AVANT, Vol. V, No. 3/2014
ISSN: 2082-6710 avant.edu.pl/en
10.12849/50302014.0112.0002

From the internal lexicon to delusional belief[5]

Max Coltheart
Department of Cognitive Science
Macquarie University
max.coltheart[]mq.edu.au

Received February 2014; accepted May 2014; published winter 2014/2015.

Editorial abstract

In this overview, Author presents the development of his approach—the two-factor account of delusions—drawing attention to the neuropsychological research on delusions (the role of brain damage in the formation of delusions), as well as to the differences between explaining monothematic and polythematic delusions (this differentiation is not analyzed in detail in the present volume). He also sketches the most promising issues in the current research on delusions.

Keywords: delusions; belief formation; brain damage; cognition; neuropsychology.

After "Access to the internal lexicon"

What has changed in the Author's approach in cognitive neuropsychology and cognitive neuropsychiatry since "Access to the internal lexicon" and "Lexical access in simple reading tasks"?[6]

In the 1970s I was a cognitive psychologist doing theoretical work on the cognitive mechanisms of skilled reading, and the two papers mentioned above are early examples of my work on this topic; they have had some influence, having been cited 1580 and 1333 times to date (Google Scholar, 29 March 2015). However, in the late 1970s I became convinced, mostly because of the work of John Marshall and Tim Shallice, that progress in understanding the cognitive mechanisms of skilled reading could best be made by detailed study

[5] "Delusional belief"—see Coltheart, M., Langdon, R. & McKay, R.T. 2011. Delusional belief. *Annual Review of Psychology*, 62: 271-298.

[6] Coltheart, M., Davelaar, E., Jonasson, J.T., and Besner, D. 1977. Access to the internal lexicon. In Dornic, S., ed. *Attention and Performance VI*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. Coltheart, M. 1978. Lexical access in simple reading tasks. In Underwood, G., ed. *Strategies of Information Processing*. London: Academic Press.

of the variety of ways in which reading breaks down in people who were formerly skilled readers but whose reading was then impaired by brain damage—people suffering from acquired dyslexia. So I began to do research with such people.

Studying people with specific impairments of cognition in an effort to learn more about the normal processes of cognition is a branch of cognitive psychology known as cognitive neuropsychology. So I, and others like Marshall and Shallice, was working on the cognitive neuropsychology of reading, using the characteristic approach of cognitive neuropsychology, which is to study individual patients in great depth (that is, to do single-case studies, not group studies) and to focus on relating data from such patients to some explicit information-processing model of the relevant domain of cognition. The joint aims of such work are to use the patient data to test predictions of the model and to use the model for seeking to understand how the patient's specific cognitive symptoms arise.

It was clear by the early 1980s that this was proving to be a very successful way of studying the cognitive psychology of reading. So cognitive psychologists interested in domains of cognition other than reading began to explore the possibility that they could learn valuable things about how processing normally proceeded in whatever domain of cognition they were interested in by studying people in whom brain damage had produced impairments in that particular domain. In this way cognitive neuropsychology, which had begun by studying the domain of reading, spread to many other cognitive domains such as spelling, spoken language comprehension and production, visual object recognition, face processing, short-term memory and many others. The insights that were achieved into normal processes of cognition by studying individuals whose cognitive processes were damaged were substantial enough that the subject of cognitive neuropsychology needed its own journal; this journal, *Cognitive Neuropsychology*, was founded in 2004.

In this early work, although a variety of different domains of cognitive (such as those listed above) were studied using the basic approach of cognitive neuropsychology, all of these domains were what one might call low-level cognitive domains. This was because in such domains cognitive psychology already offered explicit information-processing models of normal cognitive processes that could be directly tested by, and could offer plausible explanations for, the symptoms exhibited by patients with cognitive impairments consequent upon brain damage. But there are higher-level domains of cognition such as theory of mind, reasoning, belief formation and sense of agency that can also be impaired in clinical patients. Might it be possible to study even these by the methods of cognitive neuropsychology?

One major obstacle to doing this was that there were few models of normal processing in these higher cognitive domains, and even those models that did exist were far vaguer and less explicit than the models of reading, face recognition, short-term memory et cetera that cognitive psychology had been developing for decades before the rise of cognitive neuropsychology. But this did not deter John Marshall, Hadyn Ellis and Andy Young from beginning, in the early 1990s, the cognitive-neuropsychological investigation of patients with such higher-level disorders, in the hope that this might allow the development of better models of higher-order domains of cognition. Patients with disorders of these domains are typically psychiatric patients. That is why the branch of cognitive neuropsychology that applies to these higher-order domains came to be called cognitive neuropsychiatry. Soon it too needed its own journal; and *Cognitive Neuropsychiatry* was founded in 1996.

Just as acquired dyslexia was the most commonly studied form of cognitive impairment in the early days of cognitive neuropsychology, so delusional belief was the most commonly studied form of cognitive impairment in the early days of cognitive neuropsychiatry. And just as the cognitive-neuropsychological approach then rapidly spread from reading to other low-level domains of cognition, so the cognitive-neuropsychiatric approach rapidly spread from delusional belief to other high-level domains of cognition: for example, in 2014 the first two issues of the journal *Cognitive Neuropsychiatry* included papers on borderline personality disorder, anorexia nervosa, abnormalities of theory of mind, confabulation, Asperger's syndrome and autism, as well of course as papers on delusional belief.

I became a cognitive neuropsychiatrist by chance. My current long-term colleague and collaborator Robyn Langdon, having been trained as a cognitive neuropsychologist, was interested in using the cognitive-neuropsychological approach to study impairments of theory of mind in patients with schizophrenia. This was the subject of her PhD research in the 1990s. Her intended supervisor, an expert on schizophrenia, took a position in another university. So Robyn asked me to be her supervisor. When I protested that I knew nothing about schizophrenia, she said that this did not matter since she did, and could teach me what I needed to know. That is how I fell into the field of cognitive neuropsychiatry.

I fell into work on delusional belief equally fortuitously. Also in the 1990s, a clinical neuropsychologist whom I knew, Nora Breen, had come across in her practice two men with the same, remarkable, delusion. Both men (who were in the early stages of dementia, though this was not known at the time) firmly believed that when they were looking into a mirror, they were not seeing themselves, but some stranger who looked like them. Nora decided that she wanted to do doctoral research on this form of delusion, known as mirrored-self misidentification, and that I should be her supervisor. Despite never hav-

ing studied delusional belief before, I found the videotapes of these patients that she showed me so fascinating that I agreed. I have been working on delusional belief ever since.

The sources and importance of two-factor theory of delusions. Problems with monothematic delusions

William James was, I believe, the first person to express the insight that is the foundation for our two-factor theory of delusional belief (see Coltheart et al., 2011, for our most recent exposition of this theory). In 1890, James wrote "The delusions of the insane are apt to affect certain typical forms, very difficult to explain. But in many cases they are certainly theories which the patients invent to account for their bodily sensations" (James 1890/1950, chap. XIX). More recently, exactly the same idea was proposed by Brendan Maher. His view was that, when a patient is suffering from some sort of perceptual or affective impairment which leads the patient to have certain strange experiences

> strange [experiences] felt to be significant demand explanation. It is the core of the present hypothesis that the explanations (i.e. the delusions) of the patient are derived by cognitive activity that is essentially indistinguishable from that employed by non-patients, by scientists, and by people generally (Maher 1974: 103).

The James-Maher theory of delusional belief is a one-factor theory. All that is required for the genesis of delusion is the presence of a perceptual or affective impairment that leads the patient to have certain strange experiences. Normal—intact—processes of reasoning do the rest of the job in generating delusional belief.

In order to flesh out this theory, more needed to be said about the nature of the abnormal experiences that delusional patients have. What specific experiences lead to what specific delusional beliefs, and why? What particular kind of abnormal experience could cause a patient to have a delusional belief with content X rather than with content Y? Thanks to the seminal work of Ellis et al. (1997), we can now see how to answer such questions.

The particular delusion studied by Ellis and colleagues was the Capgras delusion. This is the belief that someone emotionally close to the deluded person—typically a spouse or other family member—has been replaced by a total stranger who looks just like the "missing" family member. To discover what is happening here, Ellis and colleagues exploited the fact that when anyone is looking at another person's face, that face produces a much larger response of the viewer's autonomic nervous system if the viewed face is familiar than when it is not. Ellis and colleagues reported that this was true even for non-delusional psychiatric patients. But it was not true for sufferers from Capgras delusion, who showed very little autonomic responsivity at all to faces, and no

greater responsivity when the faces were familiar to them than when the faces were entirely familiar.

This absence of autonomic response to a face that looks just like the face of someone who ought to generate a large autonomic response is to do with what William James would have thought of as a bodily sensation (though here what's unusual is the *absence* of an expected bodily sensation). So here he (and Maher) would expect the patient to invent a theory to account for this (lack of) specific bodily sensation. What could explain why a person you are seeing, who looks just like your wife, nevertheless does not generate in you a large autonomic response? Strangers do not generate large autonomic responses. So the theory that the person you are looking at is a stranger would account for your unusual experience. Here there is a highly plausible link between the specific kind of abnormal experience the patient is having and the specific content of the patient's delusional belief.

According to the James-Maher one-factor theory of delusional belief, this lack of autonomic responsivity to familiar faces is the only abnormality needed to cause the Capgras delusion. Normal processes of reasoning, used to try to explain this lack of response, do the rest of the job in generating the delusional Capgras belief. From this it follows that all patients who show no autonomic responsivity should also show Capgras delusion.

But this turns out not to be so. Tranel et al. (1995) studied five patients with damage to the ventromedial cortex of the frontal lobes (a region also know as orbitofrontal cortex). In all of these patients, their brain damage had resulted in the failure of their autonomic nervous systems to respond to familiar faces. However, they were not delusional; they were able to correctly recognize people highly familiar to them such as family members.

What are we to make of this? It might be concluded that the failure of autonomic responsivity to familiar faces seen in people with Capgras delusion is just a coincidence, something that has no causal role in the occurrence of the delusion. But this seems implausible because of the rather natural connection between the content of the Capgras delusion and the patient's lack of the autonomic response to seeing the face of the spouse, a response that the patient would have expected. So instead my collaborators and I have argued that in people who lack such autonomic responsivity, a second factor must be present if Capgras delusion is to occur.

We think of this second factor as an abnormality of a belief evaluation system that is normally used to evaluate candidates for belief and to decide whether to dismiss these or adopt them as new beliefs. This abnormality is not present in the Tranel patients, so they are able correctly to reject the candidate belief "That is not my spouse" that is suggested to them by the failure of any autonomic response the sight of the spouse's face. In Capgras patients the effect of

the second factor (an impairment of the belief evaluation system) is to prevent this candidate belief from being rejected, which permits it instead to be adopted as a new belief.

This two-factor theory of delusional belief is not just a theory about Capgras delusion; we have applied this approach to the explanation of a variety of other delusions too, such as the mirrored-self misidentification described above, the Cotard delusion (the belief that one is dead), the Fregoli delusion (the belief that people one knows are following one around, but are not recognizable because they are in disguise) and various other delusions: see Coltheart et al. (2011) for an account of our explanations of a variety of delusional conditions.

The general form of the two-factor theory of delusional belief is as follows. Just as James and Maher proposed, there is an abnormality that leads the patient to have an unexpected response to the environment that the patient seeks to explain. The explanation that the patient formulates is a candidate belief about the environment. The specific nature of this first factor is responsible for the content of this candidate belief. Since the way in which the various types of delusional belief differ is with respect to the content of the belief, the nature of the first factor differs from delusion to delusion.

By definition, a belief only counts as delusional if there is strong evidence against it, evidence that should be accepted by the deluded person but is not. It is the second factor—an impairment of the belief evaluation process—which is responsible for the candidate belief being adopted as a (delusional) belief rather than being rejected. The nature of this second factor is, we argue, common to all forms of delusional condition.

As stated above, there are patients in whom the first factor of Capgras delusion is present but who are not delusional; that is why we hold that a second factor is necessary for a delusion to arise (and why the two factors are of equal importance). A general task for the two-factor theory of delusion, then, is to identify for each delusion a plausible first factor (an impairment that has a plausible connection to the characteristic content of the delusion) and then for each proposed first factor to show that the literature contains reports of patients who suffer from the particular proposed impairment but who are not delusional. Any successful exercise of this kind justifies the claim, for the particular delusion concerned, that two factors are needed to explain the occurrence of the delusion.

Coltheart et al. (2011) have proposed first factors for a number of different delusions and provided evidence that patients have been reported who exhibited one or other of these first factors without being delusional. These demonstrations constitute strong evidence in favor of the two-factor theory of delusional belief and are very difficult to reconcile with any one-factor theory of delusion.

That is not to claim that the two-factor theory of delusional belief is not without problem. These problems are discussed by Coltheart (2007) and Coltheart et al. (2011). They include:

(a) What is the intended scope of the two-factor account of delusional belief? All of the forms of delusional belief I discuss in this paper are examples of *monothematic* delusional conditions: the patient has only a single delusional belief, or a small set of beliefs which are all about a single theme. And it is also clear that these kinds of delusion have neuropsychological bases. But some patients exhibit a *polythematic* delusional condition: they have many different and unrelated delusional beliefs. And with such patients there may be no clear evidence of neuropsychological dysfunction. Is the two-factor theory also meant to account for this kind of delusional condition?

(b) Even when a delusional condition is monothematic and clearly consequent upon neuropsychological function, the delusional belief can come and go. Sometimes the patient expresses the belief, but at other times rejects it. How can this be so if the delusion is caused by neuropsychological damage?

(c) If patients with monothematic delusions have impaired functioning of their belief evaluation systems, why don't they have a whole variety of delusional beliefs, rather than only one?

Ways in which the two-factor theory can answer these questions are discussed in detail by Coltheart (2007) and Coltheart et al. (2011).

From delusions to belief formation. The role of research on delusions in understanding non-pathological beliefs

By definition, if you are doing cognitive-neuropsychiatric work on how delusional beliefs are formed, you need a theory, no matter how sketchy, of the normal processes of belief formation. When Robyn Langdon, Nora Breen and I, together with the philosopher Martin Davies, began to develop the two-factor theory of delusions in the late 1990s (see e.g. Davies et al. 2002), we intended this theory to say something, even if not very much, about the normal processes of belief formation.

Our general idea is that what generates candidates for belief are prediction errors. By this I mean that as people go about their everyday lives, they continually use their systems of belief to make predictions about what will happen to them next. In people with intact belief systems these predictions will almost always be correct, and whenever this happens this will go unnoticed and will have no consequences for the belief system.

Occasionally, however, such predictions fail; something happens which is unexpected. Any such error of prediction is a sign that the belief system needs to be revised, since there is some feature of the world, the feature which caused

the unexpected event, that is not correctly represented in the belief system. The prediction error triggers a search for hypotheses about what this feature might be: what could the world be like such that, if it were like that, the unexpected event would actually be expected? The process by which such hypotheses are inferred is not deductive or inductive inference, but abductive inference, sometimes referred to as "inference to the best explanation" (Lipton 1991). As a rule abductive inference will yield many possible explanations for the unexpected event, and so there needs to be a belief evaluation process here—that is, a choice must be made, perhaps along Bayesian lines, concerning which hypothesis to accept. Accepting a hypothesis means adopting a new belief.

Here, then we have the following picture of the normal processes of belief generation, evaluation and adoption. Prediction error leads to the generation of a number of candidate beliefs about the world, each being a possible explanation of the unexpected event. A belief evaluation process is then applied to this set of candidate beliefs, and the winning candidate is adopted as a new belief.

All of this also happens in cases of pathological belief i.e. delusion. In the case of Capgras delusion, the patient sees a person with his wife's face and predicts that there will be an autonomic response, but there is no such response: there's the prediction error, caused by the first factor. One possible explanation for this error that could be yielded by abductive inference is that this is not the wife. If this were true, then the unexpected event would be expected, so this is an abductively adequate explanation. The belief evaluation process should reject this candidate belief because there is so much evidence against it; but because of the presence of the second factor (some form of defect in the belief evaluation process), this belief is instead adopted i.e. the person becomes delusional.

Thus the two-factor explanation for delusional belief does have embodied in it a clear, if rudimentary, theory about the non-pathological processes of normal belief formation.

New routes, emerging fields… On expectations and hopes for the future research on delusions

Several important routes for future research are evident. I will mention just two.

First is further elaboration of the theory of normal belief formation that is suggested by the two-factor theory of delusional belief. The most promising avenue here is to seek to link up the theory of delusional belief with the dual-process framework offered by Kahneman (2011). This framework posits two systems of thought. System 1 is fast, automatic and unconscious. System 2 is slow, effortful, systematic and conscious. Most of our thought is System 1 thought. One thing that brings System 2 into play is System 1 being confronted with a problem that it finds difficult or impossible to resolve.

Kahneman amasses a great deal of evidence from studies of thinking and reasoning in intact people that supports this dual-process framework. It seems likely that the two-factor theory of delusional belief can also be shown to be consistent with this framework. System 1 would seem to correspond to the processes we use—constantly, rapidly and automatically—to predict what will happen to us next. A prediction error (caused by the first factor) corresponds to a failure of System 1 processes, which leads to an invocation of System 2. System 2 in the intact mind is responsible for belief evaluation and adoption processes, and Factor 2 is an abnormality of these processes. In expositions of the two-factor theory, the nature of this abnormality is only described in the most general of terms. Future work on the two-factor theory will have to describe Factor 2 in much more specific terms, and attempting to link the two-factor theory to Kahneman's dual-process framework might achieve this.

A second important avenue for future research here concerns the neuropsychology of delusional belief. Can the abnormalities referred to as Factor 1 and Factor 2 in the two-factor theory be identified as specific neuropsychological impairments? To show that this is so, then two things would need to be demonstrated in patients with delusional beliefs.

First, for each different delusion a neuropsychological impairment would have to be demonstrated that would produce an effect corresponding to that delusion's Factor 1 (e.g. in cases of Capgras delusion, the neuropsychological impairment would have to have the effect of preventing familiar faces from evoking autonomic responses). Since the two-factor theory asserts that each type of delusion has its own type of Factor 1, what the neuropsychology would have to demonstrate is a different form of neuropsychological impairment for each different type of delusion. Quite a lot of progress has been made in identifying what form of neuropsychological impairment corresponds to Factor 1 in different forms of delusion (Coltheart 2007, 2010).

Second, a neuropsychological impairment would have to be demonstrated that would produce an effect corresponding to Factor 2. Since the two-factor theory asserts that every type of delusion has the same Factor 2, this makes the bold prediction that there will be some form of damage to the brain that will be common to all forms of delusion, regardless of the content of that delusion. This would be damage to a region of the brain that is the neural substrate for the cognitive system involved in belief evaluation. What might this region be? Coltheart (2007) reviewed some scraps of literature that suggest that this region might be located in the right dorsolateral prefrontal cortex (rDLPFC). Since then further evidence has emerged suggesting that this brain region is indeed damaged in cases of delusion (e.g. Villarejo et al. 2011).

This suggests a specific focus for future neuroimaging work with delusional patients: is it really true that rDLPFC is consistently impaired in such patients? If this does turn out to be so, what about the suggestion that rDLPFC is the neural substrate for the cognitive system involved in belief evaluation? That can be directly investigated by neuroimaging studies of cognitively intact people carrying out tasks that require belief evaluation. Does one see evidence of rDLPFC activation here?

References

Coltheart, M. 2007. The 33rd Bartlett Lecture: Cognitive neuropsychiatry and delusional belief. *Quarterly Journal of Experimental Psychology, 60:* 1041-1062.

Coltheart, M. 2010. The neuropsychology of delusions. *Annals of the New York Academy of Sciences, 1191*: 16-26.

Coltheart, M., Langdon, R. & McKay, R.T. 2011. Delusional belief. *Annual Review of Psychology, 62,* 271-298.

Davies, M., Coltheart, M., Langdon, R. & Breen, N. 2002. Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry & Psychology*, 8: 133-158.

Ellis, H.D., Young, A.W., Quayle, A.H., & de Pauw, K.W. 1997. Reduced autonomic response to faces in Capgras' delusion. *Proceedings of the Royal Society of London, Series B, 264*: 1085-1092.

James, W. 1950. *Principles of psychology* Vol. 2. New York: Henry Holt & Co. Original work published in 1890.

Kahneman, D. 2011 *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Lipton, P. 1991. *Inference to the Best Explanation*. London: Routledge.

Maher, B.A. 1974. Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30: 98–113.

Tranel, D.H., Damasio H., & Damasio, A.R. 1995. Double dissociation between overt and covert face recognition. *Journal of Cognitive Neuroscience, 7:* 425–32.

Villarejo A, Martin VP, Moreno-Ramos T, *et al.* 2011 Mirrored-self misidentification in a patient without dementia: Evidence for right hemisphere and bifrontal damage. *Neurocase* 17: 276-84.