

## **Delusional Beliefs, Two-Factor Theories, and Bizarreness**

### **Abstract**

In order to explain delusional beliefs, one must first consider what factors should be included in a theory of delusion. Unlike a one-factor theory, a two-factor theory of delusion argues that not only anomalous experience (the first factor) but also an impairment of the belief-evaluation system (the second factor) is required. Recently, two-factor theorists have adopted various Bayesian approaches in order to give a more accurate description of delusion formation. By reviewing the progression from a one-factor theory to a two-factor theory, I argue that in light of the second factor's requirements, different proposed impairments can be unified within a consistent belief-evaluation system. Under this interpretation of the second factor, I further argue that the role of a mechanism responsible for detecting bizarreness is wrongly neglected. I conclude that the second factor is a compound system which consists of differing functional parts, one of which functions to detect bizarreness in different stages of delusion; moreover, I hold that the impairment can be one or several of these functional parts.

**Keywords** delusion, two-factor theories, Bayesian theory, bizarreness

### **1. Introduction**

There are “two ways of doing philosophical psychopathology, as a foundational contribution to a scientific inquiry or as a piece of philosophy of mind.” (Murphy 2015) In regard to delusion, a two-factor theory mainly follows the former approach, focusing on such philosophical questions as how many factors should be considered in a good theory of delusion, and why a particular factor should be included.

Generally speaking, theorists who favor the one-factor theory, such as Maher 1999, argue that anomalous experience alone is sufficient to explain delusion. However, a two-factor theorist will argue that an impaired belief-evaluation system, as a second factor, is also needed. Here, the main dispute is over whether a second factor is necessary.

Among two-factor theorists, no unanimous agreement has been reached concerning the nature of the second factor. There are three prevailing proposals that adopt a Bayesian approach, represented by Coltheart et al. 2010, McKay 2012, and Davies and Egan 2013. Coltheart et al. and McKay disagree on which part of the mechanism is impaired. Meanwhile, Davies and Egan argue that both Coltheart et al. and McKay presume an ideal belief system, and that on the contrary, a normal belief system is compartmentalized; in their view, this compartmentalization of the belief system could also be the second factor.

Within a two-factor theory framework, this paper argues that if one takes the belief-evaluation system as a compound mechanism, the proposed impairments could be taken as different damaged functional parts within it. Furthermore, this belief-evaluation system still lacks a function responsible for detecting the bizarreness of delusional beliefs. This absent function's role could be *partly* fulfilled by the monitoring system proposed by Turner and Coltheart 2010.

In the following sections, this paper will first review the progression from a one-factor theory to a two-factor theory, showing why a second factor is needed in an explanatory theory of delusion. Section 3 analyses in detail the requirements of the second factor. Section 4 unifies the different candidates for the second factor within a single belief-evaluation system. This paper's most philosophical inquiry comes in Section 5, which focuses on the questions of why a function responsible for detecting the bizarreness of delusional beliefs is required, what sorts of questions it should deal with, and how it could be consistent with the other mechanisms of a belief-evaluation system.

## **2. Why Is a Second Factor Needed in an Explanatory Theory of Delusion?**

A controversial but enlightening definition of delusion by the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* is: "A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary" (DSM-IV-TR, 2000, 821). Typical examples of delusions include the Capgras delusion—"This [the subject's wife] is not my wife. My wife has been replaced by an impostor" (Capgras and Reboul-Lachaux 1923; Edelstyn and Oyebode 1999); somatoparaphrenia—"This [the subject's left arm] is not my arm" (Bottini *et al.* 2002; Halligan *et al.* 1995); and the Cotard delusion—"I am dead" (Cotard 1882; Young and Leafhead 1996).

Because of the variety and the undetermined nature of delusions, such a definition naturally causes much controversy. However, some general key features of these delusions can still be picked out, including incomprehensibility, incorrigibility, and subjective certainty (Langdon and Bayne 2010). There are several theoretical approaches to understanding the nature of delusions (Bortolotti 2013). The cognitive approach is one of the most influential, applying the methods of cognitive neuropsychology to psychiatric disorders and, in the literature, focusing on circumscribed monothematic delusions of neuropsychological origin rather than on elaborated polythematic ones.

Maher 1999 made an initial contribution to the neuropsychological understanding of delusion by arguing that delusions are hypotheses generated by normal reasoning processes to explain anomalous experience:

What this implies is that in deluded individuals the locus of the pathology is in the neuropsychology of experience, not in the neuropsychology of deductive or inferential

reasoning. (Maher 1999, 551)

It should be noted that in Maher's theory, there is a clear distinction between experience and reasoning, and belief is generated as a hypothesis by a reasoning system. The only resources available to the reasoning system are anomalous experience and normal background knowledge. For Maher, anomalous experience itself is sufficient for the individual to produce a delusion. Maher's theory is a standard one-factor theory, meaning that there is only one kind of neuropsychological deficit during the formation of the delusion.

However, there are patients who possess the proposed neuropsychological deficit but who do not have the corresponding delusion (Coltheart 2007, 2010; Coltheart, Langdon and McKay 2011). Simply put, this is the famous dissociation argument for a second factor. We may use the intensively discussed Capgras delusion as an example to illustrate this.

In a seminal 1990 paper, Ellis and Young proposed a neuropsychological deficit for the Capgras delusion. Their main idea is that in normal face processing, both face recognition units (FRUs) and an autonomic nervous system, which could be measured by skin conductance responses, will be active (Bruce and Young 1986; Tranel and Damasio 1985, 1988). Since people with the Capgras delusion could recognize the face of the so-called impostor, Ellis and Young believed that the primary face-recognition system is intact, and that the deficit should be understood as damage to the connection between the FRUs and the autonomic nervous system. This proposal predicts that the skin conductance responses of patients with the Capgras delusion would not discriminate between familiar faces and unfamiliar faces. This prediction has been confirmed by four different studies (Brighetti *et al.* 2007; Ellis *et al.* 1997, 2000; Hirstein and Ramachandran 1997; Ellis 1998). Therefore, it is highly plausible that the neuropsychological deficit of the Capgras delusion rests upon the disconnection between the primary face-recognition system and the autonomic nervous system.

However, there are patients who do not have the Capgras delusion even though their skin conductance responses do not discriminate between familiar and unfamiliar faces (Tranel *et al.* 1995). While we may admit the deficit proposed by Ellis and Young is one necessary factor for the formation of the Capgras delusion, the counter-evidence still proves that there are missing factors to be found.

Combining Maher's framework with Ellis and Young's proposal and the counter-evidence, the conclusion could be drawn that either there are more factors which are responsible for the anomalous experience, or that the reasoning system may also be impaired. Therefore, the missing factors could either be categorically similar to the disconnection between the primary face-recognition system and the autonomic nervous system, or categorically different from it.

### **3. What Is Required by the Second Factor?**

In order to deal with the dissociation of the neuropsychological deficits proposed and the corresponding delusion, a two-factor theory has been proposed (Davies and Coltheart 2000; Davies *et al.* 2001; Langdon and Coltheart 2000). The standard view of the two-factor theory is that, in order to explain the delusion, two questions must be answered:

The first question is always: where did the delusion come from?—that is, what is responsible for the content of the delusional belief?

The second question is always: why does the patient do not reject the belief? (Coltheart 2007, 1044)

In a more developed version of the two-factor theory, the second question is divided into two questions, namely, *why does the patient adopt the hypothesis in the first place?* and *why does the patient maintain such a bizarre hypothesis?* (Davies *et al.* 2001; Aimola-Davies and Davies

2009; Davies and Egan 2013).

Combining these three questions with Maher's framework, we can conclude that, in order to explain delusion, three stages must be explained by two elements:

Element 1: Anomalous experience (or abnormal data in Coltheart's terminology, which leads to anomalous experience).

Element 2: A reasoning system (or belief evaluation system in Coltheart's terminology).

Stage 1: The arising of the hypothesis

Stage 2: The adoption of the hypothesis

Stage 3: The persistence of the hypothesis

Likening the belief-formation process to machine computation, experience is like the input of a computer, the reasoning system is like the algorithm that it runs, and the belief is like its output.

Maher's theory could be interpreted as suggesting that only the input (the anomalous experience) is impaired, and that the algorithm (the reasoning system or belief-evaluation system) is intact. If the proposed anomalous experience, e.g., the absent affective response in the Capgras delusion, is sufficient to explain the impaired input, then the dissociation of the proposed anomalous experience and the corresponding delusion indicates that there must be something wrong with the reasoning system.

Coltheart has argued, in his first proposal, that there must be a second factor which is responsible for the impaired belief-evaluation system. What is more, in a second proposal, Coltheart suggests that the second factor is the same in all cases of delusions, and that it is associated with a pathology of the right lateral prefrontal cortex (e.g. Coltheart *et al.* 2011, 285).

Aimola-Davies and Davies' 2009 work shows that patients with anosognosia have a putative impairment of the executive function or working memory (or both), corroborating Coltheart's first proposal and being consistent with his second.

To explain a particular delusion, Coltheart's first proposal seems sufficient. The reason why Coltheart made his second proposal seems to be that there is something in common among various delusions, i.e. the similarity of patients' acceptance of the bizarre belief, which also demands explanation. I tend to hold, however, that this does not necessarily have to be so. The similarity may be purely phenomenal—the causes of the similar phenomena do not have to be the same. For example, a broken leg and a broken foot can both make a person unable to walk, but they are different injuries. In any event, we will for now shelve this possibility and focus on exploiting the possibility proposed by Coltheart.

Coltheart's second proposal could be interpreted in two ways. First, one might conclude that the impairment of the evaluation system is exactly the same in different cases of delusion. This entails that if a patient who already has one delusion, e.g. anosognosia, acquires a new anomalous experience, e.g. an absent affective response, then he or she will necessarily have a new delusion. Coltheart does not make such a drastic claim. The second possible interpretation proposed by Coltheart et al. is that "it seems clear that in delusional conditions the belief evaluation system is impaired rather than abolished" (Coltheart, Langdon and McKay 2011, 288–89). This proposal may be further interpreted as suggesting either that there can be various degrees of impairment of the belief-evaluation system in different delusions, or that the belief evaluation system can consist of different functional parts and that the impairment could be among these functional parts.<sup>1</sup> Because of the obvious explanatory weakness of the former interpretation, this paper will focus on the latter.

---

<sup>1</sup>The latter interpretation has been approved by Max Coltheart in a personal email to this paper's author.

It should be noted that the dissociation argument focuses on the need for a second factor. It does not have to hold a position about whether this missing factor figures in before the adoption of the belief, after it, or even before the anomalous experience. Coltheart's idea is that the missing factor should be an impairment of the evaluation system. More specifically, it is a sort of deficit occurring during the progression from anomalous data to the persistence of belief. Even though Coltheart et al. 2010 proposed that the deficit should come after the adoption of the belief, Coltheart's two proposals concerning the second factor do not preclude the possibility that the missing deficit could occur before the adoption of belief (McKay 2010), or even before the anomalous experience. In the next section, I will argue in detail about how different possible impairments could be unified within a single belief evaluation system.

#### **4. How do Different Candidates for the Second Factor be Compatible in a Unified Belief Evaluation System?**

In recent literature, the two-factor theory is combined with the Bayesian inference theory, which tries to give a more specific description of the cognitive nature of the second factor (Coltheart *et al.* 2010; McKay 2012; Davies and Egan 2013). The Bayesian approach assigns every belief a probability and presumes that the probability of the belief,  $P(H)$ , is revised based on the evidence,  $E$ . By Bayes' theorem, the posterior probability  $P'(H) = P(H|E) = P(H) * P(E|H)/P(E)$ .  $P(H)$  is the prior probability of  $H$ ;  $P(E|H)$  is the likelihood. Since a theory of delusion concerns why a delusional belief, e.g. "she is an impostor or a stranger,"  $H_S$ , is adopted and persists rather than a normal belief, e.g. "she is my wife,"  $H_W$ , the posterior probabilities of the different hypotheses can be compared:

$$\frac{P'(H_S)}{P'(H_W)} = \frac{P(H_S|E)}{P(H_W|E)} = \frac{P(H_S)}{P(H_W)} * \frac{P(E|H_S)}{P(E|H_W)}$$



Taking the Capgras delusion as an example, the evidence,  $E$ , is the absent affective response. Whether the patient should, during the process of adoption, favor the stranger hypothesis,  $H_S$ , depends on whether the ratio of likelihoods,  $P(E|H_S)/P(E|H_W)$ , can outweigh the ratio of prior probability,  $P(H_S)/P(H_W)$ .

Coltheart and his colleagues' main idea from their 2010 work is that a normal Bayesian inference would favor  $H_S$  during the process of adoption. As an example, they propose that in a normal situation the ratio of prior probabilities is 1:100 and the ratio of likelihoods is 1000:1. Therefore, the ratio of posterior probabilities is 10:1, which favors  $H_S$ . During the process of persistence, Coltheart and colleagues take friends' or relatives' assertions as the new evidence. They believe that, facing the new evidence in a normal Bayesian inference,  $H_W$  would be favored, because the ratio of likelihoods outweighs the ratio of prior probabilities. Since the patient still favors  $H_S$ , there must be a second factor at work here.

McKay 2012 challenges Coltheart and his colleagues' idea in two ways. McKay points out that, during the adoption process, the proposed ratio of prior probabilities (1:100) is "unrealistically high" (McKay 2012, 339). It would suggest that a husband should expect his wife to be replaced by an identical duplicate one time in every one hundred times he (thinks he) sees her. McKay argues that a normal Bayesian inference would favor  $H_W$ . Since the patient favors  $H_S$ , there must be a second factor during the adoption process. Another claim made by McKay is that the so-called new evidence, i.e. friends' or relatives' assertions, is actually similar to the evidence already considered during the adoption process. For example, if the impostor looks just like his wife, the patient has enough reason to believe that his friends or relatives have also been duped by the impostor. Therefore McKay argues that there is no need for an additional factor to explain the persistence of the belief in the stranger hypothesis.

Davies and Egan 2013 argue that both Coltheart et al. and McKay's Bayesian approaches

presume an ideal belief system which requires “a single coherent distribution of credences that guide action in all contexts” (Davies and Egan 2013, 705). They argue that, on the contrary, normal belief systems are “fragmented or compartmentalized” (Davies and Egan 2013, 705), and that this allows for a post-adoption evaluation of beliefs. More importantly, a compartmentalized mind makes it possible that the personal-level prior probabilities of  $H_W$  and  $H_S$  are discounted during the adoption process. Therefore, there may be no deficit in the processes leading from the anomalous data to the initial adoption of the delusional belief (stage 2). A newly adopted belief will not normally remain encapsulated and, when faced with the post-adoption new evidence, a normal person will favor  $H_W$ , while a patient with the Capgras delusion will favor  $H_S$ . So, there must be a deficit after the adoption of the belief, which is proposed as an impairment of the executive function and/or working memory (stage 3).

Given these arguments, are these proposed second factors not compatible with one another? I do not think so. In the following paragraphs I will argue that these impairments are just different functional parts of a unified belief-evaluation system.

In the Capgras delusion, the firm facts that we have in hand are the delusional belief and the absence of activity in the patient’s autonomic nervous system when confronted with familiar faces. Coltheart 2005, 2010 argues that such an absence of autonomic activity is not available to consciousness because the activity of the same autonomic nervous system cannot help patients with prosopagnosia discriminate familiar from unfamiliar faces (Tranel and Damasio 1985, 1988). Therefore, the absence of activity in the autonomic nervous system (abnormal data) is merely a candidate for the cause of the anomalous experience rather than the anomalous experience itself. A two-factor theory should explain how these two factors can answer questions about the content, adoption, and persistence of the delusional belief.

McKay (2012) argues that a bias in favor of likelihood is the second factor for answering the adoption question and that no more factors are needed to answer the persistence question,

because the so-called new evidence is in fact similar to the evidence considered in the adoption process. It seems that McKay holds the view that no hypothesis can be adopted before it is carefully evaluated, consciously or unconsciously.

However, the compartmentalized belief system, introduced by Davies and Egan (2013), allows that a belief can be adopted without an account of the prior probabilities. In another words, before it is adopted, the hypothesis does not need to be comprehensively evaluated in McKay's sense. The key question is whether all possible beliefs, ordinary or bizarre, would be encapsulated with the discounting of related beliefs.<sup>2</sup> As an example of fragmented belief, Davies and Egan write:

David Lewis describes himself as having once had fragmented beliefs about the geography of Princeton. According to one fragment, Nassau Street ran north-south and was parallel to the railway track; according to another fragment, the railway track ran east-west and was parallel to Nassau Street (Lewis 1982, 436). (Davies and Egan 2013, 705–06)

In this example, the contents of the two beliefs are related. Both of them concern the direction of the railway track and Nassau Street. They are, however, formed or used in different circumstances. Even though the two beliefs logically contradict one another, the consequences of one belief in its own circumstance does not contradict the consequences of the other belief in its own circumstance. What is more, neither of the two beliefs have much influence on other beliefs in Lewis's belief system. For example, no matter what the direction of the railway track or Nassau Street is, Lewis has the same beliefs about how to walk to the university and about where the sun rises. The two belief states are causally isolated from other belief states, and

---

<sup>2</sup> The word "related" does not mean the contents of different beliefs are connected. By related belief, I mean that the influences of the different beliefs are connected.

from one another. They are therefore naturally fragmentary.

As for the competing hypotheses,  $H_W$  and  $H_S$ , in the Capgras delusion, their contents are related and their influences are related too. Different answers to the question of whether the woman you live with every day is your wife or an imposter will have a direct influence on the patient's other beliefs or actions, especially when the delusion is elaborated. Such hypotheses should not be fragmentary within a given patient's mind.

In sum, it may be that when the belief is compartmentalized, the formation of the delusion fits Davies and Egan's theory, and that it otherwise fits McKay's theory. In both cases, the impairments function within a unified belief-evaluation system. Their only difference is which specific functional part's damage causes the impairment of the system.

## **5. Why Is a Function Responsible for Detecting the Bizarreness of Delusional Beliefs also Needed?**

In the literature of the two-factor theory, it is widely accepted that the bizarreness of delusional beliefs cries out for explanation. However, the bizarreness itself has not received enough attention. One possible reason is that it seems that the bizarreness of delusional beliefs is explained away by an impaired belief-evaluation system, which indicates that the bizarreness is equated with abnormal inferential processes. This explanation is not, in my view, persuasive.

Upon second reflection, it is clear that sometimes, even though there is nothing wrong with the inferential processes, we still find certain derived results strange or hard to accept. For example, even though quantum mechanics has been proved to be a significantly powerful and successful theory, few normal people find the world it describes natural or easy to accept. By analogy, should it not be the case that patients find the content of their derived delusional belief bizarre? Unfortunately, they do not, meaning that there must be some additional functional impairment which makes delusional patients fail to recognize the bizarreness of their beliefs.

If this analysis is on the right track, the next step is to find out what could possibly be bizarre enough to draw a normal person's attention. Two candidates naturally present themselves: bizarre content and a bizarre inferential process. Recalling the three stages of delusion—the arising, adoption, and persistence of the belief hypothesis—we can ask the following questions about bizarreness:

(1) How can such bizarre content come into the patient's mind in the first place? In other words, how can the patient create (or accept) such bizarre content?

(2) Once the content comes into the patient's mind, why does the patient take such bizarre content as a candidate hypothesis for belief?<sup>3</sup>

(3) Once the hypothesis is treated as a candidate, why does the patient adopt and maintain such a bizarre belief?

The first question focuses on the source of the content. Sometimes the content of delusion is normal. For example, the content of what patients with anosognosia may believe is not extraordinary; that is, a patient may believe that there is nothing wrong with her legs even though her legs are in fact paralyzed (Aimola-Davies *et al.* 2010). However, sometimes the content is so bizarre that a normal person can hardly imagine how such content comes into the

---

<sup>3</sup>The epistemic possibilities proposed by Parrott 2014 focus mainly on this aspect. Both Parrott and I emphasize the importance of bizarreness. I propose that bizarreness pervades the various stages of delusion, while Parrott focuses his attention on how bizarreness figures before the consideration of the hypothesis. Another difference is that Parrott argues that the consideration of epistemic possibility is a threat to a two-factor theory, while I propose that it is completely consistent with any general two-factor theory. One more thing that bears mention is that Parrott argues that epistemic possibilities are based roughly on public background knowledge. However, this requirement is unnecessary during the adoption stage of delusion, because the initial adoption of the delusion could reasonably be based on the subject's personal knowledge.

patient's mind. For instance, a patient with somatoparaphrenia may believe that her left arm belongs to her niece (Bottini *et al.* 2002). Frankly speaking, I am not sure whether having bizarre content in mind should be taken as pathological. I am inclined to hold that it is not, because if it is, would we then say that any pioneer with novel ideas is mentally ill? Merely considering the logical possibility of some extraordinary idea is better taken as a virtue of being human rather than as a mental symptom. Nevertheless, the question remains why delusional patients are more inclined to have bizarre contents. A simple answer is that they are more inclined to accept novel ideas, or are more creative (Andreasen 2006; Becker 1996). The reason why they are more creative may or may not be that they have one or more mental illnesses, but being creative itself is not pathological. In sum, the answer to the first question is that having bizarre content in mind may be an indication of mental illness—in other words, it may be caused by some mental impairment, which warrants further research—but this is not necessarily so.

Is it normal to take bizarre content in the mind as a candidate hypothesis for belief? We encounter all kinds of ideas every day, and not all of them will be taken as candidates for belief. Imagine that the laziest student, Chuckie, in John's class suddenly claims that he has won first prize in a recent examination. It is so ridiculous that no one, including John, if he is normal, will consider believing it. Here we should emphasize a distinction between epistemic possibilities and logical possibilities (Parrott 2014). A hypothesis is epistemically possible in the sense that it is not ruled out by a rational person's background knowledge. In our case, the statement that Chuckie has won first prize is epistemically impossible in the sense that this statement contradicts everything people in John's class know about Chuckie and a first prize.

Now let us further suppose that John has learned that even though Chuckie behaves lazily at school, he works very hard at home. Should it not be the case that the statement is normally epistemically possible for John? What is more, even though the statement is an epistemic

possibility for John, should it not be the case that he still is surprised to hear that Chuckie has won first prize? Whether or not John ultimately believes it, the hypothesis should normally draw extra attention from John, which means that John should evaluate it more carefully and should feel more reluctant to believe it. If John behaves like delusional patients, which is to say he does not pay extra attention to the hypothesis, then it is surely an indication of pathology. If some classmates show John that it is in fact not true that Chuckie works hard at home, then John should normally immediately discard the hypothesis as a candidate for belief, partly thanks to his constant vigilance. On the contrary, without a certain degree of vigilance, John may abnormally discount the overwhelming counter-evidence in his background knowledge and even ignore the fact pointed out by his classmates, wrongly taking the statement as he would any ordinary hypotheses, e.g., the hypothesis that Chuckie is a human being. Lacking the ability to prevent oneself from taking an unusual or bizarre hypothesis as a candidate for belief is definitely a symptom of pathology.

Someone may wonder whether this function can be included in a Bayesian two-factor theory. The debate concerns the relationship between prior probabilities and epistemic possibilities. A complete Bayesian needs to deal with how to settle prior probabilities, and it seems natural to settle prior probabilities by settling epistemic possibilities. If this analysis is right, then surely this function should be a part of a complete Bayesian theory, not to mention a complete two-factor theory.

Even though a bizarre hypothesis has been taken as a candidate for belief, does this mean that the previously-discussed Bayesian theory is powerful enough to describe what happens next through the various stages of delusion? In the following paragraphs, I will argue that it is not powerful enough to do this by showing how a monitoring system, responsible for detecting bizarreness, figures in different stages of delusion.

Turner and Coltheart 2010 propose a monitoring system which consists

[o]f (a) an unconscious checking system that either “tags” suspect thoughts as requiring extra conscious evaluation, or “passes” such thoughts, thus conferring conviction, and (b) a conscious checking system that enables elaborative and effortful evaluation of thoughts in the light of alternative evidence. (Turner and Coltheart 2010, 372)

Therefore, in McKay’s theory,<sup>4</sup> if the unconscious checking system is intact, then during the process of adoption a Bayesian inference with a bias to likelihood would unconsciously favor  $H_S$ . Since  $H_S$  is, however, tagged by the unconscious checking system, and since this tag “gives rise to the experience of doubt,”  $H_S$  is “referred to the conscious evaluation system for further work” (Turner and Coltheart 2010, 357). Therefore,  $H_S$  should be consciously rejected, whether or not  $H_S$  was initially adopted as a belief. If a patient consciously believes  $H_S$ , it means that the patient’s conscious evaluation system is impaired, which may be caused by an impaired executive function and/or working memory (Aimola-Davies and Davies 2009; Davies and Egan 2013).

If the conscious checking system is impaired, then, when new evidence or a competitive hypothesis is presented to the patient,<sup>5</sup> he or she will ignore it and still believe  $H_S$  as before. If the conscious checking system is intact, then  $H_S$  should be tagged. If, however, the evaluation system is intensively impaired, the patient will also maintain  $H_S$ . If the impairment of the evaluation system is not so serious, then repeated alerts from the conscious checking system

---

<sup>4</sup>It should be noted that McKay makes “no assumptions about the extent to which the generation and evaluation of candidate hypotheses are conscious, person-level processes. In reality Bayesian inference is likely to be occurring simultaneously at many levels of a hierarchy” (McKay 2012, 335, footnote 7).

<sup>5</sup>The evidence or competitive hypothesis is not new for the former Bayesian inference, but it is new for the conscious checking system.



will force the patient to try his/her best to reevaluate  $H_S$ , and the patient may have a chance to finally reject the hypothesis.

If the unconscious checking system is impaired, an impaired Bayesian inference will make the patient adopt  $H_S$  without initial hesitation. If the conscious system is also impaired, the patient will ignore new evidence or a competitive hypothesis and still believe  $H_S$ . If the conscious system is intact, meanwhile, the patient will reevaluate  $H_S$ . Because of the impairment of the Bayesian system, the patient may still favor  $H_S$ , but will do so warily.

Davies and Egan 2013 propose that a normal Bayesian inference, in an unconscious perceptual module which discounts the prior probabilities of a competitive hypothesis, will favor the delusional hypothesis. They assume that “the favored hypothesis would determine the content of an experience that presents or represents the world as being a certain way” (Davies and Egan 2013, 714). Furthermore, they propose that

[t]he delusional belief arises as a prepotent doxastic response and ... the processes leading to initial adoption of the belief discount personal-level prior probabilities and are biased in favor of likelihoods. (Davies and Egan 2013, 716)

I agree with the idea that a perceptual model would discount the prior probabilities. However, it seems unnecessary to insist that a prepotent doxastic tendency itself would discount the prior probabilities. Since both the experience and the prepotent doxastic tendency are conscious, an intact unconscious checking system will tag  $H_S$  such that  $H_S$  may not be adopted as a belief in the first place.

For example, most normal people have seen that when part of a stick is put into water, the stick looks bent. The perceptual model would favor the hypothesis that the stick is bent. If a normal prepotent doxastic tendency always discounts the prior probabilities, most normal

people will adopt the belief that the stick is bent. However, most normal people do not have such a belief, nor is there any indication that the belief is ever adopted. In fact, there is evidence which indicates that even babies will be surprised when bizarre things happen. This phenomenon may be explained as babies possessing the ability to notice the bizarreness. A plausible explanation of why people do not believe that the stick is bent is that such a hypothesis is so bizarre that prior probabilities should be considered before it is adopted as a belief.

By analogy, if a normal person without any other impairments had the anomalous experience, his perceptual module might favor  $H_S$ . However, before adoption, he would normally notice the bizarreness of the hypothesis, such that this hypothesis will not be adopted as a belief in the first place. The adoption of the delusional belief will indicate an impairment of the monitoring system's unconscious checking system.

In Davies and Egan's theory, a patient with impairment of both the conscious and unconscious checking system will not even bother to evaluate the hypothesis during the process of persistence. If the conscious checking system is intact, or if the patient is forced to evaluate the hypothesis, the impaired executive function and/or working memory will prevent him or her from giving the normal evaluation.

In sum, a monitoring system is consistent with both McKay's and Davies and Egan's theory. By adding a monitoring system to the belief-evaluation system, more complicated cases of delusion could be easily explained. I leave it to the reader to consider how a monitoring system could figure in more complicated cases.

My argument here is, first, that three questions need to be answered in order to explain delusion; second, that a function responsible for detecting bizarreness is needed to answer these questions; and third, that such a function is completely consistent with a two-factor theory. It should be emphasized that what I have proposed does not rely on any specific two-factor theory, since the questions it deals with are general.

Admittedly, it is still unclear how to describe the relations among a bias favoring likelihood (or a bias discounting prior probabilities), the monitoring system, and an impaired executive function and/or working memory. For example, a conscious checking system may need the help of the executive function, working memory, and long-term memory; moreover, the failure of the monitoring system may be caused by a bias that discounts the prior probabilities.<sup>6</sup> Nonetheless, in the above discussion, it is clear that each deficit plays a separate functional role. The difficulty is caused by the fact that two distinct high-level functional parts may share a common low-level functional part, or simply share the same collection of neurons (Fodor 1974). Such difficulty should not however preclude a theory of delusion that is effective at a certain level.

## References

- Aimola-Davies, A. M., and Davies, M. 2009. "Explaining Pathologies of Belief," in *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, edited by Broome and L. Bortolotti, 285–323. Oxford: Oxford University Press.
- Aimola-Davies, A. M., White, R. C., and Davies, M. 2010. "Assessment of Anosognosia for Motor Impairments," in *Handbook of Clinical Neuropsychology (Second Edition)*, edited by J. M. Gurd, U. Kischka and J. C. Marshall, 436–68. Oxford: Oxford University Press.
- American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision (DSM–IV–TR), 821. Washington, DC: American

---

<sup>6</sup> In different interpretations, a bias discounting the prior probabilities may or may not necessarily make the monitoring system fail. For example, if the prior probabilities just concern the logical possibility of the hypothesis while the monitoring system focuses on the contradiction between the hypothesis and other ordinary beliefs, the former bias may not make the monitoring system impaired.

Psychiatric Association.

Andreasen, N. C. 2006. *The Creative Brain: The Science of Genius*. Plume.

Becker, G. 1996. *Touched with Fire: Manic-Depressive Illness and the Artistic Temperament-Jamison*. Simon and Schuster.

Bortolotti, Lisa. 2013. "Delusion." *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta ed., URL = <<http://plato.stanford.edu/archives/win2013/entries/delusion/>>.

Bottini, G., Bisiach, E., Sterzi, R., and Vallar, G. 2002. "Feeling Touches in Someone Else's Hand," *NeuroReport* 13: 249–52.

Brighetti, G., Bonifacci, P., Borlimi, R., and Ottaviani, C. 2007. "'Far from the Heart far from the Eye': Evidence from the Capgras Delusion," *Cognitive Neuropsychiatry* 12: 189–97.

Bruce, V. and Young, A. W. 1986. "Understanding Face Recognition," *British Journal of Psychology* 77: 305–27.

Capgras, J. and Reboul-Lachaux, J. 1923. "L'illusion des "sosies" dans un délire systématisé chronique." *Bulletin de la Société Clinique de Médecine Mentale* 2: 6–16.

Coltheart, M. 2005. "Conscious Experience and Delusional Belief." *Philosophy, Psychiatry, & Psychology* 12: 153–7.

Coltheart, M. 2007. "Cognitive Neuropsychiatry and Delusional Belief." *Quarterly Journal of Experimental Psychology* 60: 1041–62.

Coltheart, M. 2001. "The Neuropsychology of Delusions." *Annals of the New York Academy of Sciences* 1191: 16–26.

Coltheart, M., Langdon, R., and McKay, R. 2011. "Delusional Belief." *Annual Review of Psychology* 62: 271–98.

Coltheart, M., Menzies, P., and Sutton, J. 2010. "Abductive Inference and Delusional Belief." *Cognitive Neuropsychiatry* 15:261–87.

Cotard, J. 1882. "Du délire des négations." *Archives de Neurologie* 4: 152–70, 282–95.

- Davies, M. 2000. "Interaction without Reduction: The Relationship between Personal and Sub-Personal Levels of Description." *Mind and Society* 1: 87–105.
- Davies, M., Aimola Davies, A. M., and Coltheart, M. 2005. "Anosognosia and the Two-Factor Theory of Delusions." *Mind & Language* 20:209–36.
- Davies, M. and Coltheart, M. 2000. "Introduction: Pathologies of Belief." *Mind & Language* 15: 1–46.
- Davies, M., Coltheart, M., Langdon, R., and Breen, N. 2001. "Monothematic Delusions: Towards a two-Factor Account." *Philosophy, Psychiatry, & Psychology* 8:133–58.
- Davies, M. and Egan, A. 2013. "Delusion, Cognitive Approaches: Bayesian Inference and Compartmentalization," in *The Oxford Handbook of Philosophy and Psychiatry*, edited by K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Sadler, G. Stanghellini and T. Thornton, 689–727. Oxford: Oxford University Press.
- Edelstyn, N. M. J. and Oyebode, F. 1999. "A Review of the Phenomenology and Cognitive Neuropsychological Origins of the Capgras Syndrome." *International Journal of Geriatric Psychiatry* 14: 48–59.
- Ellis, H. D. 1998. "Cognitive Neuropsychiatry and Delusional Misidentification Syndromes: An Exemplary Vindication of the New Discipline." *Cognitive Neuropsychiatry* 3:81–90.
- Ellis, H. D., Lewis, M. B., Moselhy, H. F., and Young, A. W. 2000. "Automatic without Autonomic Responses to Familiar Faces: Differential Components of Covert Face Recognition in a Case of Capgras Delusion." *Cognitive Neuropsychiatry* 5: 255–69.
- Ellis, H. D. and Young, A. W. 1990. "Accounting for Delusional Misidentifications," *British Journal of Psychiatry* 157: 239–48.
- Ellis, H. D., Whitley, J., and Luauté, J. P. 1994. "Delusional Misidentification: The Three Original Papers on the Capgras, Fregoli and Intermetamorphosis Delusions." *History of Psychiatry* 5: 117–46.

- Halligan, P. W., Marshall, J. C., and Wade, D. T. 1995. "Unilateral Somatoparaphrenia after Right Hemisphere Stroke: A Case Description." *Cortex* 31: 173–82.
- Hirstein, W. and Ramachandran, V. S. 1997. "Capgras Syndrome: A Novel Probe for Understanding the Neural Representation of the Identity and Familiarity of Persons." *Proceedings of the Royal Society: Biological Sciences* B264: 437–44.
- Fodor, J. A. 1974. "Special Sciences (or: the Disunity of Science as a Working Hypothesis)." *Synthese* 28.2: 97–115.
- Langdon, R. and Bayne, T. 2010. "Delusion and Confabulation: Mistakes of Perceiving, Remembering and Believing." *Cognitive Neuropsychiatry* 15: 319–45.
- Langdon, R. and Coltheart, M. 2000. "The Cognitive Neuropsychology of Delusions." *Mind & Language* 15:184–218.
- McKay, R. 2012. "Delusional Inference." *Mind & Language* 27: 330–55.
- Maher, B. A. 1999. "Anomalous Experience in Everyday Life: Its Significance for Psychopathology." *The Monist* 82: 547–70.
- Murphy, Dominic. 2015. "Philosophy of Psychiatry," *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta ed., URL = <http://plato.stanford.edu/archives/spr2015/entries/psychiatry/>.
- Parrott, M. 2014. "Bayesian Models, Delusional Beliefs, and Epistemic Possibilities." *The British Journal for the Philosophy of Science*, axu036: 1-26.
- Tranel, D. and Damasio, A. R. 1985. "Knowledge without Awareness: An Autonomic Index of Facial Recognition by Prosopagnosics." *Science* 228:1453–54.
- Tranel, D. and Damasio, A. R. 1988. "Non-Conscious Face Recognition in Patients with Face Agnosia." *Behavioural Brain Research* 30: 235–49.
- Tranel, D., Damasio, H., and Damasio, A. R. 1995. "Double Dissociation between Overt and Covert recognition." *Journal of Cognitive Neuroscience* 7: 425–32.

Turner, M. and Coltheart, M. 2010. "Confabulation and Delusion: A Common Monitoring Framework." *Cognitive Neuropsychiatry* 15: 346–76.

Young, A. W. and Leafhead, K. M. 1996. "Betwixt Life and Death: Case Studies of the Cotard Delusion," in *Method in Madness: Case Studies in Cognitive Neuropsychiatry*, edited by P. W. Halligan and J. C. Marshall, 147–71. Hove: Psychology Press.