleminger S. Seeing is believing: The role of 'preconscious' perceptual processing in delusional misidentification. Br J Psychiatry. 1992; 160:293–303. [PubMed: 1562856]

Fletcher PC, Frith CD. Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci. 2009; 10(1):48–58. [PubMed: 19050712]

Fonagy P, Allison E. The role of mentalizing and epistemic trust in the therapeutic relationship. Psychotherapy. 2014; 51:372–380. [PubMed: 24773092]

Friston K. The free-energy principle: a rough guide to the brain? Trends Cogn Sci. 2009; 13:293–301. [PubMed: 19559644]

Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci. 2010; 11(2):127–38. [PubMed: 20068583]

Friston K, Kiler J, Harrison L. A free energy principle for the brain 2006. J Physiol Paris. 2006; 100(1):70–87. [PubMed: 17097864]

Friston KJ. A theory of cortical responses. Philos Trans R Soc B. 2005; 360:815–36.

Friston KJ, Stephan KE. Free-energy and the brain. Synthese. 2009; 159(3):417–58.

Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. Lancet Psychiatry. 2014; 1(2):148–58. [PubMed: 26360579]

Frith C. Editorial: In praise of cognitive neuropsychiatry. Cogn Neuropsychiatry. 2008; 13(1):1–7. [PubMed: 18092222]

Frith C. The self in action: Lessons from delusions of control. Conscious Cogn. 2005; 14(4):752–70. [PubMed: 16098765]

Frith, CD. The Cognitive Neuropsychology of Schizophrenia. Hove: Lawrence Erlbaum Associates Ltd; 1992.

Frith CD, Blakemore SJ, Wolpert DM. Abnormalities in the awareness and control of action. Philos Trans R Soc B. 2000; 355(1404):1771–88.

Frith CD, Done DJ. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. Psychol Med. 1989; 19(2):359–63. [PubMed: 2762440]

Garry M, Manning CG, Loftus EF, Sherman SJ. Imagination Inflation: Imagining a Childhood Even Inflates Confidence that it Occurred. Psychon Bull Rev. 1996; 3(2):208–14. [PubMed: 24213869]

Gray JA, Feldon J, Rawlins NP, Hemsley DR, Smith AD. The neuropsychology of schizophrenia. Behav Brain Sci. 1991; 14(1):1–20.

Hardy A, Fowler D, Freeman D, Smith B, Steel C, et al. Trauma and hallucinatory experience in psychosis. J Nerv Ment Dis. 2005; 193(8):501–7. [PubMed: 16082293]

Haug E, Lien L, Raballo A, Bratlien U, Oie M, et al. Selective aggregation of self-disorders in first-treatment DSM-IV schizophrenia spectrum disorders. J Nerv Ment Dis. 2012a; 200:632–36. [PubMed: 22759943]

Haug E, Melle I, Andreassen, Raballo A, Bratlien U, et al. The association between anomalous self-experience and suicidality in first-episode schizophrenia seems mediated by depression. Compr Psychiatry. 2012b; 53(5):456–60. [PubMed: 21871617]

Haug E, Oie M, Andreassen OA, Bratlien U, Rabalo A, et al. Anomalous self-experiences contribute independently to social dysfunction in the early phases of schizophrenia and psychotic bipolar disorder. Compr Psychiatry. 2014; 55(3):475–82. [PubMed: 24378241]

Heinz A, Schlagenhauf F. Dopaminergic dysfunction in schizophrenia: salience attribution revisited. Schizophr Bull. 2010; 36:472–85. [PubMed: 20453041]

Helmsley DR. The schizophrenic experience: Taken out of context? Schizophr Bull. 2005a; 31(1):43–53. [PubMed: 15888424]

Hemsley DR. A cognitive model for schizophrenia and its neural basis. Acta Psychiatr Scand. 1994; 90:80–86.

Hemsley DR. The development of a cognitive model of schizophrenia: placing it in context. Neurosci Biobehav Rev. 2005b; 29(6):977–88. [PubMed: 15964074]

Hohwy, J. The predictive mind. Oxford: Oxford University Press; 2013.

Holmes EA, Steel C. Schizotypy as a vulnerability factor for traumatic intrusions: an analogue investigation. J Nerv Ment Dis. 2004; 192:28–34. [PubMed: 14718773]

Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016; 19:404–13. [PubMed: 26906507]

Insel TR. Rethinking schizophrenia. Nature. 2010; 468:187–93. [PubMed: 21068826]

Jardri R, Denéve S. Circular inferences in schizophrenia. Brain. 2014; 136(11):3227–41.

Johnson, K. Speaker Normalization in speech perception. The Handbook of Speech Perception. Pisoni, DB., Remezz, R., editors. Oxford: Blackwell Publishers; 2005. p. 363-389.

Johnson MK. The relation between source memory and episodic memory: Comment on Siedlecki et al. Psychol Aging. 2005; 20:529–531. [PubMed: 16248712]

Johnson MK. Memory and reality. Am Psychol. 2006; 61:760–771. [PubMed: 17115808]

Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychol Bull. 1993; 114:3–28. [PubMed: 8346328]

Johnson, MK., Hirst, W. MEM: Memory subsystems as processes. Theories of Memory. Collins, AF.Gathercole, SE.Conway, MA., Morris, PE., editors. East Sussex, England: Erlbaum; 1993. p. 241-286.

Johnson MK, Raye CL, Foley HJ, Foley MA. Cognitive operations and decision bias in reality monitoring. Am J Psychol. 1981; 94:37–64.

Johnson MK, Raye CL, Wang AY, Taylor TH. Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. J Exp Psychol: Hum Learn Mem. 1979; 5:229–40.

Kaneko Y, Keshavan M. Cognitive remediation in schizophrenia. Clin Psychopharmacol Neurosci. 10(3):125–35.

Kapur S. Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry. 2003; 160(1):13–23. [PubMed: 12505794]

Keefe RS, Arnold MC, Bayen UJ, Harvey PD. Source monitoring deficits in patients with schizophrenia; a multinomial modelling analysis. Psychol Med. 1999; 29:903–14. [PubMed: 10473317]

Keefe RS, Fenton WS. How should DSM-V criteria for schizophrenia include cognitive impairment? Schizophr Bull. 2007; 33(4):912–20. [PubMed: 17567627]

Kerns JG, Berenbaum H. Cognitive impairments associated with formal thought disorder in people with schizophrenia. J Abnorm Psychol. 2002; 111:211–24. [PubMed: 12003444]

Kirschner M, Hager OM, Bischof M, Hartmann-Riemer MN, Kluge A. Deficits in context-dependent adaptive coding of reward in schizophrenia. NPJ Schizophr. 2016; 2:16020. [PubMed: 27430009]

Knill DC, Pouget A. The Bayesian brain: The role of uncertainty in neural coding and computation. Trends Neurosci. 2004; 27(12):712–19. [PubMed: 15541511] a Kraepelin, E. Dementia Praecox and Paraphrenia. Barclay, RM., translator. New York: Robert E. Krieger; 1976.

Leavitt VM, Goldberg TE. Episodic memory in schizophrenia. Neuropsychol Rev. 2009; 19(3):312–23. [PubMed: 19639413]

LeDoux JE, Iwata J, Cicchetti P, Reis DJ. Different projections of the central amygdaloid nucleus mediate autonomic and behavioural correlates of conditioned fear. J Neurosci. 1998; 8:2517–29.

Lincoln TM, Hartmann M, Kother U, Moritz S. Do people with psychosis have specific difficulties regulating emotions? Clin Psychol Psychother. 2015; 22:637–46. [PubMed: 25256563]

Lindsay DS, Johnson MK. False memories and the source monitoring framework: Reply to Reyna and Lloyd (1997). Learn Individ Differ. 2000; 12:145–161.

Lysaker PH, Lysaker JT. Narrative structure in psychosis: Schizophrenia and disruptions in the dialogical self. Theory & Psychol. 2002; 12:207–22.

Lysaker PH, Vohs J, Minor KS, Irarrazaval L, Leonhardt B, Hamm J, et al. Metacognitive Deficits in Schizophrenia Presence and Associations With Psychosocial Outcomes. J Nerv Ment Dis. 2015; 203:530–36. [PubMed: 26121151]

Maher BA. Delusional thinking and perceptual disorder. J Indiv Psychol. 1974; 30(1):98–113.

Mathys C, Daunizeau J, Friston KJ, Stephan KE. A Bayesian foundation for individual learning under uncertainty. Front Hum Neurosci. 2011; 5(39)

McKay R. Delusional inference. Mind Lang. 2012; 27(3):330–55.

Miller R. Schizophrenic psychology, associative learning and the role of forebrain dopamine. Med Hypotheses. 1976; 2(5):203–211. [PubMed: 9558]

Mintz S, Alpert M. Imagery vividness, reality testing, and schizophrenic hallucinations. J AbnormPsychol. 1972; 79(3):310–16.

Mitchell KJ, Johnson MK. Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? Psychol Bull. 2009; 135:638–677. [PubMed: 19586165]

Moore, BCJ. An Introduction to the Psychology of Hearing. Sixth Edition. Bingley, England: Emerald Group Publishing Ltd; 2012.

Moore JW, Fletcher PC. Sense of agency in health and disease: a review of cue integration approaches. Conscious Cogn. 2012; 21(1):59–68. [PubMed: 21920777]

Moseley P, Ellison A, Fernyhough C. Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. Neurosci Biobehav Rev. 2013; 37:2794–805. [PubMed: 24125858]

Nassar M, Wilson R, Heasly B, Gold J. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J Neurosci. 2010; 30(37):12366–78. [PubMed: 20844132]

Nelson B, Parnas J, Sass LA. Disturbance of minimal self (ipseity) in schizophrenia: clarification and current status. Schizophr Bull. 2014b; 40:479–82. [PubMed: 24619534]

Nelson B, Thompson A, Yung AR. Basic self-disturbance predicts psychosis onset in the ultra high risk for psychosis “prodromal” population. Schizophr Bull. 2012; 38:1277–1287. [PubMed: 22349924]

Nelson B, Whitford TJ, Lavoie S, Sass LA. What are the neurocognitive correlates of basic self-disturbance in schizophrenia?: Integrating phenomenology and neurocognition. Part 1 (Source monitoring deficits). Schiz Res. 2014a; 152(1):12–19.

Nienow TM, Docherty NM. Internal source monitoring and thought disorder in schizophrenia. J Nerv Ment Dis. 2004; 192:696–700. [PubMed: 15457113]

Nienow TM, Docherty NM. Internal source monitoring and communication disturbance in patients with schizophrenia. Psychol Med. 2005; 35:1717–26. [PubMed: 16300687]

Nordgaard J, Parnas J. Self-disorders and the schizophrenia spectrum: a study of 100 first hospital admissions. Schizophr Bull. 2014; 40(6):1300–07. [PubMed: 24476579]

Oertel V, Rotarska-Jagiela A, van de Ven V, Haenschel C, Grube M. Mental imagery vividness as a trait marker across the schizophrenia spectrum. Psychiatry Res. 2009; 167(1):1–11. [PubMed: 19345421]

Pacherie E, Green M, Bayne T. Phenomenology and delusions : Who put the 'alien' in alien control? Conscious Cogn. 2006; 15(3):566–77. [PubMed: 16403655]

Pantelis C, Barnes TR, Nelson HE, Tanner S, Weatherley L, Owen AM, Robbins TW. Frontal-striatal cognitive deficits in patients with chronic schizophrenia. Brain. 1997; 120(10):1823–43. [PubMed: 9365373]

Parnas J, Handest P, Saebye D, Jansson L. Anomalies of subjective experience in schizophrenia and psychotic bipolar illness. Acta Psychiatr Scand. 2003; 137(6):434–25.

Parnas J, Henriksen MG. Subjectivity and schizophrenia: another look at incomprehensibility and treatment nonadherence. Psychopathology. 2013; 46:320–29. [PubMed: 23860468]

Parnas J, Raballo A, Handest P, Jansson L, Vollmer-Larsen A, Saebye D. Self-experience in the early phases of schizophrenia: 5-year follow-up of the Copenhagen Prodromal Study. World Psychiatry. 2011; 10:200–04. [PubMed: 21991279]

Pearson J, Naselaris T, Holmes EA, Kosslyn SM. Mental imagery: Functional mechanisms and clinical applications. Trends Cogn Sci. 2015; 19(10):590–602. [PubMed: 26412097]

Pezzulo G. Why do you fear the Bogeyman? An embodied predictive coding model of perceptual inference. Cogn Affect Behav Neurosci. 2013; 14(3):902–11.

Premack D. The infant's theory of self-propelled objects. Cognition. 1990; 36:1–16. [PubMed: 2383967]

Preuschoff K, Bossaerts P. Adding prediction risk to the theory of reward learning. Ann N Y Acad Sci. 2007; 1104:135–46. [PubMed: 17344526]

Ratcliffe M. The interpersonal world of psychosis. World Psychiatry. 2015; 14(2):176–78. [PubMed: 26043330]

Remington G, Foussias G, Fervaha G, Agid O, Takeuchi H, Lee J, Hahn M. Treating negative symptoms in schizophrenia: An update. Curr Treat Options Psychiatry. 2016; 3:133–150. [PubMed: 27376016]

Rescorla, RA., Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II. Black, AH., Prokasy, WF., editors. New York: Appleton Century Crofts; 1972. p. 64-99.

Ridley RM. The psychology of perseverative and sterotyped behaviour. Prog Neurobiol. 1994; 44:221–31. [PubMed: 7831478]

Roediger HL, McDermott KB. Creating false memories: remembering words not presented in lists. J Exp Psychol: Learn Mem Cogn. 1995; 21:803–14.

Roessler, J. Thought insertion, self-awareness and rationality. Oxford Handbook of the Philosophy of Psychiatry. Fulford, B.Davies, M.Gipps, R., et al., editors. Oxford: Oxford University Press; 2013.

Rushworth MFS, Behrens TE. Choice, uncertainty, and value in prefrontal and cingulate cortex. Nat Neurosci. 2008; 11:389–97. [PubMed: 18368045]

Sack AT, van de Ven V, Etschenberg S, Schatz D, Linden DEJ. Enhanced vividness of mental imagery as a trait marker of schizophrenia? Schizophr Bull. 2005; 31(1):97–104. [PubMed: 15888429]

Schafer I, Fisher HL. Childhood trauma and psychosis – what is the evidence. Dialogues Clin Neurosci. 2011; 13(3):360–65. [PubMed: 22033827]

Schlagenhauf F, Huys QJM, Deserno L, Rapp MA, Beck A, et al. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. Neuroimage. 2014; 89:171–80. [PubMed: 24291614]

Schmack K, Gomez-Carrillo de Castro A, Rothkirch M, Sekutowicz M, Rossler H, Haynes JD, Heinz A, Petrovic P, Sterzer P. Delusions and the role of beliefs in perceptual inference. J Neurosci. 2013; 33(34):13701–12. [PubMed: 23966692]

Schneider, K. Clinical Psychopathology. N.Y.: Grune & Stratton; 1959.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275(5306):1593–99. [PubMed: 9054347]

Schwartenbeck P, FitzGerald TH, Dolan R. Neural signals encoding shifts in beliefs. Neuroimage. 2016; 125:578–86. [PubMed: 26520774]

Schwartz, DA. Indexical constraints on symbolic cognitive functioning. Proceedings of the twentieth annual conference of the cognitive science society. Proceedings of the 20th annual meeting of the Cognitive Science Society. Gernsbacher, MA., editor. Mahwah, N.J.: Derry SJ, Erlbaum; 1998. p. 935-938.

Shohamy D, Mihalakos P, Chin R, Thomas B, Wagner AD, Tamminga C. Learning and generalization in schizophrenia: effects of disease and antipsychotic drug treatment. Biol Psychiatry. 2010; 67:926–32. [PubMed: 20034612]

Shergill SS, Samson G, Bays PM, Frith CD, Wolpert DM. Evidence for sensory prediction deficits in schizophrenia. Am J Psychiatry. 2005; 162(12):2384–86. [PubMed: 16330607]

Shergill SS, White TP, Joyce DW, Bays PM, Wolpert DM, Frith CD. Functional Magnetic Resonance Imaging of Impaired Sensory Prediction in Schizophrenia. JAMA Psychiatry. 2014; 71(1):28–35. [PubMed: 24196370]

So SH, Freeman D, Dunn G, et al. Jumping to Conclusions, a Lack of Belief Flexibility and Delusional Conviction in Psychosis: A Longitudinal Investigation of the Structure, Frequency, and Relatedness of Reasoning Biases. J Abnorm Psychol. 121(1):129–39.

Steel C, Fowler D, Holmes EA. Trauma-related intrusions and psychosis: An information processing account. Behav Cogn Psychother. 2005; 33:139–52.

Steel C. Hallucinations as a trauma-based memory: implications for psychological interventions. Front Psychol. 2015; 6:1262. [PubMed: 26441698]

Stephens GL, Graham G. Self-Consciousness, Mental Agency, and the Clinical Psychopathology of Thought Insertion. Philos Psychiatry Psychol. 1994; 1:1–12.

Suddendorf T, Corballis MC. The evolution of foresight: What is mental time travel and is it unique to humans? Behav Brain Sci. 2007; 30:299–313. [PubMed: 17963565]

Synofzik M, Their P, Leube DT, Schlotterbeck P, Lindner A. Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. Brain. 2010; 133:262–71. [PubMed: 19995870]

Szpunar KK, Schacter DL. Get real: Effects of repeated simulation and emotion on the perceived plausibility of future experiences. J Experimental Psychol: Gen. 2013; 142:323–27.

Teufel C, Fletcher PC. The promises and pitfalls of applying computational models to neurological and psychiatric disorders. Brain. (in press).

Teufel C, Nanay B. How to (and how not to) think about top-down influences on visual perception. Conscious Cogn. 2016

Varese F, Smeets F, Drukker M, Lieverse R, Lataster T, et al. Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. Schizophr Bull. 2012; 38(4):661–71. [PubMed: 22461484]

Vinckier F, Gaillard R, Palminter S, Rigoux L, Salvador A, et al. Confidence and psychosis: a neuro-computational account of contingency learning disruption by NMDA blockade. Mol Psychiatry. 2015; 21:945–55.

Voss M, Moore J, Hauser M, Gallinat J, Heinz A, Haggard P. Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. Brain. 2010; 133:3104–12. [PubMed: 20685805]

Waltz JA, Kasanova Z, Ross TJ, Salmeron BJ, McMahon RP, Gold JM, Stein EA. The roles of reward, default, and executive control networks in set-shifting impairments in schizophrenia. PLoS One. 2013; 8:57257.

Waltz JA. The neural underpinnings of cognitive flexibility and their disruption in psychotic illness. Neuroscience. 2016

Wilkinson S. Levels and kinds of explanation: lessons from neuropsychiatry. Front Psychol. 2014; 5:373. [PubMed: 24808882]

Wolpert DM, Flanagan JR. Motor prediction. Curr Biol. 2001; 11(18):729–32.

Woodward, TS., Menon, M. Misattribution Models (II): Source Monitoring in Hallucinating Schizophrenia Subjects. The Neuroscience of Hallucinations. Jardri, R.Cachia, A.Thomas, P., Pins, D., editors. New York: Springer; 2012. p. 169-184.

Woodward TS, Moritz S, Menon M, Klinge R. Belief inflexibility in schizophrenia. Cognitive Neuropsychiatry. 2008; 13:267–77. [PubMed: 18484291]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Summary

- Predictive processing models of brain function, though diverse and expressed at somewhat different levels, have at their core the idea that the brain engages in prediction-based inferences about the causes of its sensory inputs. Without such experience-based prediction, these inputs by themselves are essentially ambiguous. Within predictive processing models, what we experience on a moment-to-moment basis largely reflects our informed predictions of the world rather than its direct reality.
- A great deal of recent work has attempted to understand the altered reality of psychosis in terms of a shift in predictive processing. Delusions and hallucinations - false or sub-optimal inferences about the world- are held to result from a (potentially subtle) alteration in the ability to integrate incoming data with predictions based on prior knowledge.
- Earlier formulations of these ideas, including those invoking altered prediction error signalling as a fundamental driver of the altered integrative processes, provided a limited account of psychosis. In this paper we show how the further consideration of the predictive processing system in terms of its hierarchical arrangement and its dynamic adaptivity improves substantially upon these accounts. This extended account provides an explanation of certain key clinical features of psychotic experiences that would otherwise be very puzzling, (such as their bizarreness and negative emotional content) and accounts for typical evolution of these experiences over time.
- In developing the model in this way, we also highlight the degree to which predictive processing ideas provide a deeper and more comprehensive understanding of source monitoring processes and the consequence of their derangement in psychosis.
- We argue that the predictive processing and source monitoring accounts are deeply compatible. The kind of high-level source monitoring deficits documented in the empirical literature are precisely those that would be expected to emerge, given the underlying computational abnormalities postulated by predictive processing accounts. Further, these emergent source monitoring problems would themselves impact predictive processing as part of a vicious circle that may go a long way towards explaining the phenomenology.
- Having established the relationship between source monitoring and predictive processing ideas, we show further how the latter may provide a comprehensive account of features of psychosis through their impact on a person’s capacity to identify and use source information in distinguishing internally- from externally-generated as well as agentic from non-agentic inputs.

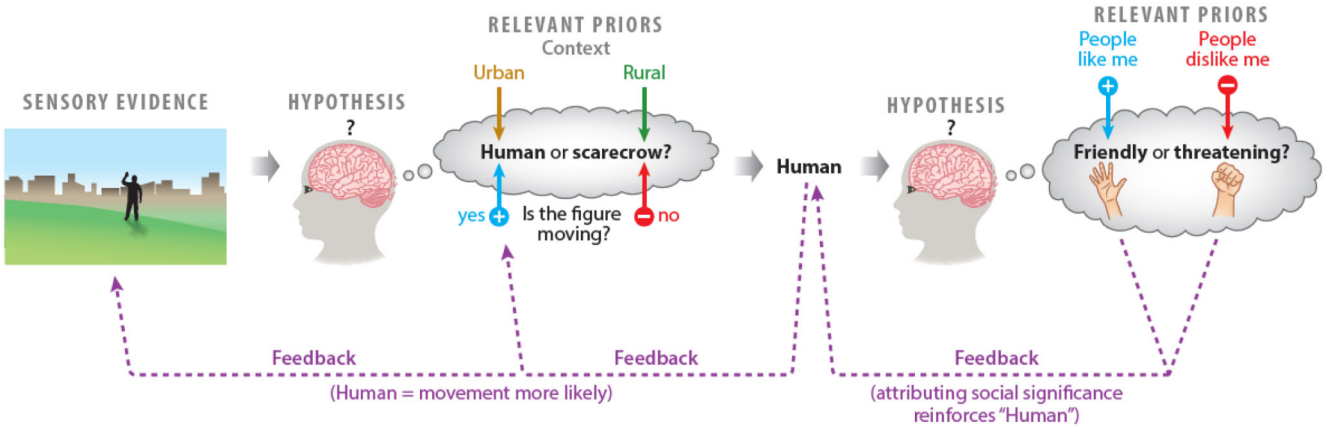

Figure.

This toy example illustrates how sensory evidence is shaped by expectations, and how ensuing inferences govern further evidence-gathering biased towards confirming initial hypotheses.

It shows, in simplified form, a recursive stream of processing leading from ambiguous sensory experience to an inference that one is being threatened. Several levels of inference are depicted, with top-down expectations based on past experience (referred to as "priors") shaping the interpretation of bottom-up inputs from each previous level. Thus, the very first impression of the distant silhouette ("is it a human or a scarecrow?") will be affected by the context (a scarecrow perhaps becomes more probable in an isolated rural setting), and guided by top-down predictions about which bottom-up signals will most reliably discriminate between likely candidate hypotheses. (Here for example, the "human" and "scarecrow" hypotheses generate very different predictions about the presence of agentic movement – motion signals are therefore highly informative (precise), and are upregulated accordingly by top-down gain control mechanisms including gaze direction and covert attention)

Importantly, the winning inference (here, human) generates modified expectations for ensuing bottom-up input (for example, movement may perhaps be more readily perceived and attended to when the over-arching hypothesis is "human"). Put simply, the inference that the causal source of the sense data is "human" means that further inputs are processed in a way that tend to confirm "human" as the correct hypothesis.

Moreover, the winning hypothesis generates specific predictions about which aspects of the percept are behaviourally relevant, constraining the array of hypotheses entertained at the next level of inference (for example, the attribution of social significance becomes relevant if the existence inference includes "human" and "movement"). And, at this higher level, different priors play a role in inferring the social significance. Importantly, as well as leading to a further updating of ever higher-level inferences, the interpretation of social intent provides further confirmatory evidence that the lower-level inferences ("movement" and "human") are correct and meaningful. After all, if the percept has social significance it confirms that the figure is human, and if it is human, movement is highly probable.