

# The Causal Role Argument against Doxasticism about Delusions

**Kengo Miyazono**, Philosophy Department, University of Birmingham  
*k.miyazono@bham.ac.uk*

**Lisa Bortolotti** Philosophy Department, University of Birmingham  
*l.bortolotti@bham.ac.uk*

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

## Abstract

In this paper we consider an argument that is very influential in the philosophical literature, the argument from causal role against the view that delusions are beliefs. The argument has two premises, that many delusions fail to play belief-roles and that playing belief-roles is necessary for a mental state to be a belief. We assess both premises and suggest that they can be resisted.

**Keywords:** delusions; beliefs; double bookkeeping; motivation; teleo-functionalism.

## Introduction

Doxasticism about delusion, “doxasticism” hereafter, is the claim that delusions are beliefs. Delusions are usually regarded as beliefs in psychiatry. For instance, in the DSM-5 delusion is defined as a “false belief based on incorrect inference about external reality that is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary” (American Psychiatric Association 2013: 819). Moreover, cognitive scientists working on theories of delusion formation assume that the mechanisms responsible for the formation of delusions are also the mechanisms responsible for the formation of beliefs (Coltheart 2007; Corlett et al. 2010).

However, doxasticism is not very popular among philosophers. Greg Currie and colleagues (Currie 2000; Currie and Jureidini 2001) argue that delusions are not beliefs but imaginings. Andy Egan (2009) thinks that they are “bimagnations” (i.e. they are the states with some belief-like features and some imagination-like features). Eric Schwitzgebel (2012) suggests that they are “in-

between beliefs” (i.e. they are not beliefs, but not non-beliefs). Jakob Hohwy and colleagues (Hohwy and Rosenberg 2005; Hohwy and Rajan 2012) propose that they are perceptual inferences.

The main argument against doxasticism is what we call “the argument from causal role” and what Tim Bayne (2010) calls “the functional role objection.” We believe that the argument is, at best, wanting despite its numerous supporters within philosophy. The aim of this paper is to show some of the limitations of the argument from causal role. Doxasticism and the argument from causal role will be introduced in Section 1. The argument has two premises; premise (1) says that many delusions fail to play belief-roles, and premise (2) says that playing belief-roles is necessary for a mental state to be a belief. Premise (1) and premise (2) are critically examined in Sections 2 and 3 respectively. Our conclusion will be that both premises can be resisted.

## 1. Doxasticism and the Argument from Causal Role

Delusions are beliefs, according to doxasticism. There are some *prima facie* reasons for this claim.

(1) People with delusions sincerely assert the content of their delusions. This gives a *prima facie* reason for doxasticism, since a person’s sincere asserting that P gives us a *prima facie* reason to think that she believes that P. If a person does not believe that P, but, for instance, merely imagines that P, then why does she sincerely assert that P? We do not usually sincerely assert that P when we merely imagine that P.

(2) People with delusions typically regard their delusions as beliefs, which is another *prima facie* reason for doxasticism. LA-O, a woman with anosognosia (denial of illness), denies that her left hand belongs to her (Bisiach and Gemiani 1991). If we ask LA-O: “Do you believe that the left hand doesn’t belong to you?” she will definitely say “Yes, I do believe it.” She will not say, for instance, “No, I don’t. I am just imagining it.” Although self-reports of mental states are fallible, it would be fair to assume that they give *prima facie* reasons.

(3) Delusions are regarded as beliefs not only by people reporting delusions, but also by psychiatrists. In introductory psychiatry textbooks and research papers, delusions are usually described as beliefs (Berrios 1991 is a notable exception). In general, when x, who knows y well (y’s behaviour, preferences, value, etc.), judges that y believes that P, then it gives a *prima facie* reason to think that y believes that P. In our case, many psychiatrists, who know people with delusions well, judge that their patients believe the content of their delusions. This gives us a *prima facie* reason to think that delusions are beliefs.

(4) Delusions are a pathological phenomenon and the truth of doxasticism seems to be part of the reason why delusions are pathological. For instance, LA-O's mental condition is pathological partly because she seriously denies that her left hand belongs to her. If she did not believe it, but merely imagined it, there would not be anything particularly pathological about her condition, as acts of imagination do not necessarily reflect how things are for the person engaging in the imagining. It is *a strange thing* for LA-O to imagine that her left hand does not belong to her, but we can easily entertain various kinds of strange possibilities in our imagination without losing mental health.

(5) Doxasticism helps us make sense of the distinction between delusions and other pathological mental phenomena. First, one can distinguish delusion from hallucination by saying that the former is a belief, while the latter is a perceptual state. In fact, this is the normal way in which psychiatry textbooks introduce the distinction between them. Second, doxasticism helps us make sense of the distinction between delusion and some linguistic disorders such as jargon aphasia. People with jargon aphasia make strange utterances due to linguistic impairments. In contrast, LA-O makes strange claims, such as "This is not my hand, but someone else's!" not because of linguistic impairments, but because of the fact that she really believes strange things. Doxasticism might also enable us to make sense of the distinction between delusion and obsessive thought. People with obsessive thoughts about contamination by germs have more acute awareness of the strangeness of their thoughts than people with delusions, and this supports the view that they do not actually believe that they have been contaminated.

Despite of these *prima facie* reasons, there are some philosophical arguments against doxasticism. Among others, the following seems to be playing the central role in recent debates.

#### **Argument from Causal Role**

- (1) Many delusions fail to play belief-roles.
- (2) A mental state is a belief only if it plays belief-roles.
- (3) Therefore, many delusions are not beliefs.

Here, "belief-roles" refers to causal roles that are distinctively belief-like. Premise (2) is entailed by functionalistic theories of belief, including standard functionalism (i.e. to believe is to be in a state that plays belief-roles), standard representationalism (to believe is to have a representation that plays belief-roles), and dispositionalism (to believe is to have some belief-like dispositions). Eric Schwitzgebel (2012) calls them "token-functionalisms." They are *token* functionalisms because according to them every single token of a belief-type needs to play belief-roles. Anti-doxastic philosophers tend to accept one of these theories. For instance, Currie and colleagues, Egan, and Schwitzgebel

accept standard functionalism, standard representationalism, and dispositionalism respectively.

Premise (1) is supported by clinical observations. It is commonly assumed that playing belief-roles includes being more or less sensitive to evidence, being more or less coherent with other beliefs, guiding action, and causing appropriate affective responses. What clinical observations often reveal is that delusions lack these features.

First, delusions do not easily respond to evidence. This is usually regarded as an essential, definitional feature of delusion. For instance, the earlier quote from the DSM-5 says that delusion “is firmly held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary.”

Second, delusions often fail to be coherent with non-delusional beliefs and commitments. For instance, a young woman with Cotard delusion, LU, was committed to the idea that she was dead, although she recognized that dead people cannot move and talk (McKay and Cipolotti 2007).

Third, delusions sometimes fail to guide action. As Stone and Young put it,

[...] although in some cases of Capgras delusions patients act in ways that seem appropriate to their beliefs, in many other cases one finds a curious asynchrony between the firmly stated delusional belief and actions one might reasonably expect to have followed from it. [...] This failure to maintain a close co-ordination of beliefs and actions may be typical of the delusions that can follow brain injury. (Stone and Young 1997: 334)

Fourth, delusions sometimes fail to cause appropriate affective responses. For instance, a man with Capgras delusion, who believed that his wife had been replaced by a “double”, never became hostile or aggressive to the “double”, but rather treated her in a very gentle manner, showing some positive affective feelings (Lucchelli and Spinnler 2007).

Those clinical observations are certainly important, but should not be exaggerated. It is not the case that all delusions have all of the features above. For instance, a delusion might be irresponsive to evidence, but guide action. People with Capgras delusion who report that their loved ones are replaced by imposters can act violently towards the “imposters” and harm or even kill them (Silva et al. 1994). Some delusions show none of the features above and seem to behave just like paradigmatic beliefs. For instance, some persecutory delusions impact on people’s cognitive and affective life, are not obviously disconfirmed by the available evidence, lead to action, and seem to be in a relation of mutual support with some non-delusional beliefs (Payne 1992). Even if delusions have some of the features above, it does not mean that they are different from paradigmatic beliefs in the relevant aspects. For instance, some delusions fail to guide action. Still, those delusions are typically reflected

in appropriate verbal behaviour. In other words, people with delusions verbally behave as if they seriously believe the content of their delusions.

In the rest of the paper, we shall examine how good the causal role argument actually is. We shall discuss premise (1) and premise (2) in Section 2 and 3 respectively.

## 2. Examining Premise (1)

Premise (1) says that many delusions fail to play belief-roles. In this section, we shall consider whether the premise is sound, focusing on the relationship between beliefs and action.

### 2.1. Do delusions really fail to guide action?

The most influential claim in support of premise (1) is that delusions fail to guide action in the relevant circumstances. This phenomenon is described and accounted for in different ways. In the context of schizophrenia, people with delusions are charged with “double bookkeeping” (Sass 2001; Gallagher 2009). Keith Frankish (2009) talks about “behavioural inertness” as the relative lack of influence that certain professed beliefs have on behaviour.

The notion of double bookkeeping suggests that people with delusions are simultaneously committed to their delusional report and to a non-delusional belief, which conflicts with their delusional report. An example would be that of a person with Cotard delusion (like LU in the case studied by McKay and Cipolotti 2007) who asserts sincerely that she is dead, but at the same time does not find it surprising that she can move and talk. One version of the double bookkeeping view is endorsed by Shaun Gallagher (2009). He argues that the person with delusions inhabits *multiple realities* at the same time, that is, the delusional reality where the delusional report is genuinely endorsed and the actual reality where the delusional report is at some level recognised as false. For instance, in the patient’s delusional reality it is true for him that the hospital nurses are attempting to poison him, but in his actual reality he realises that the nurses pose no threat—and his behaviour is consistent with the actual reality as the patient eats the food the nurses give him.

Similarly, the notion of behavioural inertness is about the delusion failing to give rise to the actions we would expect a person to initiate if she did genuinely believe the content of her delusion. Relying on a dual-process theory of the mind, Keith Frankish (2009; 2012) likens delusions to *acceptances* rather than beliefs, where acceptances are policies the person openly endorses but would not be ascribed to the person on the basis of her observable behaviour. The person with Capgras delusion who claims that his wife has been replaced by

an impostor but shows affection to the alleged impostor (Lucchelli and Spinnler 2007) seems to exemplify behavioural inertness. An external observer would not ascribe to the person the belief that his wife has been replaced by an impostor.

But not all cases of delusions illustrate double bookkeeping or behavioural inertness. We should not forget that delusions are diagnosed on the basis of their behavioural manifestations, including stress, depression, preoccupation, social withdrawal, and impaired functioning more generally. In some cases, behavioural manifestations are specific to a type of delusion and dictated by its content: for instance, in persecutory delusions people manifest “safety behaviours”, that is, they avoid situations that they perceive as threatening given their delusions, and they are emotionally distressed with respect to the content of their delusions (Freeman, Garety and Kuipers 2001).

There are also circumstances in which people behave in accordance with their delusions even though the ensuing actions have very significant costs associated with them. Here are some examples. Affected by perceptual delusional bicephaly, the delusion that one has two heads, a man who believed that the second head belonged to his wife’s gynaecologist attempted to attack it with an axe. When the attack failed he attempted it to shoot it and as a consequence he was hospitalised with gunshot wounds (Ames 1984). A man who believed that a lizard was inside his body due to the scaly appearance of his skin tried to remove the lizard with a knife, harming himself as a result (Browning and Jones 1988). Some people with Cotard delusion remain motionless, refuse to eat or wash, and even speak in sepulchral tones (Weinstein 1996).

To sum up, there are cases in which delusions fail to guide action, and cases in which they drive (specific and costly) actions. Where does this leave us with respect to premise (1)?

## **2.2. The role of motivation**

Several strategies to explain the apparently conflicting evidence have been attempted, but here we want to put some pressure on the claim that delusions fail to influence or drive action because people lack doxastic commitment to the content of the delusions. We do so by referring to the hypothesis that double bookkeeping and behavioural inertness are due to a failure of supporting the motivation to act. This is plausible because schizophrenia negatively impacts on motivation (Bortolotti 2010; Bortolotti and Broome 2012).

In the debate about the nature of delusions, philosophers impressed by the argument from causal role have not paid much attention to motivation either in the generation of an intention to act given a certain belief, or in the conver-

sion of an intention into an action. Put simply, the suggestion is that one of the reasons why in some circumstances delusions fail to give rise to appropriate action is that the person genuinely believes the content of the delusions, but cannot acquire or sustain the motivation to act on it. Lack of motivation can inhibit action in paradigmatic beliefs, but there are very good reasons to suppose that in the case of people with delusions the factors undermining motivation are much more powerful. At least in delusions emerging in the context of schizophrenia, motivation is adversely affected for a variety of reasons, including widespread co-morbidity with depression (Siris 1995), flat affect, neuropsychological deficits, emotional disturbances, and lack of support from the social and physical environment in which the agent finds herself.

How would these factors impact on the observed inconsistencies between delusional reports and behaviour? First, intentions to act may not be generated due to conflicting attitudes towards the content of the delusion that give rise to incompatible goals and paralyse action. Second, relevant and appropriate goals may be initially identified, but become inaccessible at a later stage due to a deficit in meta-representation, such as limitations in imagination and projection—Chris Frith (1992) calls this *poverty of action*. Third, fluctuating conviction in the content of the delusion may undermine the stability of the person's goals.

Moreover, even when the intention to act is generated and the person's goals preserve stability, other factors may interfere, and this applies to a population with schizophrenia more than to non-clinical or other clinical populations. For instance, there may be no appropriate emotional reaction towards the goal, and thus the goal may not be found sufficiently desirable to lead to action (Gard et al. 2007; Kring and Elis 2013). More generally, the emotions that would support relevant action are disturbed due to stress and anxiety (Freeman and Garety 2003; Freeman, Garety and Kuipers 2001). In addition, due to a reduced sense of self-efficacy and competence, and a reduced sense of autonomy (both very common in schizophrenia), the person can lose confidence in her ability to achieve the relevant goals and doubt her capacity to carry out the relevant actions (Ryan and Deci 2010).

Finally, the impact of the external environment on the person's motivation to act on her delusion should not be underestimated. Plausibly, for a person who often experiences isolation, there are fewer opportunities to engage in activities that would promote the pursuit of her goals and fewer social incentives to act, given that the people surrounding her would not be supportive of the delusion and the actions that would ensue from it. Also, we should not forget that, given the absurd content of some delusions, acting consistently with the delusion is often not possible (a person with Cotard delusion can *act dead*, but she will not succeed in behaving *as if she were dead*).

Examining the role of motivation in the case of delusions in schizophrenia may further persuade us that premise (1) does not warrant anti-doxastic accounts of delusions, even if we were to embrace premise (2). It seems to us that some of the initial plausibility of premise (1) is due to the claim that people do not act on their delusions. Thus delusions fail to play the most central of all belief-roles, that which distinguishes beliefs from other mental states, i.e. the capacity to influence behaviour and drive action. But the claim that people do not act on their delusions is unwarranted when we consider the clinical evidence available to us. Even when we acknowledge that people do not behave in the way we would expect if they genuinely believed the content of their delusion, philosophers have underestimated how inaction or inconsistent behaviour may be due to factors that undermine motivation in schizophrenia and possibly other disorders of which delusions are a symptom. Premise (1) works within the argument from causal role only if we assume that the best explanation for a delusion failing to play an action-guiding role is that the person does not genuinely believe the content of the delusion. There are other explanations for failure of action guidance. The role of belief is to guide action in conjunction with the right kind of motivation. If the capacity to acquire or preserve motivation is compromised in schizophrenia, then it is not surprising that delusions may fail to guide action, even if they are belief states.

### **2.3. Interim conclusions**

Glenn Roberts reminds us of the continuity between delusional reports and belief reports:

It is also worth emphasising that irrational beliefs are by no means limited to the deluded. [...] Rachman (1983) has demonstrated the propensity for ‘normals’ to produce irrational thinking and form false conclusions from the available information. He gives a vivid picture of the occurrence of cognitive illusions which are strenuously defended as true, even when the subjects are informed that they are false, and he summarises the biases found in cognitive studies of otherwise normal subjects. Similarly, Rothstein (1980) traces the distortion of scientific theories by irrational elements introduced through the ‘narcissistic investment’ of the theorists in their own creations. This may be directly comparable with the relationship the deluded have with their beliefs. He says that when the narcissistically invested theory is challenged, it is characteristically defended by “devising numerous articulations and ad hoc modifications in order to eliminate any apparent conflict. As such it assuages his sense of vulnerability and helplessness.” (Roberts 1992: 304)

We believe that the described continuity between delusions and non-delusional beliefs also extends to the way in which reported attitudes are



manifested in behaviour and lead to action. Contrary to the spirit, if not the letter, of the argument from causal role, we suggest here that we should think about *delusions and action* in the same way we think about *beliefs and action*. Rational agents have beliefs that are not shaped by their desires, are well-supported by evidence, do not conflict with their other beliefs, and drive action in a consistent way. But rational agents are sparse both in the clinical and non-clinical population. Actual agents can be subject to positive illusions or self-deception when they form beliefs that fill an emotional need or play a defensive function. They have beliefs that are sensitive to evidence only some of the time and tend to strenuously defend previously endorsed beliefs in the face of counterevidence. They have beliefs that can and often do conflict with some of their other beliefs as we see from the literature on cognitive dissonance and preference reversals. Most important to us, actual agents do not consistently act on their beliefs.

Accepting a more psychologically realistic account of what agents can do will reduce the perceived gulf between people with delusions and people without, and will also contribute to deflating some popular arguments against doxasticism, including the argument from causal role.

### **3. Examining Premise (2)**

Premise (2) says that a mental state is a belief only if it plays belief-roles. This is a consequence of token-functionalisms. In this section, we consider two questions. First, are there any good arguments for (2)? Second, are there any good alternatives to token-functionalisms that do not entail (2)? We shall answer the first question negatively and the second question positively. In other words, there is no good argument for premise (2) and there is at least a good alternative theory that does not entail (2). We take them to show that premise (2) can be resisted.

#### **3.1. Any arguments for (2)?**

Let us call the possibility of beliefs without belief-roles “multiple-functionability of belief”. In this terminology, premise (2) says that belief is not multiply-functionable. Now, are there any good arguments against the multiple-functionability of belief?

A possible argument against multiple-functionability of belief would go like this: How can we identify beliefs without belief-roles as beliefs? How can we know that they are beliefs? We usually figure out what mental states people are in by looking at their causes and effects. For instance, we come to think that one believes that it is raining outside on the basis of the fact that one behaves as if it is the case that it is raining outside. If there are some beliefs

without belief-roles, then we will have serious troubles in identifying them as beliefs.

However, this argument is not very compelling. We usually know the temperature of a room by looking at the thermometer in the room. When the temperature of the room fails to have the right kind of impact on the thermometer's readings (because, for instance, the thermometer is broken), it will be difficult to know the temperature. But nobody takes this to show that it is essential for the temperature of the room that it has a right kind of impact on the thermometer's readings. Now, certainly, if some beliefs fail to play belief-roles, then it would be difficult to identify them as beliefs. But it would not show that it is essential for beliefs that they play belief-roles any more than in the case of thermometer (Sober 1985).

Here is another argument. One might think that the multiple-realizability argument, which is the standard argument for functionalism, rules out the multiple-realizability of mental states. Beliefs seem to be multiply-realizable. After all, we do not deny beliefs to some non-human creatures just because those creatures are neurophysiologically different from us. But then it looks as though the only thing that is shared by all beliefs is the fact that they play the same kind of causal roles, namely, belief-roles. Thus, beliefs would simply be regarded as the mental states with belief-roles.

However, this argument fails. The multiple-realizability does not rule out multiple-functionability, because there might be some theories of mental states that allow for multiple-realizability and multiple-functionability at the same time. For instance, David Lewis (1983) proposed such a theory. Lewis believes that pain is multiply-realizable because a Martian, who is neurophysiologically different from us, can be in pain. He also thinks that pain is multiply-functionable because a guy whose pain does not play pain-roles can be in pain. According to Lewis, a successful theory of pain needs to allow for the multiple-realizability and multiple-functionability of pain at the same time. Lewis's proposal is this:

x feels pain iff x is in a physical state of the type whose tokens statistically-normally play pain-roles in the appropriate population to which x belongs.

The theory certainly allows for those possibilities. First, the Martian can be in pain because it is perfectly possible that he is in a physical state of the type whose tokens statistically-normally play pain-roles in his population. Maybe the Martian has some inflated cavities in his hydraulic brain, and inflated cavities statistically-normally play pain-roles in the Martian population. Second, the mad guy can be in pain because it is perfectly possible that he is in a physical state of the type whose tokens statistically-normally play pain-roles in his population. Maybe the mad guy has some firing C-fibres in his head, and C-fibre firing statistically-normally play pain-roles in the human population. It

is not our claim here that this is a good theory of pain. The point of this is merely to illustrate that multiple-realizability is perfectly compatible with multiple-functionability.

### **3.2. An alternative theory of belief**

Are there any independently motivated theories of belief that do not entail premise (2)? In other words, are there any independently motivated theories that allow for multiple-functionability of belief? There are several candidates (e.g. an application of Lewis's theory to beliefs), but we focus on one in particular.

Peter Godfrey-Smith introduces a useful distinction between two basic forms of functionalism:

Most recent philosophy of mind has been "functionalist" in some sense or other. We can distinguish two basic forms of functionalism in philosophy of mind. First, there is the more orthodox view which I will call "dry functionalism." This view understands function in terms of causal role, and it identifies mental states in terms of their typical causal relation to sensory inputs, other mental states, and behavioral outputs. Second there is "teleo-functionalism." The view makes use of a richer, biological concept of function more closely allied to traditional teleological notions, a concept often analyzed with the aid of evolutionary history. For the dry functionalists, one essential property of any mental state is the pattern of behavioral outputs which the state, in conjunction with the rest of the system, tends to cause in various circumstances. For the teleo-functionalists, what is essential to the mental state is not what it tends to do but what it is supposed to do. (Godfrey-Smith 1998: 13)

According to dry functionalism, mental states are defined by the causal roles they play. Beliefs, for example, are the mental states that play (or tend to play) belief-roles. This basically corresponds to token-functionalism in Schwitzgebel's terminology. According to teleo-functionalism, on the other hand, mental states are defined by the etiological functions they have. Beliefs, for example, are the mental states that have distinctively belief-like etiological functions. Let us call them "doxastic functions." Etiological functions of something are determined by their evolutionary history. To have a certain etiological function is to have a certain evolutionary history. For example, a very simple definition of etiological function is that *x* has the etiological function of doing *F* just in case the ancestors of *x* were selected for doing *F*. (Although the definition can be complicated in several different ways, those complications are not directly relevant in our discussion.)

A number of philosophers defended teleo-functionalist theories for various theoretical reasons (Lycan 1987; Sober 1985; Sterelny 1990): (1) teleo-functionalism can deal with some counterexamples to token functionalism; (2) teleo-functionalism is more promising in accounting for consciousness than

token functionalism; (3) teleo-functionalism is more consistent than token functionalism with the function-analytic explanatory strategy which is widely used in psychology. Relatedly, teleosemantics is arguably the most influential naturalistic account of the content of mental states (Dretske 1986; Millikan 1984). The core idea of teleosemantics is that the possibility of misrepresentation can be explained naturalistically in terms of the possibility of the failure of etiological functions. Although teleosemantics is a theory about the content as opposed to attitude (which we are interested in here) of beliefs, the supporters of the view sometimes blur the distinction between them. For example, Ruth Millikan writes: “[The] content of the representation turns out to be an abstraction from a fuller affair intrinsically involving an imbedding mood or propositional attitude. Put simply, there is no such thing as content without mood or attitude; content is an aspect of attitude” (Millikan 1995: 155). The teleo-functionalist definition of mental states can also be seen in empirical literature. For instance, Randolph Nesse defines emotions as “specialized modes of operation shaped by natural selection to adjust the physiological, psychological, and behavioral parameters of an organism in ways that increase its capacity and tendency to respond adaptively to the threats and opportunities characteristic of specific kinds of situations” (Nesse 1990: 268).

A feature of teleo-functionalism about beliefs is that it allows for the multiple-functionability of beliefs. Teleo-functionalism defines beliefs in terms of doxastic functions and to have doxastic functions is to have a certain evolutionary history. Thus, in effect, teleo-functionalism defines beliefs in terms of their evolutionary history. A consequence of this is that playing belief-roles is not necessary for a mental state to be a belief. Playing particular causal roles is not necessary in order to have a certain evolutionary history. Here is an analogous case. One might define kidneys in terms of their etiological function, namely, filtering metabolic wastes from blood. To have the function of filtering metabolic wastes from blood is, again, to have a certain evolutionary history. Thus, in effect, this definition characterises kidneys in terms of their evolutionary history. A consequence of this is that actual or dispositional performances are irrelevant for an item to be a kidney. This rightly allows for the possibility of malfunctioning kidneys, namely, kidneys that fail to filter metabolic wastes from blood.

“Doxastic functions” is just a placeholder notion. Here we only provide a very rough idea about what doxastic functions are, relying on Frank Ramsey's (1931) famous idea that a belief is a “map of neighbouring space by which we steer.” On the one hand, Ramsey says that a belief is a “map”, which suggests that beliefs, just like maps, are supposed to represent facts more or less accurately. On the other hand, he also says that “we steer by” beliefs, which suggests that beliefs are supposed to guide our actions. According to what we call the “Ramseyan account”, doxastic functions are the functions of (more or less) accurate representing and action guiding. Beliefs, therefore, are the mental

states with the functions of accurate representing and action guiding. The Ramseyan account of doxastic functions is probably too simplistic. It needs to be revised in several ways. For example, there would be some types of beliefs whose evolutionary value has little to do with accuracy (McKay and Dennett 2009). However, we shall simply leave this issue open since the perfect understanding of doxastic functions is not essential for our purpose. What is crucial here is that teleo-functionalism does allow for the multiple-functionability of beliefs, whatever doxastic functions are.

Teleo-functionalism allows for the multiple-functionability of beliefs. Even if delusions fail to play belief-roles, on this view they could be beliefs because nothing yet rules out the possibility that they have doxastic functions. Thus, there is at least an independently motivated theory of belief that does not entail premise (2). But, here is a further question. Do we have reasons to believe that delusions are in fact beliefs according to teleo-functionalism? In other words, do we have reasons to believe that delusions in fact have doxastic functions? This, we believe, is an open empirical issue. In the following, we provide a way in which we can approach this issue with the help of (future) relevant empirical findings.

So far, we have assumed that etiological functions can be attributed to beliefs. The assumption is in fact problematic since etiological functions are attributed only to the items that were selected in the past. In most cases, mental states such as beliefs are not the products of natural selection. As Godfrey-Smith pointed out, “[s]tructural features of the visual apparatus are products of an evolutionary history, a history of heritable variation in fitness. But, states of visual system are not the right sort of things to have such a history” (Godfrey-Smith 1989: 542). The problem can be fixed by attributing functions to mental states in an indirect way. The idea is that all the talk about the functions of mental states is translated into a talk about the functions of the mechanisms that produce and consume those states. In turn, the functions of those mechanisms are defined in terms of their evolutionary history (Millikan 1984). For example, the statement that a mental state has the function of accurate representing is translated into the statement that its producer mechanisms have the function of producing accurate representations. And, this statement is cashed out in terms of the evolutionary history of the mechanisms: the ancestors of the mechanisms were selected for producing accurate representations in the past.

The statement that a mental state has doxastic functions is, thus, translated into a statement about the producers and the consumers of the state having some corresponding functions. Let us call the functions that are attributed to the producer and consumer mechanisms “doxastic producer functions” and “doxastic consumer functions” respectively. For instance, if accurate representing is one of the doxastic functions, then producing accurate representa-

tions is one of the doxastic producer functions. Now, our target hypothesis is that delusions have doxastic functions. The hypothesis is translated into the following: *Delusions are produced by mechanisms with doxastic producer functions and consumed by mechanisms with doxastic consumer functions*. Let us call this “doxastic mechanisms hypothesis” (DMH). What is nice about DMH is that, on the one hand, it is equivalent to our target hypothesis and, on the other hand, we can examine its truth once we know (a) the mechanisms that produce and consume delusional states and (b) the functions of those mechanisms.

According to teleo-functionalism, paradigmatic beliefs, such as my belief that there is a cup of tea on the table, have doxastic functions. (Or, more precisely, “doxastic functions” are fixed by the etiological functions that those paradigmatic beliefs have.) This means that paradigmatic beliefs are produced by mechanisms with doxastic producer functions and consumed by mechanisms with doxastic consumer functions. Then, if we can show that delusions are (a) *produced by exactly the same mechanisms that produce paradigmatic beliefs* and (b) *consumed by exactly the same mechanisms that consume paradigmatic beliefs*, then we in effect show that DMH is correct. Let us call (a) and (b) “same producer hypothesis” (SPH) and “same consumer hypothesis” (SCH) respectively. Please note that, for instance, SPH does not say that there is no difference between the producers of paradigmatic beliefs and those of delusions. It is certainly possible that paradigmatic beliefs are produced by mechanisms that are intact, while delusions are produced by mechanisms that are malfunctioning. SPH does not deny this. It rather says that those mechanisms are nonetheless the same mechanisms for the same reason that healthy kidneys and malfunctioning ones are the same organs. The same thing is true about SCH.

Now, we do not have enough empirical knowledge yet to come to a definite conclusion about these hypotheses. Still, it would be fair to say that SPH is very likely to be true in light of what we already know about the delusion formation process. Arguably, one of the mechanisms that are responsible for the production of paradigmatic beliefs is the one that outputs beliefs in response to perceptual or experiential inputs. The current empirical literature strongly suggests that delusions are also formed by such a mechanism. “Empiricism about delusions” is the view that delusions are formed in response to some kind of abnormal experience. A delusion might arise as the endorsement of the content of abnormal experience (endorsement version of empiricism) or the explanation of it (explanation version of empiricism). Empiricism is well supported by evidence, especially when applied to monothematic delusions. Capgras delusion is probably the most famous example. Hadyn Ellis and Andrew Young (1990) propose the very influential hypothesis that Capgras delusion arises from a deficit in face processing involving loss of affective responses to familiar faces. This deficit is hypothesised as being caused by the

disrupted connection between the face recognition system and the autonomic nervous system. This hypothesis is supported by the finding that people with Capgras delusions have a reduced galvanic skin response to faces and do not show asymmetrical responses between familiar and unfamiliar faces (Ellis et al. 1997).

One might think that empiricism is not very plausible when it comes to typical delusions in the context of schizophrenia such as persecutory delusion or delusion of reference. However, this is probably not the case. Shitij Kapur (2003) proposes an influential empiricist account of delusions in schizophrenia. He argues that the abnormality in dopamine transmission in schizophrenia leads to an inappropriate attribution of salience, where “salience” refers to, roughly, the attention-grabbing quality of events. When salience is attributed inappropriately to events that are not very interesting as a matter of fact, they grab special attention. Such events, all of a sudden, seem to be important and are invested with special meaning. This hypothesis is coherent with the fact that many people with delusions in the context of schizophrenia report that some events have “special meaning” for them, that they have an “altered experience” of the world, that their awareness is “sharpened”, and so on. The delusion, according to Kapur, is produced in response to the experience of abnormal salience: “Delusions in this framework are a “top-down” cognitive explanation that the individual imposes on these experiences of aberrant salience in an effort to make sense of them” (Kapur 2003: 15).

This account has recently turned into a more sophisticated theory which is often called “prediction-error theory” of delusions (Corlett et al. 2010; Fletcher and Frith 2009). The account assumes that the operation of the brain relies in part on the signals that tell us that what we experience does not match our predictions. The prediction-error signals indicate that the internal model of the world from which the predictions are derived is incorrect and needs to be updated. Prediction-error signals play a crucial role in, among others, allocating attention; we allocate attention to the events that defy our expectations. This makes sense intuitively since the predictable events do not bring any new information and, thus, do not deserve attention. According to prediction-error theory, due to prediction-error signalling abnormalities (or abnormal weighting of prediction errors), people with delusions allocate attention to events that do not actually deserve it. This is the process where uninteresting events become inappropriately salient, and delusions are produced as explanatory responses to the inappropriately salient events due to aberrant prediction-error signals. This hypothesis is supported by the study by Phil Corlett and colleagues (2007) which suggests the existence of prediction-error signalling abnormalities in people with delusions due to schizophrenia.

Now, these facts and theories suggest that SPH is true; delusions are produced by the same mechanisms that are responsible for the production of paradigm-

matic beliefs (in response to experiential inputs). Of course, there could be a debate about whether or not that mechanism is working properly. Brendan Maher seems to be arguing that the mechanism is working properly in his famous statement: “a delusion is a hypothesis designed to explain unusual perceptual phenomena and developed through the operation of normal cognitive processes” (Maher 1974: 103). On the other hand, Tony Stone and Andrew Young seem to be sceptical about this when they wrote: “the perceptual deficit account needs to be augmented by a theory of the reasoning biases that lead to the delusional interpretation of the perceptual anomalies produced by the perceptual deficit” (Stone and Young 1997: 341). In any case, the crucial point here is that delusions are produced by exactly the same mechanism (working properly or not) that is responsible for producing paradigmatic beliefs.

#### **4. Conclusion**

The argument from causal role, which is the most influential argument against doxasticism, depends on two premises. According to premise (1), many delusions fail to play belief-roles. According to premise (2), a mental state is a belief only if it plays belief-roles. The former is said to be supported by clinical observations. The latter seems to follow token-functionalist theories of belief. We argued, however, that both premises can in fact be resisted.

Supporters of premise (1) tend to make two mistakes. Firstly, they tend to idealise belief-roles in such a way that “playing belief-roles” is almost synonymous with “being a rational belief.” If we are committed to this idealisation of beliefs, it is not difficult to give good evidence for premise (1), because it can be easily shown that delusions are not rational beliefs. But the problem with idealising beliefs in this way is that many non-delusional beliefs would also fail to play belief-roles and, hence, we would need to accept anti-doxasticism about them too. And, if we are not committed to the idealisation of beliefs, then it is not obvious that clinical observations give good evidence for premise (1). Secondly, anti-doxasticists tend to argue that premise (1) is strongly supported by the fact that delusions sometimes fail to guide action. This argument ignores the possibility that failure of action guidance can be explained by hypotheses that are perfectly compatible with doxasticism. In particular, a coherent and empirically supported explanation of the failure to act on delusions is that the capacity to acquire or preserve motivation is compromised in people with delusions, especially in the context of schizophrenia.

Premise (2) can also be resisted. Firstly, we did not find any good argument for premise (2). The fact that belief is multiply-realizable does not support premise (2) because multiple-realizability is perfectly compatible with multiple-functionability. The fact that we usually identify beliefs by looking at their causes and effects does not support premise (2) because what the nature of



belief is does not need to coincide with how we identify beliefs. Secondly, premise (2) is usually accepted by anti-doxastic philosophers simply because it is entailed by the token-functional theories they accept. However, there might be some alternative theories that do not entail premise (2). Teleo-functionalism, according to which beliefs are the mental states with doxastic functions, seems to be an example. Given that there is at least one theory of belief with some independent motivations that does not entail premise (2), anti-doxasticists cannot assume the truth of premise (2) simply because it is entailed by their favourite theories.

Of course, to show that the argument from causal role can be resisted is not to show that doxasticism is true. Doxasticists have more work to do. Doxasticists who resist premise (1), for instance, need to argue that the compromised motivation hypothesis is not just coherent, but is also more plausible than the alternative hypotheses favoured by anti-doxasticists. Doxasticists who resist premise (2) on the basis of teleo-functionalism need to show that teleo-functionalism supports (and not just permits) doxasticism. To do so, they need to show that delusions have doxastic functions. We described a way in which this could be done by mentioning relevant empirical findings. But we need to know more about the delusion formation process and the delusion consumption process before we can reach a more definite conclusion.

**Acknowledgements:** In the preparation of this paper, Kengo Miyazono acknowledges the support of Japan Society for the Promotion of Science Post-doctoral Fellowship for Research Abroad. Lisa Bortolotti acknowledges the support of an Arts and Humanities Research Council Fellowship (The Epistemic Innocence of Imperfect Cognitions, grant number: AH/K003615/1).

## References

- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders* (DSM 5). Fifth edition. Washington, DC: American Psychiatric Publishing.
- Ames, D. 1984. Self-shooting of a phantom head. *British Journal of Psychiatry*, 145(2): 193-4.
- Bayne, T. & Pacherie, E. 2005. In defence of the doxastic conception of delusions. *Mind & Language*, 20(2): 163-188.
- Bayne, T. 2010. Delusions as doxastic states: Contexts, compartments and commitments. *Philosophy, Psychiatry & Psychology*, 17(4): 329-336.

- Berrios, G. E. 1991. Delusions as "wrong beliefs": A conceptual history. *The British Journal of Psychiatry*, 159 (Suppl. 14): 6-13.
- Bisiach, E. & Geminiani, G. 1991. Anosognosia related to hemiplegia and hemianopia. G. P. Prigatano & D. L. Schacter, eds. *Awareness of deficit after brain injury: Clinical and theoretical issues*: 17-39. New York, NY: Oxford University Press.
- Bleuler, E. 1924. *Textbook of Psychiatry*. Transl. A. A. Brill. New York, NY: Macmillan.
- Bortolotti, L. 2009. *Delusions and other irrational beliefs*. Oxford: Oxford University Press.
- Bortolotti, L. 2010. Double bookkeeping in delusions: Explaining the gap between saying and doing. K. Frankish, A. Buckareff & J. Aguilar, eds. *New waves in the philosophy of action*: 237-256. Basingstoke: Palgrave Macmillan.
- Bortolotti, L. & Broome, M. R. 2012. Affective dimensions of the phenomenon of double bookkeeping in delusions. *Emotion Review*, 4(2): 187-191.
- Browning, S. M. & Jones, S. 1988. Ichthyosis and delusions of lizard invasion. *Acta Psychiatrica Scandinavica*, 78(6): 766-767.
- Corlett, P., Murray, G., Honey, G., Aitken, M., Shanks, D., Robbins, T., et al. 2007. Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain*, 130(9): 2387-2400.
- Coltheart, M. 2007. Cognitive neuropsychiatry and delusional belief: The 33rd Sir Frederick Bartlett lecture. *The Quarterly Journal of Experimental Psychology*, 60(8): 1041-1062.
- Corlett, P., Taylor, J., Wang, X., Fletcher, P. & Krystal, J. 2010. Toward a neurobiology of delusions. *Progress in Neurobiology*, 92(3): 345-369.
- Cummins, R. 1975. Functional analysis. *Journal of Philosophy*, 72(20): 741-765.
- Currie, G. 2000. Imagination, delusion and hallucinations. M. Coltheart & M. Davies, eds. *Pathologies of belief*: 167-182. Oxford: Blackwell.
- Currie, G. & Jureidini, J. 2001. Delusions, rationality, empathy: Commentary on Davies et al. *Philosophy, Psychiatry & Psychology*, 8(2-3): 159-162.
- Dretske, F. 1986. Misrepresentation. R. Bogdan, ed. *Belief: Form, Content and Function*: 17-36, Oxford: Oxford University Press.
- Egan, A. 2009. Imagination, delusion, and self-deception. T. Bayne, & J. Fernandez, eds. *Delusions and self-deception: Affective and motivational influences on belief formation*: 263-280. Hove: Psychology Press.
- Ellis, H. D. & Young, A. W. 1990. Accounting for delusional misidentifications. *The British Journal of Psychiatry*, 157(2): 239-248.
- Ellis, H. D., Young, A. W., Quayle, A. H. & De Pauw, K. W. 1997. Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1384): 1085-1092.

- Fletcher, P. C. & Frith, C. D. 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1): 48-58.
- Frankish, K. 2009. Delusions: A two-level framework. M. R. Broome & L. Bortolotti, eds. *Psychiatry as cognitive neuroscience: Philosophical perspectives*: 269-284. Oxford: Oxford University Press.
- Freeman, D. & Garety, P. A. 2003. Connecting neurosis and psychosis: The direct influence of emotion on delusions and hallucinations. *Behaviour Research and Therapy*, 41(8): 923-947.
- Freeman, D. Garety, P. A. & Kuipers, E. 2001. Persecutory delusions: Developing the understanding of belief maintenance and emotional distress. *Psychological Medicine*, 31(7): 1293-1306.
- Frith, C 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hove: Psychology Press.
- Gallagher, S. 2009. Delusional realities. M. R. Broome & L. Bortolotti, eds. *Psychiatry as cognitive neuroscience: Philosophical perspectives*: 245-266. Oxford: Oxford University Press.
- Gard, D. E., Kring, A. M., Gard, M. G., Horan, W. P. & Green, M. F. 2007. Anhedonia in schizophrenia: Distinctions between anticipatory and consummatory pleasure. *Schizophrenia Research*, 93: 253-260.
- Godfrey-Smith, P. 1989. Misinformation. *Canadian Journal of Philosophy*, 19(4): 533-550.
- Godfrey-Smith, P. 1998. *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
- Hohwy, J. & Rajan, V. 2012. Delusions as forensically disturbing perceptual inferences. *Neuroethics*, 5(1): 5-11.
- Hohwy, J. & Rosenberg, R. 2005. Unusual experiences, reality testing and delusions of alien control. *Mind & Language*, 20(2): 141-162.
- Hume, D. 1978. *A Treatise of Human Nature*. ed. A. Selby-Bigge. Sccond edition. Revised by P. H. Nidditch. Oxford: Clarendon Press.
- Kapur, S. 2003. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry*, 160(1): 13-23.
- Kring, A. M. & Elis, O. 2013. Emotion deficits in people with schizophrenia. *Annual Review of Clinical Psychology*, 9: 409-433.
- Lewis, D. 1983. Mad pain and Martian pain. *Philosophical Papers*, vol. I: 122-130. Oxford: Oxford University Press.
- Lucchelli, F. & Spinnler, H. 2007. The case of lost Wilma: A clinical report of Capgras delusion. *Neurological Sciences*, 28(4): 188-195.
- Lycan, W. G. 1987. *Consciousness*. Cambridge, MA: MIT Press

- Maher, B. A. 1974. Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30: 98-113.
- McKay, R. & Cipolotti, L. 2007. Attributional style in a case of Cotard delusion. *Consciousness and Cognition*, 16(2): 349-59.
- McKay, R. T. & Dennett, D. C. 2009. The evolution of misbelief. *Behavioral and Brain Sciences*, 32(6): 493-561.
- Millikan, R. G. 1984. *Language, Thought and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
- Millikan, R. G. 1995. Pushmi-pullyu representations. *Philosophical Perspectives*, 9: 185-200.
- Neander, K. L. 1991. Functions as selected effects: The conceptual analyst's defence. *Philosophy of Science*, 58(2): 168-184.
- Nesse, R. M. 1990. Evolutionary explanations of emotions. *Human Nature*, 1(3): 261-289.
- Payne, R. (1992). My Schizophrenia. *Schizophrenia Bulletin* 18(4): 725-728.
- Ramsey, F. P. 1931. *The Foundations of Mathematics and Other Logical Essays*. London: Routledge.
- Ryan, R. M. & Deci, E. L. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1): 68-78.
- Roberts, G. 1992. The Origins of Delusion. *The British Journal of Psychiatry*, 161: 298-308.
- Sass, L. A. 2001. Self and world in schizophrenia: Three classic approaches. *Philosophy, Psychiatry, & Psychology*, 8(4): 251-270.
- Schwitzgebel, E. 2012. Mad belief? *Neuroethics*, 5(1): 13-17.
- Siris, S. G. 1995. Depression and schizophrenia. S. R. Hirsch & D. Weinberger, eds. *Schizophrenia*: 128-145. Oxford: Blackwell.
- Silva, J. A., Leong, G. B., Weinstock, R., Sharma, K. K. & Klein, R. L. 1994. Delusional misidentification syndromes and dangerousness. *Psychopathology*, 27(3-5): 215-219.
- Sterelny, K. 1990. *The Representational Theory of Mind*. Oxford: Blackwell.
- Stone, T. & Young, A. W. 1997. Delusions and brain injury: The philosophy and psychology of belief. *Mind & Language*, 12(3-4): 327-364.
- Sober, E. 1985. Panglossian functionalism and the philosophy of mind. *Synthese*, 64(2): 165-193.
- Tumulty, M. 2012. Delusions and not-quite-beliefs. *Neuroethics*, 5(1): 29-37.

- Weinstein, E. A. 1996. Reduplicative misidentification syndromes. P. W. Halligan & J. C. Marshall, eds. *Method in madness: Case studies in Cognitive Neuropsychiatry*: 13–36. Hove: Psychology Press.
- Wessely, S., Buchanan, A., Reed, A., Cutting, J., Everitt, B., Garety, P. & Taylor, P. 1993. Acting on Delusions (I): Prevalence. *British Journal of Psychiatry*, 163(1): 69-76.
- Wilson, T. 2002. *Strangers to ourselves*. Cambridge, MA: Harvard University Press.