Explaining delusions: a cognitive perspective

Vaughan Bell¹, Peter W. Halligan² and Hadyn D. Ellis²

¹Institute of Psychiatry, King’s College London, Psychology Department Box 78, De Crespigny Park, London, SE5 8AF, UK
²School of Psychology, Cardiff University, Park Place, Cardiff, CF10 3YG, UK

There is now considerable evidence for reasoning, attention, metacognition and attribution biases in delusional patients. Recently, these findings have been incorporated into a number of cognitive models that aim to explain delusion formation, maintenance and content. Although delusions are commonly conceptualized as beliefs, not all models make reference to models of normal belief formation. This review considers those models that explain delusions as a breakdown of normal belief formation (belief-positive models), approaches that explain the pathology only (belief-negative models) and approaches that view delusions as one end of a distribution of anomalous mental phenomena (the continuum view). A cognitive theory that includes the ‘pragmatic pathology’ of delusions will be able to address both the phenomenology and the treatment of delusion-related distress.

Introduction

Delusions are a key clinical manifestation of psychosis and have particular significance for the diagnosis of schizophrenia. Although common in several psychiatric conditions, they also occur in a diverse range of other disorders (including brain injury, intoxication and somatic illness). Delusions are significant precisely because they make sense for the believer and are held to be evidentially true, often making them resistant to change.

Although an important element of psychiatric diagnosis, delusions have yet to be adequately defined (see existing diagnostic criteria; Box 1). Despite this operational ambiguity, the concept of a ‘delusion’ as a core psychopathological feature, indicative of a substantial break with reality (where patients might claim to be under the influence of non-existent or impossible machines or persecuted by invisible malign agents, to name but two examples; see Table 1) continues to have widespread clinical acceptance.

Notwithstanding issues of definition, the last decade has witnessed a particular intensification of research on delusions, with cognitive neuroscience-based approaches providing increasingly useful and testable frameworks from which to construct a better understanding of how cognitive and neural systems are involved [1]. These in turn have informed cognitive behavioural therapy, which has been shown to be an effective therapeutic approach [2].

A number of models of delusion formation have arisen from these research efforts, although little attempt has been made to bring together approaches from disparate traditions to assess critically their assumptions and contributions to a wider understanding of the phenomena. This review evaluates these models and draws attention to ambiguities in the empirical evidence, as well as highlighting areas needing further theoretical development.

Evidence for cognitive dysfunction and bias in delusions

Considerable data have been gathered over the past decade that suggest the presence of cognitive biases in people with delusions or those prone to delusional thinking. As many of the broader models of delusion formation draw on the same evidence (despite differing interpretations) recent work in this area is briefly reviewed to provide a context for the later discussion of the competing theoretical approaches.

Probabilistic reasoning

One early finding was that delusional patients showed a ‘jumping-to-conclusions’ (JTC) reasoning bias, whereby initial probabilistic estimates and the subsequent revision of hypotheses were made on less evidence than that required by controls [3]. Recent studies have shown this association to be more complex, however, as a JTC bias has also been found in non-delusional patients with schizophrenia [4], patients where the delusions had remitted [5] and unaffected first-degree relatives of psychotic individuals [6]. Although these studies suggest that JTC bias may be a ‘trait’, rather than a ‘state’ variable, relationships have also been found between strength of delusional conviction [7], level of psychosis-proneness [6,8] and presence of acute delusional symptoms [4], suggesting that the JTC reasoning bias might be related to delusional thinking in a dose-response relationship. Although usually represented as evidence for a probabilistic reasoning bias, it is still not clear how well the JTC evidence generalizes to all probabilistic reasoning, given the paucity of good research studies in this area.

Attributional style

As with JTC biases, attributional biases have been extensively studied, although particularly with regard to their role in persecutory delusions. Bentall et al. [9] have
Box 1. Problems defining delusions

The DSM diagnostic criteria for a belief to be considered delusional consists of the following:
‘A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everybody else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person’s culture or subculture (e.g. it is not an article of religious faith).’

This concise definition belies the fact each of the diagnostic criteria is either incoherent or subject to significant counter-examples. Delusions might not necessarily be false beliefs [63,64]. They could be value judgements (e.g. ‘I am a fantastically talented poet’), not amenable to practical falsification (e.g. ‘Satan is listening to my thoughts’), or be untestable in practice as the available relevant evidence is either limited, cannot be ascertained within the confines of the consulting room, or lies beyond the forensic capabilities of the clinician. On occasion, delusions can turn out to be ‘serendipitously’ true, as is sometimes the case with delusions of marital infidelity (‘Othello syndrome’). Delusions might not necessarily be about ‘external reality’ as demonstrated by passivity delusions of thought or action control, that typically involve claims of wildly abnormal or impossible mental states. A study by Myin-Germeys et al. [65] that sampled patients’ experiences established that conviction in a delusion’s veracity can vary even over the course of a day, suggesting that delusions are not always firmly sustained. Garety et al. [7] reported that about half of delusional patients will accept the possibility that they might be mistaken about their beliefs; and Kuhn [66] famously reported that scientists can hold fixed, incorrigible beliefs about external reality, despite overwhelming evidence to the contrary and without being considered delusional (at least, by clinicians). This suggests that resistance to widely accepted contradictory evidence is not a defining feature of delusions. The criterion that a delusion is ‘not (a belief) ordinarily accepted by other members of the person’s culture or subculture’ is typically not based on empirical evidence of how widely accepted a belief might be. Furthermore, there is recent evidence that subcultures can be based on the content of delusional beliefs [67], violating this aspect of the definition.

distinguished between an externalizing bias (a general tendency for non-self attributions, including people and circumstances, for negative events), and a personalizing bias (a specific tendency to blame others for negative events). The evidence for both biases being consistently present in deluded patients is equivocal, however; although the existence of an externalizing bias for negative events seems to be a more robust finding [10]. Little evidence for such biases is found in studies on non-delusional individuals with paranoid traits, including groups of non-clinical participants [11], or people with Asperger’s syndrome [12,13]. One further study did find externalizing biases in those with moderate or severe, but not mild, persecutory delusions [14], suggesting that the effect might be specific to the extent of psychosis.

Attention and metacognition

It is now well accepted that patients with persecutory delusions preferentially attend to threatening stimuli [1,9]. Studies investigating the link between latent inhibition (the effect whereby pre-exposure to a stimulus impairs its later association with another stimulus) and the wider concepts of psychosis or psychosis-proneness have typically found that latent inhibition is particularly disrupted in acute psychosis. This effect has also been found in healthy controls who score highly on measures of schizotypy, but not usually in patients in the chronic phases of illness, suggesting an impairment in attentional filtering might underlie symptom development [15]. In studies looking at metacognitive performance, several have shown impaired source monitoring (the ability to distinguish internally from externally generated experiences) in patients with both chronic and first-episode delusions [16,17,18]; although it should be noted that these studies typically consider delusions and hallucinations together as ‘reality distortion’, or have found similar effects for each. Reviews of the literature on putative impairments to ‘theory of mind’ (ToM; the ability to represent others’ mental states) typically suggest that, although often present in delusional patients, ToM impairments can resolve after remission [19,20], and are generally not present in unaffected relatives [21], suggesting a state-like relationship.

Table 1. Commonly reported themes of clinical delusions

<table>
<thead>
<tr>
<th>Theme</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defined by DSM</strong></td>
<td></td>
</tr>
<tr>
<td>Persecutory</td>
<td>“My food is being poisoned by the police”</td>
</tr>
<tr>
<td>Grandiose</td>
<td>“I have the power to heal all illnesses”</td>
</tr>
<tr>
<td>Jealous (Othello syndrome)</td>
<td>“My partner is cheating on me”</td>
</tr>
<tr>
<td>Erotomanic (De Clerambault’s syndrome)</td>
<td>“A famous pop star secretly signals her love to me over the radio”</td>
</tr>
<tr>
<td>Somatic (e.g. delusional parasitosis/Ekbom’s syndrome)</td>
<td>“I am infected by tiny parasites”</td>
</tr>
<tr>
<td>Bizarre</td>
<td>“My mother’s thoughts are being carried on raindrops that fall on the air conditioner”</td>
</tr>
</tbody>
</table>

| Misidentification | |
| Capgras syndrome | “My relatives have been replaced by identical looking impostors” |
| Fregoli syndrome | “The same person is disguising himself as others” |
| Reduplicative paramnesia | “My present location exists in two places simultaneously” |
| Mirrored self-misidentification | “The reflection in the mirror is another person” |

| Other | |
| Thought insertion/withdrawal | “Thoughts are being inserted into/withdrawn from my mind” |
| External control | “My mind/body is being controlled by an external agent” |
| Guilt | “I am responsible for the AIDS epidemic” |
| Religious | “I am the reincarnation of Solomon” |
| Cotard delusion | “I am dead/do not exist,” or “My body is decaying” |
| Lycanthropy | “I am/ have transformed into an animal” |
Belief-positive models have their origins in cognitive neuropsychiatry, where the study of psychopathology is used to infer normal function, and normal models provide the framework for understanding the phenomena. Therefore, many belief-positive models explain delusions within wider models of the neuropsychology of normal belief formation. Traditionally, research in this area has focused on brain-injury related ‘monothematic’ delusions that typically have ‘bizarre’ (i.e. impossible) content.

Perhaps the most detailed belief formation model is provided by Langdon and Coltheart [22] who make a distinction between factors that explain the content of delusions and those that explain their presence (Box 3). These authors suggest that perceptual distortion (the ‘first factor’) is a necessary component, at least for ‘bizarre’ delusions, but argue that cognitive biases, although potentially contributing to content, are neither necessary nor sufficient to account for the presence of delusions. They argue, however, that such biases could explain the generation of unusual hypotheses (‘magical thinking’), but that an additional deficit (the ‘second factor’) is needed to explain why such hypotheses might result in full-blown delusions. This, these authors suggest, is a deficit in the rational evaluation of hypotheses as candidate explanations. In a subsequent ‘two-factor’ analysis, Davies et al. [23] argue that pathological beliefs could arise via two possible routes: either by producing a pathological explanation for an anomalous experience (potentially affected by both pathological and non-pathological cognitive biases), or simply believing a pathological perception, with the ‘second factor’ needed to account for why such explanations are not rejected in light of everything else the patient knows (see Figure 1).

The necessity of perceptual aberration in delusion formation remains an ongoing matter of debate [1]. There is now growing evidence that levels of anomalous perceptual experiences are distributed throughout the normal population, and that a significant minority of delusional patients do not report such phenomena [24], suggesting that this can be a contributory factor, but not a necessary condition.

McKay et al. [25] further claim that the ‘first factor’, previously considered to be largely perceptual, should be reconsidered to include ‘defences, desires and motivations’ – also thought to be important in delusions following neurological injury [26]. The second factor might then involve a deficit or bias that gives undue credence to such evidence. Furthermore, the second factor could be explained by damage to proposed right-hemisphere processes that re-organize beliefs in response to new information [27]. One problem with this hemispheric explanation, however, is that, although it is consistent with some of the findings from pathological beliefs arising from brain injury and dementia [28,29,30], the evidence for such a clear functional asymmetry in idiopathic psychosis has yet to be found [31].

Thus far, the models discussed are either explicitly or implicitly based on ‘representationalist’ views of belief, that equate beliefs with explicitly-held propositions in (presumably semantic) memory. In contrast, Bayne and Pacherie [32] argue for a ‘liberal dispositional’ model of belief [33]. Here, beliefs are seen not only as propositions, but dispositions to respond and experience internal mental events or states, including conscious experiences and emotions. These are considered to be mediated by pre-existing schemas and biases which can include, but are not limited to, propositional or semantic content. Although

Box 2. What is cognitive neuropsychiatry?

Cognitive neuropsychiatry brings together cognitive and neurobiological research to improve understanding of mental disorder. By drawing inspiration from cognitive neuroscience and cognitive neuropsychology, it aims to explain psychiatric and neuropsychiatric symptoms within normal models of cognitive function and uses the study of psychopathology to inform existing models. Crucially, the discipline also attempts to link such functional explanations to relevant brain structures and their pathology [68]. The focus is typically on symptoms rather than diagnoses, as the study of symptoms is thought more likely to elucidate functional relationships than the broad diagnostic approach [69].
 speculative, we feel Bayne and Pacherie’s approach is more closely in tune with current and substantial philosophical objections to the representationalist view of belief [34]. Moreover, it links well with both cognitive theories of schema and memory, and the ideas underlying cognitive behavioural therapy, whereby beliefs can be changed through tackling maladaptive appraisals and pre-existing assumptions [35].

**Belief-negative models of delusions**

Belief-negative models are largely focused on the more common, idiopathic or ‘functional’ psychoses, rather than delusions following brain injury. Typically, they attempt to explain the pathological process only and do not make explicit links to theories of normal belief formation.

Blakemore and colleagues [36,37] have outlined a model of delusions of control, although it might be better framed as a model of the anomalous experience. Despite clear predictions and growing empirical support [38,39], this account fails to explain why such experiences result in a delusional belief – simply stating that ‘In parallel, the patient’s belief system is faulty so that he interprets this abnormal sensation in an irrational way’ [37].

The Bentall et al. ‘attribution/self-representation’ model [9] focuses on persecutory delusions and argues that they result from trying to keep ‘actual-self’/‘ideal-self’
discrepancies to a minimum by making excessive external-personal attributions (‘blaming others’). The authors maintain that such attributions defend against latent negative beliefs about the self which can lower implicit self-esteem, even if explicitly reported self-esteem is normal or high. Reliable evidence for a clear difference between implicit and explicit self-esteem has been elusive, however. Recent studies using the emotional Stroop task [40], the Pragmatic Interference Test [41,42] and the Implicit Attribution Test [43] have, at best, only partly supported this prediction. Bentall et al. [9] have suggested that self-esteem is not stable in all patients. Although there is evidence that self-esteem is dynamically affected by attributions in both healthy and delusional participants [44,45], this dynamic aspect of the model has not yet been comprehensively tested. Recently, aspects of Bentall and colleagues’ model have been tested and explored using neuroimaging in an attempt to provide useful parameters and limits for this approach [46,47].

Freeman and Garety [48] also focus on persecutory delusions, which they explain in what amounts to two separate, but overlapping models (delusion formation and delusion maintenance). The formation model (see Figure 2) is based on a ‘stress-vulnerability’ approach and argues that persecutory delusions arise from an anomalous experience (potentially precipitated by a biological or psychological stressor), especially when interpreted by someone with extreme cognitive biases or maladaptive pre-existing beliefs. The authors argue that delusions are maintained by cognitive dissonance-style ‘relief’ at gaining an explanation for an unusual experience [49], biases for collecting confirmatory evidence and avoiding or discounting evidence to the contrary [50] and ongoing affective disturbance. In contrast to the Bentall et al. [9] model, where delusions are thought to arise as a ‘defence’ or attempt to keep negative affect from consciousness, Freeman and Garety argue for a ‘direct’ role of emotion, where the delusion shares much of its content and structure with the proposed form of emotional disturbance. This is based on evidence for thematic similarities between common delusions and emotions [48], and studies showing high levels of paranoia-related anxiety in clinical [7,51] and non-clinical populations [52].

Morrison et al. [35] consider delusions to arise from ‘intrusions into awareness’, in the form of thoughts, hallucinations or bodily sensations, that are interpreted in culturally unacceptable ways (e.g. a hallucinated voice interpreted as evidence for being subject to microwave mind-control technology). Such interpretations are thought to be generated by faulty self and social knowledge, and maintained by mood, physiological changes (such as poor sleep) and maladaptive cognitive and behavioural responses. The models of Freeman and Garety [48] and Morrison et al. [35] are often pragmatically focused on a formulation of delusions used in cognitive behavioural therapy. Perhaps, because of this, there is little overt integration with neuropsychological evidence, unlike with the other models described in this section.

**Continuum approaches to delusions**

The continuum approach to psychosis draws inspiration from epidemiological approaches to look for predictors and correlates of delusional ideation and/or frank delusions within the wider population. Although not a model of delusion formation per se (and hence compatible with both

---

**Figure 2.** Persecutory delusion formation model, adapted from Freeman and Garety [48].

www.sciencedirect.com
beliefs, and how intrusive trauma-related memories can maintain and exacerbate psychotic states. Longitudinal studies have produced data suggesting that more general states of negative affect, when associated with a hallucination or intrusive perception, vastly increase the later risk of delusion. Interestingly, this seems a particular risk for those with a hyperreactive dopamine system.

Defining new criteria
Currently, delusions are largely treated as core psychopathological symptoms (although Gileen and David [1] have sounded a dissenting voice) suggesting that delusions are more like multifaceted syndromes rather than indivisible symptoms (see also Box 4). Similarly, we feel that, because both belief and delusion are understood and expressed in such diverse ways (both scientifically and when used in everyday clinical settings), they are unlikely to yield to accounts based on explanations of a unitary phenomenon.

One further complicating issue in dealing with the theoretical accounts is that different approaches have not always based their conclusions on the same patient populations. For example, belief-positive theories are largely based on the study of impossible, monothematic, brain-injury-related delusions; whereas belief-negative theories are largely drawn from the study of plausible but persecutory delusions, associated with idiopathic psychosis. Consequently, belief-positive theories are often more concerned with explaining how an impossible belief comes to be accepted, whereas belief-negative theories (the Blakemore model excepted) are concerned with how a belief becomes associated with negative affect.

One way of potentially integrating these models, together with the wider array of empirical findings in the delusions literature, would be to arrive at an agreed model (or related models) of normal belief formation. This has the advantage of allowing a classification of delusional pathology based on cognitive similarities rather than diagnostic tradition.

Importantly, however, a theory of delusions must be more than simply the articulation of an isolated impairment in a ‘cold’ information processing model. It is striking that none of the criteria outlined in Box 3, nor the DSM definition, nor many of the cognitive models (particularly in the belief-positive tradition) includes any explanation of delusion-related distress or pragmatic disability that often characterizes and provides for clinical involvement. It is tempting to think that those models that best address this issue (most notably Bentall et al. [9] and Freeman and Garety [48]) do so by concentrating solely on persecutory delusions, which, by their very nature, involve intense negative emotion. As a result, these models, however successful in their own domains, might not generalize well to other types of delusion (although with the caveats outlined above with regard to unitary theories, this might not necessarily be a disadvantage).

Concluding remarks
This review highlights a striking neglected issue: current models that make no distinction between beliefs resulting in pragmatic distress and impairment, and those that are simply impossible or anomalous, are unable to distinguish between beliefs that are clinically significant, and those that may be unusual or ‘magical’, but otherwise benign. Hence, an adequate theory should also explain the ‘pragmatic pathology’ of delusions. Cognitive theories that achieve this will be able both to address the phenomenology of anomalous belief and to

Figure 3. Distribution of delusional ideation scores in general population and patients, adapted from Peters et al. [56].

Box 4. Questions for future research

- Do delusions in the context of psychiatric illness and those in the context of brain-injury arise from pathologies to the same cognitive mechanisms?
- What is the exact role of emotion in the formation and maintenance of delusions?
- Are delusions best thought of as indivisible symptoms or multifaceted syndromes?
- What mediates the transition from simply ‘magical’ or anomalous beliefs to those that cause distress and impairment?
- Can we adequately define a delusion in terms of pathology to existing belief systems or should the term be more theoretically neutral, assuming that several different pathologies to normal cognitive processes could be involved?
inform the treatment of delusion-related distress and disability.

References
64 Bell, V. et al. (2003) Beliefs about delusions. *The Psychologist* 16, 418–423