



In Defence of the Doxastic Conception of Delusions

Elisabeth Pacherie, Tim Bayne

► To cite this version:

Elisabeth Pacherie, Tim Bayne. In Defence of the Doxastic Conception of Delusions. *Mind & Language*, 2005, 20 (2), pp.163-188. ijn_00353586

HAL Id: ijn_00353586

https://hal.science/ijn_00353586

Submitted on 15 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

In Defence of the Doxastic Conception of Delusions

TIM BAYNE AND ELISABETH PACHERIE

Abstract: In this paper we defend the doxastic conception of delusions against the metacognitive account developed by Greg Currie and collaborators. According to the metacognitive model, delusions are imaginings that are misidentified by their subjects as beliefs: the Capgras patient, for instance, does not believe that his wife has been replaced by a robot, instead, he merely imagines that she has, and mistakes this imagining for a belief. We argue that the metacognitive account is untenable, and that the traditional conception of delusions as beliefs should be retained.

1. Introduction

Delusions have traditionally been characterized as beliefs, as in the DSM definition:

Delusion: 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, p. 821).

Such a conception is open to a number of objections:

No-content

Some delusions appear either obviously false or incoherent. One patient claimed that his mother changed into another person every time she put her glasses on

This paper has been presented to audiences at the 2003 ESPP meeting in Turin, the Joint International Cognitive Science Conference (Sydney, July 2003), and the Philosophy departments of the University of Sydney and the Australian National University (RSSH). We also wish to thank David Chalmers, Max Coltheart, Martin Davies, Anne Aimola Davies, Jérôme Dokic, Andy Egan, Pascal Engel, Melissa Green, Pierre Jacob, Ryan McKay, David Braddon-Mitchell, Joëlle Proust, Tony Stone, and Caroline West for their comments on this paper. Thanks also to Robyn Langdon and an anonymous referee at *Mind & Language* for their many helpful comments and suggestions. This paper was supported by the SCMP research fund (Macquarie University).

Address for correspondence: Department of Philosophy, Macquarie University, North Ryde, Sydney NSW 2109 AUSTRALIA

Email: tbayne@scmp.mq.edu.au

Address for correspondence: Institut Jean-Nicod, CNRS-EHESS-ENS, 1 bis, avenue de Lowendal, 75007 Paris FRANCE

Email: pacherie@ehess.fr

(De Pauw and Szulecka, 1988); another had the delusion that there was a nuclear power station inside his body (David, 1990); a third had the delusion of being in both Boston and Paris at once (Weinstein and Kahn, 1955).

Pragmatically self-defeating

Other delusions express propositions that seem to be believable, but it is difficult to see how the delusional patient themselves could believe them. In extreme forms of the Cotard delusion, patients say that they are dead. This claim seems to be pragmatically self-defeating (Bermúdez, 2001).

Lack of evidence

Deluded patients appear to lack reasons or evidence for their delusional state. A patient looked at a row of empty marble tables in a café and became convinced that the world was coming to an end. As Campbell (2001) points out, it is difficult to understand how an experience of marble tables could verify the proposition ‘The world is ending’. The lack of evidence for delusions appears to be a problem for the doxastic account if—as many hold—there is a constitutive relationship between belief and evidence.

Theoretical reasoning

A subject will normally accept the obvious logical implications of her beliefs—at least when these are pointed out to her. And when she realizes that some of her beliefs are inconsistent, she will normally engage in a process of revision to restore consistency. In contrast, deluded patients often fail to draw the obvious logical consequences of their delusions and show little interest in resolving apparent contradictions between their delusion and the rest of their beliefs. Breen *et al.* (2000) describe the case of patient DB, who suffered from reduplicative paramnesia after a right parietal stroke. DB both affirmed (correctly) that her husband had died and been cremated four years before and also (incorrectly) that he was a patient on the ward in the same hospital that she was in. DB did not seem upset by the inconsistency between her claims about her husband and did nothing to resolve it. Brett-Jones *et al.* (1987) cite the case of a patient who was happy to converse with his psychologist despite insisting that she was deaf and dumb.

Practical reasoning

It is said that delusional patients fail to act in ways that they ought to act if they believed their delusion (Sass, 1994). ‘[Delusional] patients rarely follow up the logic

to act accordingly, as, for instance, to bark like a dog when they profess to be a dog. Although they may refuse to admit the truth, they behave as if the expression is only to be taken symbolically' (Bleuler, 1924).

Lack of appropriate affect

Delusional patients often fail to exhibit the affective responses that would be appropriate to their delusion if they were beliefs. Capgras patients are often unmoved by the fate of their relatives whom, according to the doxastic interpretation of this delusion, they believe to have been abducted.

These last four objections are aspects of a more general encapsulation objection: delusions lack the kind of holistic character that beliefs are supposed to have: they do not interact with perceptual input, other cognitive states or behaviour in the way beliefs should.

Although not all of these objections count equally against all kinds of delusions—the pragmatically self-defeating objection, for example, applies only to the Cotard delusion—their combined force constitutes a powerful challenge for the doxastic account of delusion.

2. The Metacognitive Model of Delusions

In the face of the difficulties confronting the doxastic account, a number of rival approaches to delusions have been suggested. For example, Jaspers (1963) and Berrios (1991) have suggested that delusions are not contentful states. Campbell (2001) has suggested that delusions are beliefs, but they don't appear to have the content they seem to have.¹ In this paper we leave these accounts to one side, and examine an account of delusions developed by Currie and collaborators (Currie, 2000; Currie and Jureidini, 2001; Currie and Ravenscroft, 2002).² Currie and co-authors argue that delusions—or at least an important range of delusions—are cognitive hallucinations: they are imaginative states that are misidentified by their subjects as beliefs. As Currie puts it, 'what we normally describe as the delusional belief that P ought sometimes to be described as the delusional belief that I believe that P' (Currie, 2000, p. 175). Because Currie's account involves reference to deficits in the monitoring of mental states we will call it the *metacognitive model*.

We take the metacognitive model to involve the following three claims (where P is the content of the delusional state):

- (i) Delusional patients who seem to believe P do not actually believe P;

¹ We discuss Campbell's model of delusions in Bayne and Pacherie (2004a).

² Stephens and Graham (2004, forthcoming) have developed an account of delusions that bears important similarities to the metacognitive account.

- (ii) Delusional patients who seem to believe P actually imagine P;
- (iii) Delusional patients who seem to believe P believe that they believe P.

Because these three claims are logically independent, the task of evaluating Currie's account is rather complicated: each claim has to be assessed on its merits. In section 2.1 and 2.2 we will focus on (i) and (ii), while in section 2.3 we turn our attention to (iii).

The metacognitive model promises to account for those features of delusions that put pressure on the doxastic account. It can account for the pragmatically self-defeating objection, for there is nothing incoherent about imagining a world in which one does not exist. It can account for the fact that delusions seem to be generated on the basis of little or no evidence, for this is also true of imaginings. It can account for the fact that deluded patients appear to have beliefs that are inconsistent with their delusions, for there is nothing unusual about having imaginings that are inconsistent with one's beliefs; indeed, this would appear to be the norm. And it can account for the fact that delusions typically fail to issue in direct actions or strong affective responses, for this is also true of imaginings. Finally, the metacognitive model promises to explain the patient's verbal behaviour: the patient says P because she believes that she believes P. The metacognitive account does not have a straightforward response to the no-content objection—arguably one can no more imagine an incoherent proposition than believe it—but the no-content objection is perhaps the weakest of the objections to the doxastic account that we surveyed.

One point that is somewhat unclear is the intended scope of the metacognitive account. Currie and co-authors apply it to the florid and polythematic delusions typical of schizophrenia, but they are less explicit about whether or not it applies to monothematic delusions (such as the Capgras delusion). In some places Currie implies that the model should be extended to include monothematic delusions (Currie and Jureidini, 2001), while in other places he suggests that a different account of monothematic delusions might be appropriate (Currie, 2000).

We will take the metacognitive account as an account of delusions in general. Our primary reason for taking this approach is that the arguments Currie and collaborators give for thinking that delusions are cognitive hallucinations apply not just to polythematic delusions but to delusions fairly generally. In addition, taking the debate between the two approaches to concern the analysis of delusions in general also means that it is of more interest than it would be were we to restrict its domain to polythematic delusions.

2.1. Beliefs, Imaginings, and Delusions

What distinguishes imaginings from beliefs? This is a surprisingly difficult question. We can start by following Currie and Ravenscroft (2002) in distinguishing various kinds of imaginings. Some states of imaginings have imagistic (quasi-perceptual) content; we can call such imaginings *perceptual imaginings*. The content of the state

of imagining looking at one's mother resembles the content of actually looking at one's mother. Closely akin to (and perhaps a species of) perceptual imagining is motor imagining; what it is like to imagine raising one's hand resembles what it is like to experience raising one's hand.

Although perceptual and motor imaginings are not unrelated to the metacognitive account (see 2.2), they are not at its heart. Rather, the metacognitive account centers on *propositional imagining*: such states as imagining that the first person on the moon was a woman. Propositional imaginings need not be accompanied by any form of quasi-pictorial imagery, and even when they are, this imagery does not fix their representational content.

Although Currie and collaborators distinguish propositional imaginings from perceptual imaginings, they overlook certain important distinctions between different forms of propositional imagining. In one sense, to imagine a proposition is simply to entertain it, independently of any attitude to the truth-value of the proposition in question. One might imagine that the population of Nepal has doubled in the last 25 years without having any attitude towards the truth-value of this thought. Call this *simple imagination*. One can also entertain in imagination a scenario in which a proposition holds while believing of this scenario that it is not actual. This is the sense of imagination at work in our understanding of fiction. Call this *counterfactual imagination*. There is yet a third form of propositional imagining, according to which to imagine P is to have some inclination—however slight—to think that P is the case. Call this *indicative imagination*. It is this sense of imagination that is at work in such claims as, 'I can't imagine that the Mafia killed JFK'. Someone who says this is not meaning to deny that she can *entertain* the proposition <the Mafia killed JFK>, rather, she means to communicate the thought that it is extremely unlikely that the Mafia killed JFK. Indicative imagination and belief seem to be on a continuum, insofar as both are attitudes to the way the world actually is; simple and counterfactual imagination, by contrast, are not on a continuum with belief.

Which of these three notions of imagination do Currie and co-authors have in mind when they say that delusions are imaginings? Although they appear to slide between these various notions of imagination, we suspect that Currie and co-authors *usually* mean what we are calling 'simple imagination'. The 'counterfactual imagination' reading of their position seems implausible, for to imagine something counterfactually involves taking the contents of one's imagining as concerning a counterfactual world, and that is in tension with the thought that delusional patients believe that they believe their imaginings. And reading their position in terms of indicative imagination is at odds with their insistence that the onset of an imaginative state leaves the person's previous beliefs intact:

... when we imagine something, we do not cease to believe things that are inconsistent with what we imagine, and we do not feel any pressure to resolve clashes between what we believe and what we imagine (Currie and Jureidini, 2001, p. 160, our emphasis).

These comments suggest that, according to the metacognitive model, the Capgras patient imagines that his wife has been replaced by an imposter but continues to believe that she has not been replaced.³ So that leaves the 'simple imagination' reading of the metacognitive account as the most plausible.

What distinguishes belief from simple imagination? (Hereafter we will drop the modifier 'simple'.) There are two things we need here: (i) an account of what it is in virtue of which believing that P and imagining that P have the same content (namely P); and, (ii) an account of what makes the former state a belief and the latter an imagining.

According to Currie, imaginings and beliefs have the same content when their inferential roles mirror each other. Content is determined by (or at least tracks) inferential relations, and since the belief that P and the imagination that P license the same types of inferential moves, they share the same kinds of content. This mirroring of inferential roles extends to practical reasoning. The pattern of practical inferences supported by imagining that P mirrors that supported by believing that P. Finally, imaginings have affective and emotional consequences similar in content (but perhaps weaker in strength) to those that the corresponding beliefs possess.

How then do beliefs and imaginings *differ*? One might have expected Currie and co-authors to appeal to the sense of subjective certainty (or conviction) at this point. Arguably, to believe P is to think that P is (or is likely to be) true when one reflects on it, whereas no such sense of truth accompanies the state of merely imagining P. But although Currie and Ravenscroft grant that 'under normal conditions, P's being a state that carries conviction for its subject can be an almost infallible sign that P is a belief of the subject' (2002, p. 176), they reject the view that beliefs can be defined in terms of the sense of conviction. Instead, they hold that beliefs and imaginings are distinguished in the following three ways:

- (i) the relations that they have to perceptual states (input);
- (ii) the consistency constraints that they are subject to;
- (iii) the relations they have to action (output).

Roughly, beliefs are generated by a person's environment and by other beliefs, whereas imaginings are generated autonomously (beliefs are not actions, imaginings are); beliefs ought to be consistent with other beliefs, whereas imaginings need not be consistent with beliefs; and beliefs are action-guiding in ways in which imaginings are not.

It is not clear how Currie and co-authors take these three criteria to be related. It seems possible for a state to qualify as an imagining according to one of these three criteria (say, the first), and yet qualify as a belief according to one of the other criteria (say, the third). Would Currie and co-authors regard such a state as an imagining or a belief? It is quite unclear. We suspect that they incline to the view

³ We thank John Campbell for this point.

that action-guidance and consistency with other beliefs are jointly necessary and sufficient for belief, but even this position fails to clarify what we should say when action-guidance and consistency come apart. Although these issues are theoretically important, we need not resolve them here, for we will argue that delusions are better thought of as beliefs according to each of these three criteria.

2.1.1. Perception, Belief, and Imagination. Currie and Ravenscroft suggest that imaginings and beliefs are generated in very different ways. Sometimes they say that imaginings are actions—they are doings—whereas beliefs are not actions, they are things that happen to us (Currie and Ravenscroft, 2002, p. 25). In other places they make the weaker claim that imaginings are autonomously generated, whereas beliefs are formed ‘in response to perceptual information, or by inference from other beliefs we already have’ (Currie and Ravenscroft, 2002, p. 173).

We think that both ways of drawing this contrast are problematic. Although we agree that imaginings are often actions, we are not convinced that the distinction between imagination and belief lines up with the distinction between what we do and what happens to us. As Currie and Ravenscroft themselves note, imagination is not always under the control of the will; we sometimes cannot stop ourselves from imagining certain things. And, conversely, there is something to be said for the claim that forming (and maintaining) a belief can be an action, as when we deliberately apply ourselves to exploring the consequences of our beliefs and form new beliefs as a result.

Even the weaker (and more plausible) thesis that the distinction between imagination and belief lines up with the distinction between endogenous and exogenous states is problematic. Self-deception and wishful thinking demonstrate that beliefs can be generated and maintained by the emotional functions they serve (Mele, 1993). And it seems entirely conceivable that certain beliefs (and not just concepts) might be innate. If these non-evidential routes to belief are possible, there seems little reason to rule out the possibility of belief being formed as the direct result of brain damage.⁴ Beliefs are *typically* formed in response to perceptual information or by inference from other beliefs, but we see no reason to think that these are the only routes to belief. (We should also note that beliefs about one’s own mental states are endogenous, but that does not prevent them from being beliefs!)

Even if belief were constitutively tied to perception, there is little reason to think that this would support the metacognitive account of delusions. Building on Maher’s work (Maher, 1974; Maher, 1999), many theorists have argued that unusual experiences play an important role in triggering monothematic delusions (Ellis and Young, 1990; Langdon and Coltheart, 2000; Davies, Coltheart, Langdon

⁴ Although we think that top-down accounts of delusions are *coherent*, we do not think that they are *plausible* (see Bayne and Pacherie (2004a)).

and Breen, 2001; Ramachandran and Blakeslee, 1998; Gerrans, 2002). Experience-based accounts of delusion appear to be entirely consistent with—indeed, even confirmatory of—doxastic accounts.

Currie and collaborators are not convinced. They claim, rather, that experiential models of delusion-formation are grist for the *metacognitive* mill:

It is characteristic of imagining to be much more easily triggered by perception than is belief. We might readily imagine that a stranger's gaunt appearance signals that he has AIDS but be rightly very resistant to believing it without further evidence. Indeed, if imagination were not easily triggered by mere appearances, pictures, plays, and movies would have very little appeal for us (Currie and Jureidini, 2001, p. 159f).

They go on to say that a 'rational agent with an odd experience of faces, which we think Capgras patients have, might well be prompted to imagine that his loved ones had been replaced by imposters' (Currie and Jureidini, 2001, p. 160).

We are puzzled by these claims. How could it be true both that imaginings are autonomously (voluntarily) generated *and* that they are more easily triggered by perception than beliefs are?

We suspect that Currie and co-authors are equivocating between different forms of imagination. When Currie and Jureidini say that imaginings are autonomously generated they seem to have simple and/or counterfactual imagination in mind. And when they say that 'it is characteristic of imagining to be much more easily triggered by perception than is belief' they appear to have indicative imagination in mind. It is clearly indicative imagination rather than simple or counterfactual imagination that is at work in the example of the AIDS patient. One does not just entertain or conceive a scenario in which this stranger has AIDS—rather, one experiences some temptation to think that the stranger actually has AIDS. One is drawn to it, even if, as Currie and Jureidini say, one is resistant to believing it. Of course, it is true that certain appearances can trigger counterfactual imagination, but they typically do so only within certain contexts: contexts in which one takes up the counterfactual attitude towards certain events. Indicative imagination, by contrast, is routinely triggered by perception. One looks at a colleague's face and imagines that they are anxious; one looks at the clouds and imagines that it will rain; one reads the news and imagines that hostilities will soon commence.

But to say that indicative imagination is easily triggered by perception is not obviously inconsistent with the doxastic account, for indicative imagination, unlike counterfactual imagination, is continuous with belief. To imagine P in the indicative sense is to think that P might actually be true, it is to take P as a contender for the truth. Someone who leaps to conclusions—someone who fails to evaluate a hypothesis in a careful and considered manner—might easily proceed from regarding P as a contender for the truth to thinking that P is true.

2.1.2. Consistency, Belief, and Imagination. We turn now to the second of the three criteria for distinguishing belief from imagination that Currie and co-authors advance: the consistency criterion.

Beliefs are supposed to obey the constraints of inferential rationality long-term and globally; ideally, all our beliefs are always consistent. The most we ask of imaginings is that they carve out a consistent subset of our beliefs with which to connect in the short term (Currie and Ravenscroft, 2002, p. 16).

Call this constraint—that applies to beliefs but not to imaginings—the *consistency constraint*. Clearly, Currie and Ravenscroft see it as definitional of belief:

If someone says that he has discovered a kind of belief that is peculiar in that there is no obligation to resolve or even to be concerned about inconsistencies between these beliefs and beliefs of any other kind, then the correct response to him is to say that he is talking about something other than belief (Currie and Ravenscroft, 2002, p. 176).

And they see the consistency constraint as providing another reason for thinking that delusions are imaginings rather than beliefs.

Imaginings seem just the right things to play the role of delusional thoughts; it is of their nature to coexist with the beliefs they contradict, to leave their possessors unwilling to resolve the inconsistency, and to be immune to conventional appeals to reason and evidence (Currie and Jureidini, 2001, p. 160).

A number of distinctions will prove helpful in assessing this argument. First, we can distinguish between a *normative* and a *constitutive* conception of the consistency constraint. Clearly some form of consistency constraint applies to belief in the sense that one ought to want to have consistent beliefs. This is a *normative* conception of the consistency constraint: epistemic agents ought to take consistency into account in revising their beliefs. A second conception of the consistency constraint is *constitutive*: roughly, one cannot believe P if one believes things that are obviously inconsistent with P (or, perhaps, if one is indifferent to the question of whether P is consistent with the rest of one's beliefs). Now, if the consistency constraint is understood normatively, it can tell us only how beliefs and imaginings *ought* to differ; it will not necessarily help us decide whether someone believes P or merely imagines P. Our cognitive behaviour does not always meet the normative standards that it ought to. Since the present question is to determine whether deluded patients believe P or only imagine P, it is the constitutive reading that is relevant. But, as a constitutive point, Currie's claim about consistency appears to be implausibly strong. Self-deception and the partial encapsulation of belief are commonplace phenomena. As social psychologists (and those we live with) never

tire of telling us, most of us fall short of satisfying or even approximating the normative requirements of global and long-term consistency.

A second distinction that is relevant here is that between inconsistency and the awareness of inconsistency. It is one thing for a delusional state to *be* inconsistent with other things that the subject believes, and it is quite another thing for the deluded patient to *realize* (or believe) that the delusional state is inconsistent with other beliefs she holds. Currie seems to suggest that what is particularly troubling for the doxastic account is not that it attributes inconsistent states to delusional patients, but that such patients are indifferent to the alleged inconsistency when confronted with it (Currie, 2002, p. 160).

Currie's argument, we take it, is this:

- (1) Delusional patients often acknowledge that their delusions are inconsistent with other things that they believe.
- (2) Delusional patients are indifferent to acknowledged inconsistencies between their delusions and the rest of their beliefs.
- (3) This indifference can best be explained by reference to the fact that the delusional state is an imagining rather than a belief.

Currie and Jureidini motivate (1) by pointing out that deluded subjects often acknowledge that their delusional claims are 'implausible, inexplicable, and unlikely to be believed by anyone else' (Currie and Jureidini, 2002, p. 161). But there are a number of things that patients could mean by such admissions, and it is not obvious that any of them supports (1). There is nothing *per se* strange about having a belief that one thinks is unlikely to be believed by anyone else if one also thinks that one has evidence that is not available to, or is unlikely to be taken seriously by, other people. After all, philosophers often believe things that they think others are unlikely to believe. Thinking that one's beliefs are implausible, inexplicable, and unlikely to be believed by anyone else does not demonstrate that one regards them as inconsistent with one's other beliefs.

Nonetheless, there is some evidence that many delusional patients are aware, if only implicitly, of tensions between their delusion and other beliefs they hold.⁵ Delusional patients often revise their background beliefs when confronted by tensions between them and the content of their delusion(s). Consider the following well-known report of a conversation between JK, a Cotard patient, and Young and Leafhead.

We asked her during the period in which she claimed to be dead whether she could feel her heart beat, whether she could feel hot or cold and whether she could feel whether her bladder was full. JK said that since she had such

⁵ Delusional patients with polythematic delusions often fail to acknowledge that their delusions are inconsistent with other things that they believe, but in such cases it is less clear that the distinction between their delusional beliefs and their non-delusional beliefs can be sustained.

feelings even though she was dead they clearly did not represent evidence that she was alive (Young and Leafhead, 1996, p. 158).

JK is obviously *not* indifferent to the tension between her delusional state and her background beliefs. Other deluded patients exhibit discomfort and attempt to change the topic of conversation when confronted with inconsistencies between their background beliefs and their delusion (see e.g. Halligan and Marshall, 1995). This sort of behaviour is also evidence that the delusion is playing the role of a belief—if it were an imagining the delusional patient would not be troubled by tensions between it and her background beliefs.

Another challenge to the consistency constraint concerns how it is meant to apply to unconscious beliefs. Currie and co-authors grant that unconscious states can qualify as beliefs on the grounds that they can be action-guiding, but it is at best unclear how states that are not easily accessible to consciousness—if accessible at all—could be subject to constraints of consistency.

2.1.3. Action, Belief, and Imagination. Imagination tends to have less effect on behaviour than belief. If, for example Capgras patients are only imagining that their relatives have been replaced by imposters, this would explain why they so often fail to act on the basis of their delusions.

This is an intriguing suggestion, but it faces problems. An obvious problem concerns the patient's verbal behaviour: if the delusional patient only imagines P why then does she *say* P? (This problem is especially acute if, as seems likely on the metacognitive account, the onset of the delusion leaves the patient's previous beliefs intact, for in many cases the patient will believe not-P.) We return to this problem below (see section 2.3).

But it is not only the patient's verbal behaviour that creates problems for Currie's account. As a number of authors have pointed out, delusions also generate non-verbal behaviour (see Coltheart and Davies, 2000; Young, 1999). A recent view of 260 Capgras cases found delusion-related violence in 18% of the cases that they examined (Förstl *et al.*, 1991). Young and Leafhead (1996) note that all of Cotard's patients showed some form of delusion-related behaviour (e.g. refusing to move, eat, or defecate). O'Dwyer reports that patients with erotomania usually act on their delusion (O'Dwyer, 1990). Wessely *et al.* (1993) found that 77% of their group of fifty-nine subjects acted on their delusional belief in the month before admission. If the doxastic account faces a problem of *inactivity*, the metacognitive model faces a problem of *over-activity*.

Currie is aware of this objection, and suggests that the doxastic account might apply to those cases in which patients act on their delusions (Currie, 2000, p. 176). He goes on to suggest that believing that one believes P might bring it about that one actually believes P. There is certainly room for hybrid models of this kind, but there is a real worry that they fail to get to the heart of the issues. The problem is that there are *particular* patients of whom there is reason to think that they believe P and reason to think that they do not believe P (or even believe not-P). This

problem is not addressed by holding that some deluded patients believe *P* and others merely imagine *P*. We return to this issue below (see section 3).

Let us summarize the results of section 2.1. We have examined each of the three criteria Currie and co-authors advance as grounds for distinguishing beliefs from imaginings. None of these three criteria supports the meta-cognitive account over the doxastic alternative: either the criterion in question is itself problematic (as is the case with (2.1.1) and (2.1.2)) or the criterion fails to support the metacognitive view over the doxastic account (as is the case with (2.1.3)).

2.2. How Do We Distinguish Imaginings From Beliefs?

Why might delusional patients misidentify their imaginings as beliefs? In answering this question it is useful to ask how we succeed in identifying our (current) imaginings and beliefs as such. Currie and co-authors follow Wittgenstein in holding that we identify our imaginings as imaginings on the basis of a sense of agency:

Suppose that . . . our primary ground for identifying something as an imagined thought is our identification of it as the outcome of will or agency. And suppose that imaginings are significantly similar to beliefs in their inferential role. And suppose, finally, that Frith is right that schizophrenia involves a loss of the sense of agency. We have here the makings of an explanation of why someone with schizophrenia would take an imagined thought to be a belief (Currie, 2000, p. 176).

This account rests on three assumptions:

- (1) Imaginings are typically active (and beliefs are typically passive).
- (2) We identify our imaginings as imaginings on the basis of the sense of agency we experience towards them (and we identify beliefs as beliefs on the grounds that they are unaccompanied by the sense of agency).
- (3) Delusional patients misidentify their imaginings as beliefs because of an impairment in the experience of agency towards them.

We regard all three assumptions as problematic.

We have already cast doubt on the first assumption. Although counterfactual imagination is often consciously self-directed—as when we deliberately imagine flying pigs—, imagination can also be passive (spontaneous, involuntary). We can suddenly realize that we have been daydreaming. Yet the fact that such daydreams are not accompanied by a sense of agency does not prevent us from identifying them as daydreams. Similarly, imagination is easily triggered by perception (as Currie and Jureidini point out): the shape of the clouds can prompt one to imagine that they are monsters, or animals, or gods. But the fact that an act of imagination is stimulus-driven rather than self-generated does not prevent us from identifying it as an act of imagination.

Second, the absence of a sense of agency would not account for the fact that we experience a sense of conviction when introspecting our beliefs. Why wouldn't the awareness of the content P when unaccompanied by the feeling of doing generate simply an awareness of the content P, rather than the thought that one believes P? (Consider stray thoughts that occur to one without the sense of doing.) It is difficult to see how a sense of truth might result from the absence of a sense of agency.

The third assumption in the account put forward by Currie and Ravenscroft (2002) derives its inspiration from an hypothesis about schizophrenia developed by Chris Frith. The hypothesis is that schizophrenic subjects suffer from impairment in efference-copying processes. When a motor instruction is created preparatory to movement, a copy of the instruction (the efference copy) is also created to be compared with reafferences from the movement. The efference copy is then fed to an internal forward model that uses it to predict the sensory consequences of executing the motor command. This serves two purposes. First, a comparison of the predicted outcome with the sensory reafferences will yield a sense of agency when there is a match. Second, the output of the forward model is compared with the intended outcome and thus functions as internally generated feedback, allowing for a faster online correction of errors than would be possible with peripheral feedback. Frith's proposal is therefore that an impairment in efference-copying processes in schizophrenia would lead to deficits in action-monitoring and intention-monitoring. As a result of this impairment a subject may perform an action and yet have no sense of having acted.

Currie and Ravenscroft go on to suggest that an impairment in efference-copying may also explain why schizophrenic patients would mistake their imaginings for beliefs. In making this suggestion they appeal to the idea that imagination draws extensively on systems initially designed for motor control. Although internal forward models, which they call emulators, were initially designed for motor control, they have been put to new uses through the evolutionary history of cognitive systems. In particular, they can be used offline to simulate the effects of carrying out a motor command without the command actually being executed. They function as a planning tool allowing us to imagine the effects of a hypothetical movement or action. Currie and co-authors suggest that if imagination exploits the capacities of systems designed for motor control, a breakdown in the monitoring of the motor system could lead to similar difficulties in the monitoring of imagination-based thoughts.

We see several problems with Currie and Ravenscroft's third assumption. First, even assuming that an impairment of the kind postulated by Frith may explain why some imaginings are not identified as such by schizophrenics, we do not think that it would vindicate the metacognitive account. An impairment in efference-copying mechanisms may possibly explain why internally generated sensory expectations are mistaken for actual sensory input—and thus explain why hallucinations in various perceptual modalities are taken for genuine perceptual experiences—but imaginings mistaken for perceptions would be *perceptual* imaginings rather than

propositional imaginings, and it is unclear how such an account could be extended to explain the genesis of propositional imaginings. Following Campbell (1999), Currie and Ravenscroft point out that there are problems with the idea that thinking is a motor process.⁶ Motor processes seem to be dependent on thoughts rather than the reverse. If we were to consider thinking as a kind of action then we would seem to be committed to the idea that there can be intentions to think and a vicious infinite regress would threaten, for prior intentions are themselves thoughts, which would demand prior intentions of their own (and so on). Currie and Ravenscroft think that one might avoid these difficulties by appealing to emulators that are recruited for new purposes. Yet, if the emulators were initially recruited for motor control, it is difficult to see how they could yield propositional outputs. If, on the other hand, we extend the notion of emulators and simply conceive of them as internal models of reality, quite independently of a possible use for motor control, we may perhaps make sense of their yielding propositional outputs, but then the existence of efference-copying mechanisms of the kind postulated by Frith does not appear to be a necessary requirement for their proper functioning. Perhaps the solution to these difficulties consists in developing a plausible account of thought-monitoring that does not equate thinking with a kind of acting. But Currie and Ravenscroft provide no such account and in its absence an extension of their proposal from perceptual imaginings to belief-like, propositional imaginings is problematic.⁷

Second, assuming, for the sake of argument, that some delusions start their career as either *perceptual* or *propositional* imaginings, it is at best unclear that they are imaginings of the active variety. To qualify as active an imagining would probably have to satisfy at least one of the following two conditions: (1) the imaginative project is self-initiated, i.e. the subject decides to start imagining that P and sees what ensues from there; or (2) the imaginative process is self-directed: even if the imagining was externally triggered rather than self-generated, the subject is in a

⁶ Campbell (1999) notes that there are problems with the idea that our thoughts are the results of conscious intentions to think. Yet he argues that there is a sense in which we control and monitor our thoughts. Our occurrent thoughts are normally 'caused by a combination of our background of beliefs, desires and interests together with current external stimuli' (p. 617). Subpersonal monitoring of this causal process would ensure that the ongoing stream of occurrent thoughts is kept on track and would be the basis of the sense of authorship we normally experience towards our own thoughts. Gerrans (2001) points out that failure of this monitoring process could also explain the thought disorganization frequently encountered in schizophrenic patients. Yet, we commonly experience the occurrence of stray thoughts, thoughts that come to mind unbidden and that do not appear consistent with our background psychology, without thereby attributing the authorship of such thoughts to an alien agency. What Campbell's proposal, as indeed Frith's, does not explain is the attribution of one's thoughts to some *alien* agency.

⁷ Note that the doxastic account of delusions is perfectly consistent with the notion that many delusions involve impairments in efference copying. Doxastic models will regard such impairments as generating abnormal perceptual and affective states, which give rise to various delusional beliefs.

position to inhibit the act of imagining thus triggered or to control its course to some degree.

An important subset of delusions do not satisfy condition (1) insofar as they seem to be triggered by abnormal perceptual or emotional experiences rather than being generated in a stimulus-independent way. Nor do most delusions seem to satisfy condition (2). Delusional subjects do not seem to be able to stop themselves from imagining things that, we may suppose, they would very much like not to imagine.

It may be more plausible to describe delusional patients as suffering from an impairment in the *control* rather than the *monitoring* of imagination.⁸ They seem unable to inhibit stimulus-driven acts of imagining, to voluntarily shape their course and somehow guide, control, and prune out the flow of associations that are automatically triggered. Thus, it may be suggested that one important problem delusional patients encounter is loss of control over their imagination as a result of a defect in inhibitory processes.⁹ Delusional patients would then misidentify their imaginings for beliefs not as a consequence of an impaired sense of agency, but, more fundamentally, as a consequence of impaired agency.

A third difficulty for the metacognitive account is that it fails to explain why delusional patients mistake only *some* of their imaginings—viz., those that form the focus of their delusion—as beliefs and not others.¹⁰ This question is especially acute for patients with monothematic delusions. Such patients seem to be able to imagine, say, that the first person on the moon was a woman without mistaking this imagining for a belief, so why do they misidentify their delusional imaginings? It seems implausible to suppose that deficits in attitude-monitoring might be content-specific.

The objections we have raised so far in this section concern the plausibility of Currie and co-authors' specific account of the mechanisms by which we distinguish imagination from belief and of the kind of impairment in these mechanisms suffered by delusional subjects. Currie and co-authors may reply that the fact that normal subjects are able to distinguish their beliefs from their imaginings is sufficient evidence for the existence of such mechanisms, whatever their exact nature, and that therefore the idea that delusional subjects may misidentify their imaginings for beliefs as a result of an impairment in these mechanisms remains plausible.

The crucial issue then is not whether some delusions might start their career as imaginings, whether perceptual or propositional, active or passive, but whether an impairment in identification-mechanisms would result in these imaginings being

⁸ This might suggest that we are sympathetic to some components of the metacognitive account. Indeed we are.

⁹ See for instance Maher (2003) for evidence of such a defect in inhibitory processes in schizophrenic patients.

¹⁰ Note that Frith's proposal regarding delusions of control faces the same difficulty. Schizophrenic patients suffering from such delusions do not experience all their actions as being controlled by alien forces all the time.

mistaken for beliefs (second-order mistaken belief) or in their being transformed into, or giving rise to, first-order mistaken beliefs.¹¹ The first option is consistent with a meta-cognitive approach, the second with a doxastic account of delusions. Absent a persuasive account of identification-mechanisms and their possible impairments, the plausibility of the metacognitive account rests on the plausibility of (i) the claim that in delusions that begin as imaginings, these imaginings remain imaginings rather than becoming beliefs and (ii) the claim that they give rise to mistaken meta-beliefs. We have already argued that none of the three criteria Currie and co-authors advance as grounds for distinguishing beliefs from imaginings supports the meta-cognitive account over the doxastic alternative, hence (i) is doubtful. In the next section, we will argue that claim (ii) is also doubtful.

2.3 Verbal Behaviour and Second-Order Beliefs

We presented the metacognitive account as involving the conjunction of three claims: delusions are not (first-order) beliefs; delusions are imaginings; and deluded patients who seem to believe *P* actually believe that they believe *P*. This third claim is motivated by the need to explain the deluded patients' verbal reports that reflect the content of their delusions. Whereas the doxastic theorist holds that these reports reflect the patient's first-order beliefs, the metacognitive theorist holds that they reflect her second-order belief that she believes the content of the delusion.

We think that there are two difficulties with this story. First, the account appears to be in danger of conflating the distinction between first-order beliefs and second-order beliefs with the distinction between implicit beliefs and explicit beliefs. The former distinction is a distinction of content: first-order beliefs are not about other doxastic states, second-order beliefs are beliefs about first-order beliefs. The latter distinction is a distinction of functional role. Roughly, implicit beliefs are beliefs that the subject is unaware of or fails to acknowledge; explicit beliefs are beliefs that the subject is aware of or acknowledges. We could say that what is distinctive of implicit beliefs is that they are not assented to verbally (although they might be manifest in verbal behaviour, such as changing the subject). Verbal behaviour with explicit delusional content provides evidence that the delusional state is explicit

¹¹ Although we object to the specific account Currie and co-authors offer of delusions as misidentified active propositional imaginings, we do not deny that some delusions start their career as imaginings. As pointed out by one of our referees, findings of empirical studies which use principle components analysis to investigate the factor structure of clinical schizophrenic symptoms indicate that delusions and hallucinations typically co-occur in people with schizophrenia. A general impairment of monitoring the self-generated source of perceptual and propositional imaginings would explain why. But this in itself does not support the view that delusions that begin as imaginings remain imaginings rather than being transformed into beliefs.

rather than merely implicit, but it does not provide evidence that its content is second-order rather than (merely) first-order.¹²

The second difficulty with the view is this. As we have seen, Currie and Ravenscroft hold that a state must satisfy both the consistency and action-guidance constraints in order to be a belief. Since these constraints apply to beliefs in general, they should apply equally to second-order beliefs—such beliefs as the Capgras patient's belief that he believes that his wife has been replaced by an imposter. Assuming the consistency constraint on belief, this state must have precisely the kinds of impact on the patient's behaviour (both verbal and non-verbal) that Currie and co-authors point out is often absent. The belief <I believe that P> has many of the same rational implications that the belief <P> has; in particular, there is a tension between having the belief <I believe P> and not having the belief <P>. If the encapsulation objection in its various forms is a problem for the doxastic account, it would appear to be no less a problem for the metacognitive account in so far as it is a meta-doxastic account. (In a similar vein we might note that a patient with Cotard's delusion—who on the metacognitive account, believes that he believes that he is dead—appears to be no less pragmatically irrational than the patient who merely believes that he is dead.)

3. Delusions as Doxastic states: Towards a Defence¹³

3.1 A Sketch of a Theory of Belief

We now turn to a defence of the doxastic conception of delusions. Our ambitions here are quite limited; we certainly do not intend to argue that all delusional states are beliefs. Instead, we hope to sketch a model of belief—drawing from Cherniak (1986) and Schwitzgebel (2002)—on which it is plausible to suppose that many delusions qualify as beliefs.

The category 'belief' is far from homogeneous. As with many terms of folk psychology (desire, emotion, intention) the states that we are prepared to accept as beliefs differ in various ways. Unfortunately, dominant models of belief obscure the multi-dimensional nature of belief. Classical decision theory as well as classical theories of belief-ascription, in particular interpretationism (Davidson 1973, Dennett, 1971, Quine, 1960) emphasize rationality constraints to the exclusion of almost all other considerations. We accept that considerations of rationality have a role to play in constraining belief-ascription, but we deny that they are the only—or even the most important—constraints on belief-ascription.

¹² The need to keep the implicit/explicit distinction separate from the first-order/second-order distinction is apparent when we consider the doxastic states of young children. Three year-olds have explicit (verbally articulable) beliefs, but it is at best controversial that they have second-order beliefs.

¹³ Here we are much indebted to Stone and Young's (1997) important discussion of delusions as beliefs, although we take issue with their approach in a number of places.

Following Cherniak (1986), we reject the thought that the rationality constraints on belief-ascription should be derived from a model of ideal rationality. Given our finitary predicament—the computational, memory, and time limitations we are subject to—, it is actually irrational for us to aspire to ideal rationality.

As a consequence of this finitary predicament and of the structure of human memory, our belief system is, Cherniak argues, “quilted” into a patch of relatively independent subsystems’ (1986, p. 51). In rough outline, his argument for this claim is as follows. Human memory includes both long-term and short-term or working memory. Whereas the storage capacity of long-term memory appears not to have well-defined upper bounds, there are sharp limits to the number of items we can simultaneously hold in short-term memory. When assessing a new piece of information or deciding on a course of action, relevant beliefs must be recalled to short-term memory. Because of time and processing limitations, the memory search cannot be complete. Given these constraints and limitations, the compartmentalization of beliefs held in long-term memory into more or less permeable subsets of related beliefs appears as a necessary condition for minimal rationality rather than as a departure from rationality.¹⁴

The first lesson to be drawn from this picture of human ‘minimal’ rationality is that the existence of some degree of compartmentalization or encapsulation—and the concomitant failure to draw certain inferences or notice and resolve certain inconsistencies—need not count as evidence that an agent is irrational or that his purported beliefs do not really qualify as beliefs. Rather, given our finitary predicament, some local irrationality may be the price to pay for global rationality.

The second lesson is that there is an important sense in which—contrary to a prominent doxastic ideal—beliefs are not context-independent. If one thinks of beliefs as dispositions—or, as Schwitzgebel (2002) proposes, *clusters* of dispositions—then it is not too much of a stretch to suppose that what someone believes may be context-sensitive. If what a person believes depends on the dispositions she manifests, and if in turn the dispositions she manifests depend on the contents of her short-term memory, and these contents on the structure of her long-term memory together with the cues for recall provided by the current situation, then there is indeed an important sense in which her beliefs are context-dependent. (Here the notion of context-dependence does not refer just to the external context, but also to the current motivational and affective set of the agent.)

There is a large body of literature in social psychology that shows that people’s beliefs—or at least their verbal behaviour—is heavily influenced by mood and other non-evidential factors. In a representative example of this work, Forgas and Moylan (1987) interviewed subjects who had either just seen a happy film or who had just seen a sad film. On all questions happy subjects reported having

¹⁴ Cherniak is careful to stress that doxastic compartmentalization does not imply that anything goes. A belief set can be compartmentalized in the ‘wrong way’, not organized into subsets of related beliefs and thus leaving the system unable to recognize inconsistencies or make useful inferences.

significantly more positive beliefs and produced more optimistic judgments than did the sad subjects.

Affect is also important in structuring long-term memory and influencing recall. One shouldn't think that beliefs are grouped together simply on the basis of logical relations among them, emotion and affect also play an important role in creating and reinforcing connections among beliefs. If one has been mugged and the assailant was wearing a green jumper, thoughts of urban violence and thoughts of green clothing may become strongly associated for the victim despite there being no logical relations between the two families of thoughts.

One important aspect of Schwitzgebel's dispositional account of belief is its appeal, in contrast to traditional dispositional accounts, to phenomenal dispositions—dispositions to undergo certain sort of conscious experiences, including emotional experiences—in addition to behavioural and cognitive dispositions (Schwitzgebel, 2002). By insisting on phenomenal dispositions Schwitzgebel highlights a dimension of belief that may be thought of as characteristic of human beliefs. For humans at least, there can be something it is like from the inside to believe that *P*, a sense of conviction, of stronger or weaker commitment ranging from mere suspicion or what we called indicative imagination—where one has some inclination to think that *P* might be the case—to certainty—where one is absolutely sure that *P* is the case. Furthermore, it may be conjectured that these phenomenal dispositions associated with beliefs are part of what makes it possible for humans to have meta-beliefs, beliefs about their beliefs.

Let us recap the model of belief that we have just sketched. We defend two central claims. The first is that the link between belief and rationality is much looser than classical decision theory or interpretationist theories of belief ascription would maintain. Given their finitary predicament, human beings are answerable only to minimal rationality constraints. Given that an agent has a particular belief-desire set, she should undertake some of the appropriate actions; she should make some of the appropriate inferences; and, if inconsistencies arise in her belief set, eliminate some of them.

Our second and related claim is that beliefs are context-dependent in a number of ways. First, which dispositions are actualized is a function of several factors: (1) the way the long-term memory of the individual is structured, something that depends in turn both on the cognitive organization of the species and on the personal history of the individual, (2) the current external context, and (3) the current motivational and affective set of the individual. Second, belief-ascription is also context-dependent. According to Schwitzgebel, we have dispositional stereotypes for beliefs, specific clusters of behavioral, cognitive, and phenomenal dispositions we associate with given beliefs and expect to be manifested in standard situations. We attribute to a subject full belief that *P* if he conforms to the associated stereotype in standard situations and if his deviations from the stereotype are readily explainable or excusable by appeal to some non-standard feature of the situation in which they occur. When a deviation from the stereotype cannot be excused or explained in this way, whether or not the attributor ascribes the belief will depend on the context of the belief ascription and what her interests are.

3.2 Revisiting the Objections to the Doxastic Account

Let us now revisit the three classes of objections to the doxastic account that we introduced in §1. Content-based objections locate the source of the difficulty in what is putatively believed; evidence-based objections locate the difficulty in the patient's supposed lack of evidence for the belief; and commitment-based objections locate the source of the difficulty in the subject's lack of commitment to the truth of P.

Belief and Content

Content-based objections to the doxastic account can take a number of forms. Probably the most basic proceeds in two steps: first, one links the notion of believability with that of meaningfulness; second, one links the notion of meaningfulness with that of logical possibility. The result is a position according to which one cannot believe something that is not logically possible. Of course not all delusional claims concern logically impossible states of affairs, but some do.

An obvious response to this objection is that it ties belief and logical possibility together too tightly. If sound, the argument would show that it is impossible to believe any logical falsehood. We think that issues of belief-ascription are best approached via the question of predictive leverage rather than claims about logical possibility. Even if the delusion of inter-metamorphosis—which involves the belief that one person has changed into another—is necessarily false, it is no less capable of guiding one's action than is the thought <there are no irrational numbers>.

What about the Cotard delusion—which (often) involves the belief that one is dead? Is this belief pragmatically self-defeating? We think not. To believe that one doesn't exist might be pragmatically self-defeating (although perhaps not unbelievable), but it is not so clear that believing that one is dead is. The semantics of a word may reflect the fact that it is part of several scientific or folk-psychological domains. For 'death' these domains include biology, the subjective realm of consciousness, religious beliefs concerning life after death, supernatural entities, etc. In his use of words, a delusional patient may break the rules governing one of these intersecting language games while retaining the meaning his words have in other language games.

Belief and Evidence

A common theme in accounts of belief is that belief is constitutively connected to evidence (see e.g. Price, 1967). This, of course, generates a straightforward objection to the doxastic account, for one of the distinguishing marks of delusions appears to be that they are held in the absence of evidence (and, we might add, in the face of counter-evidence).

We deny that there is a constitutive connection between belief and evidence. As we pointed out earlier, folk-psychology—and indeed, scientific psychology—has little trouble with the thought that beliefs can be innate, and can be formed by motivation mechanisms. It seems entirely possible for brain damage to cause states that have the sorts of inferential relations to other cognitive states and action that beliefs have. Why such states shouldn't be called beliefs is unclear to us—after all, they could guide theoretical and practical reasoning in much the way that standard beliefs do.¹⁵

Further, there is a case to be made for thinking that at least some delusions are grounded in evidence of a sort. Here it is useful to distinguish between first-person (or private) evidence and third-person (or inter-subjective) evidence. It is clear that delusional patients lack third-person evidence for their beliefs: they are unable to produce evidence that is available to others. But there is some reason to think that delusional patients often have first-person evidence for their beliefs. As we mentioned earlier, there is a plausible experiential account of the genesis of a number of delusions (see section 2.1; for reviews see Langdon and Coltheart, 2000; Davies *et al.*, 2001).¹⁶

We might also add that the mere sense that a proposition is true can itself count as a source of evidence for it and incline one to believe it even in the face of what one acknowledges are good reasons to the contrary. We are often in this position with respect to philosophical claims. Someone might think that she has no evidence for the claim that there is a physical world other than that there seems to be a physical world. One can be affectively committed to a proposition—and thus believe it—despite confessing that one lacks any evidence in its favor.

Belief and Commitment

Perhaps the strongest objection to the doxastic account is that delusional patients often fail to demonstrate the kind of commitment to the content of the delusional state that they would if they believed it. This lack of commitment can manifest itself in theoretical reason, in practical reason, and in affective responses. In short, the objection is that delusional subjects typically fail to manifest the behavioural and cognitive dispositions that they would manifest if they really believed the content of their delusions.

¹⁵ Of course, one might think that it is unlikely that concepts could be structured in a belief-like way by random processes of brain damage. Perhaps brain damage is no more likely to cause a coherent thought than randomly selecting scrabble tiles is likely to spell out a coherent sentence. But remember that our claim here is only that it is *possible* for brain damage to directly cause belief.

¹⁶ Empiricist approaches to delusions can be conceptualized in different ways. See Bayne and Pacherie (2004a, 2004b) for discussion.

One way to think of this objection is as follows: delusional persons who seem to believe *P* deviate so much from the dispositional stereotype associated with the belief that *P* that there is little to no point in thinking of them as believers of *P*. But as we saw in the previous section, whether one should be ascribed the belief that *P* is not just a matter of whether the target manifests enough of the dispositions in the relevant cluster but also of whether his not manifesting some of these dispositions can be satisfactorily excused or explained by reference to some non-standard aspects of his situation.

And in the case of many delusional patients an appeal to relevant non-standard factors can be made. If, as the empiricist claims, monothematic delusions such as the Capgras delusion are grounded in unusual perceptual and affective experiences, then the patient's belief might be continually reinforced (see Bayne and Pacherie, 2004a for further discussion of this point). These non-standard perceptual and affective conditions may be thought to excuse the patient from manifesting the cognitive dispositions stereotypically associated with their belief.

In other delusions, it is less clear that there is much that the delusional patient can do to verify or evaluate the hypothesis in question. For example, it is not obvious what one might do to evaluate the notion that someone else is putting thoughts into one's head. Of course, this thought is antecedently highly implausible in the context of normal conceptions of causation, but normal conceptions of causation are themselves grounded in normal experiences of agency. It is not clear what evidential force normal conceptions of agency ought to have in the context of abnormal experiences of agency. (We might turn this issue on its head by asking whether and how we evaluate our belief that we are the thinkers of our own thoughts.)

A second way in which delusional patients fail to conform to the stereotypical cluster is emotionally and affectively. To address the objection, one either needs to argue that delusional patients really do satisfy the affective conditions, or that the affective conditions are too stringent.

We are inclined to take the latter position. The cluster of dispositions that mark out particular beliefs typically includes dispositions to certain emotional responses, but we resist the thought that emotional and affective dispositions are constitutive elements of the belief stereotype. We should also note that the ascription of emotional states is itself far from straightforward. As Bentall (2003, p. 225) points out, the subjective experience of emotion can dissociate from the behavioural features of emotion: some depressive patients have the former but not the latter. Given the close connection between some forms of delusion (such as Cotard's delusion) and depression, it might also be the case that some delusional patients have the subjective experiences associated with certain emotions even when they lack their behavioural manifestations. If so, the question of whether such individuals have a certain emotion might not admit of a definitive answer.

Finally, we turn to the objection from practical agency. Some familiar points dull the force of this objection. For one thing, delusion-generated action is not as

rare as is often thought. Furthermore, action is not caused by cognitive states alone but by cognitive states in conjunction with motivational states. As Stone and Young remind us (1997), deluded patients have disrupted affective and emotional states, and they know that acting on their beliefs might result in hospitalization.

In addition to these familiar points, the notion of doxastic context-sensitivity might prove useful here. Consider the curious fact that some Capgras patients appear to think that their spouse has been replaced by an imposter when in visual contact with them but not when talking to them on the telephone (see Hirstein and Ramachandran, 1997).¹⁷ Perhaps the content of the person's visual state is such that it leads them to endorse the imposter hypothesis. When, however, this visual evidence is absent, the person's normal disposition to believe that their spouse is their spouse is triggered. Here, unqualified ascription of any belief concerning the identity of the person's spouse is problematic. Instead, the tempting thing to say is the person's beliefs concerning the identity of their spouse are dependent on their current perceptual information: to a first approximation we might say that the person has the imposter belief when, and only when, he is in visual contact with her. (We might liken this behaviour to the field-dependent behaviour that one sees in frontal lobe patients, who are unable to inhibit their pre-potent responses to environmental cues. Perhaps some delusional patients are doxastically field-dependent.)

4. Conclusion

We finish with some reflections on the practical importance of the debate between Currie and ourselves. The metacognitive account raises a number of important issues for philosophical psychology, but what implications does it have for the treatment of delusions?

Consider cognitive behavioural therapy (CBT), an important form of therapy for delusions (see Dickerson, 2000). An essential component of CBT involves questioning the consistency and plausibility of the patient's delusions (Chadwick *et al.*, 1996). This form of therapy seems to accord with the doxastic account, in that the therapist treats the delusional patient as a believer of P, and he or she gently invites the patient to question whether P is the thing that ought to be believed. But CBT seems to be puzzling from the metacognitive perspective: after all, according to the metacognitive account, the patient doesn't believe P. Instead of the first-order approach of CBT, the metacognitive account would seem to suggest that the point of therapy should be to disabuse patients of their false beliefs about

¹⁷ As one of the referees reminds us, there are cases in which Capgras patients seem to endorse the imposter belief when talking to their spouse on the telephone. And there are also cases of Capgras in blind persons (Hermanowicz, 2002; Dietl *et al.*, 2003). We are certainly not suggesting that there is any necessary dependence of the Capgras delusion on visual perception.

the nature of their delusional (imaginative) states; and/or to remove the imaginative states themselves. It is not clear how either of these goals could be achieved by CBT. To disabuse, say, a Capgras patient of his false meta-belief, the therapist would be better off pointing out to the patient that he does not behave in ways that he would behave if he really believed that his wife had been replaced by imposters. To the best of our knowledge therapists do not take this approach, nor does it seem to be an approach that they would welcome.

Department of Philosophy
Macquarie University
Institut Jean-Nicod
CNRS-EHESS-ENS

References

- American Psychiatric Association 2000: *Diagnostic and Statistical Manual of Mental Disorders, Text Revision*. Fourth Edition. Washington, D.C.: American Psychiatric Association.
- Bayne, T. and Pacherie, E. 2004a: Bottom-up or top-down: Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry & Psychology*, 11/1, 1–11.
- Bayne, T. and Pacherie, E. 2004b: Experience, belief, & the interpretive fold. *Philosophy, Psychiatry & Psychology*, 11/1, 81–86.
- Bentall, R. 2003: *Madness Explained*. London: Allen Lane.
- Bermúdez, J.L. 2001: Normativity and rationality in delusional psychiatric disorders. *Mind & Language*, 16/5, 457–93.
- Berrios, G. 1991: Delusions as 'wrong beliefs': A conceptual history. *British Journal of Psychiatry*, 159, 6–13.
- Bleuler, E. 1924: *Textbook of Psychiatry* (4th ed), trans. AA Brill. New York: Macmillan.
- Breen, B., Caine, D., Coltheart, M., Hendy, J. and Roberts, C. 2000: Towards an understanding of delusions of misidentification: Four case studies. In M. Coltheart and M. Davies (eds.) *Pathologies of Belief*, Oxford: Blackwell, 75–110.
- Brett-Jones, J., Garety, P. and Hemsley, D. 1987: Measuring delusional experiences: A method and its application. *British Journal of Clinical Psychology*, 26, 257–65.
- Campbell, J. 1999: Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, 82, 609–25.
- Campbell, J. 2001: Rationality, meaning, and the analysis of delusion. *Philosophy, Psychiatry, & Psychology*, 8/2,3, 89–100.
- Chadwick, P., Birchwood M., and Trower, P. 1996: *Cognitive Therapy for Delusions, Voices and Paranoia*. New York: Wiley.
- Cherniak, C. 1986: *Minimal Rationality*. Cambridge, MA: MIT Press.
- Coltheart, M. and Davies, M. (eds.) 2000: Introduction, *Pathologies of Belief*. Oxford: Blackwell.
- Currie, G. 2000: Imagination, delusion and hallucinations. In M. Coltheart and M. Davies (eds.), *Pathologies of Belief*. Blackwell, 167–182.

- Currie, G. and Jureidini, J. 2001: Delusion, rationality, empathy. *Philosophy, Psychiatry & Psychology*, 8/2,3, 159–62.
- Currie, G. and Ravenscroft, I. 2002: *Recreative Minds*. Oxford: Oxford University Press.
- David, A. 1990: Insight and psychosis. *British Journal of Psychiatry*, 156, 798–808.
- Davidson, D. 1973: On the very idea of a conceptual scheme. *Proceedings of the American Philosophical Association*, 47, 5–20.
- Davies, M., Coltheart, M., Langdon, R. and Breen, N. 2001: Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry & Psychology*, 8/2,3, 133–58.
- Dennett, D.C. 1971: Intentional systems. *Journal of Philosophy*, 8, 87–106.
- De Pauw, K.W. and Szulecka, T.K. 1988: Dangerous delusions: Violence and the misidentification syndromes. *British Journal of Psychiatry*, 152, 91–96.
- Dickerson, F.B. 2000: Cognitive behavioural psychotherapy for schizophrenia: A review of recent empirical studies. *Schizophrenia Research*, 43/2,3, 71–90.
- Dietl, T., Herr, A., Brunner, H., and Friess, E. 2003: Case report: Capgras syndrome—out of sight, out of mind? *Acta Psychiatrica Scandinavica*, 108/6, 460.
- Ellis, H.D. and Young, A.W. 1990: Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157, 239–248.
- Forgas, J.P. and Moylan, S. 1987: After the movies: The effects of transient mood states on social judgments. *Personality and Social Psychology Bulletin*, 13, 478–489.
- Förstl, H., Almeida, O.P., Owen, A.M., Burns, A. and Howard, R. 1991: Psychiatric, neurological and medical aspects of misidentification syndromes: A review of 260 cases. *Psychological Medicine*, 21, 905–10.
- Gerrans, P. 2001: Authorship and ownership of thoughts. *Philosophy, Psychiatry & Psychology*, 8/2,3, 231–237.
- Gerrans, P. 2002: A one-stage explanation of the Cotard delusion. *Philosophy, Psychiatry & Psychology* 9,1, 47–53.
- Halligan, P.W. and Marshall, J.C. 1995: Supernumerary phantom limb after right hemisphere stroke. *Journal of Neurology, Neurosurgery and Psychiatry*, 59, 341–2.
- Hermanowicz, N. 2002: A blind man with Parkinson's disease, visual hallucinations, and Capgras syndrome. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 14/4, 462–463.
- Hirstein, W. and Ramachandran, V.S. 1997: Capgras syndrome: A novel probe for understanding the neural representation of identity and familiarity of persons. *Proceedings of the Royal Society of London (Biology)*, 264, 437–444.
- Jaspers, K. 1963: *General Psychopathology*, trans. J. Hoenig and M.W. Hamilton. Manchester: Manchester University Press.
- Langdon, R. and Coltheart, M. 2000: The cognitive neuropsychology of delusions. In M. Coltheart and M. Davies (eds.), *Pathologies of Belief*. Oxford: Blackwell.
- Maher, B. 1974: Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30, 98–113.
- Maher, B. 1999: Anomalous experience in everyday life: Its significance for psychopathology. *The Monist*, 82, 547–70.

- Maher, B. 2003: Schizophrenia, aberrant utterance and delusions of control: The disconnection of speech and thought, and the connection of experience and belief. *Mind & Language*, 18/1, 1–22.
- Mele, A.R. 1993: Motivated belief. *Behaviour & Philosophy*, 21/2, 19–27.
- O'Dwyer, J.M. 1990: Coexistence of the Capgras and de Clérambault's syndromes. *British Journal of Psychiatry*, 156, 575–77.
- Price, H.H. 1967: Some considerations about belief. In: A. Phillips Griffiths (ed.), *Knowledge and Belief*. Oxford: Oxford University Press.
- Quine, W.V.O. 1960: *Word and Object*. Cambridge, MA: MIT Press.
- Ramachandran, V.S. and Blakeslee, S. 1998: *Phantoms in the Brain: Human Nature and the Architecture of the Mind*. London: Fourth Estate.
- Sass, L. 1994: *The Paradoxes of Delusion: Wittgenstein, Schreber and the Schizophrenic Mind*. Ithaca, N.Y.: Cornell University Press.
- Schwitzgebel, E. 2002: A phenomenal, dispositional account of belief. *Notls*, 36/2, 249–275.
- Stephens, G.L. and Graham, G. 2004: Reconceiving delusions. *International Review of Psychiatry*, 16/3, 236–241.
- Stephens, G.L. and Graham, G. forthcoming: The delusional stance. In M. Chung, K.W.M. Fulford and G. Graham (eds.) *The Philosophical Understanding of Schizophrenia*. Oxford University Press.
- Stone, T. and Young, A. 1997: Delusions and brain injury: The philosophy and psychology of belief. *Mind & Language* 12, 327–64.
- Weinstein, E.A. and Kahn, R.L. 1955: *Denial of Illness*. Illinois: Charles C. Thomas.
- Wessely, S., Buchanan, A., Reed, A., et al 1993: Acting on delusions. I: Prevalence. *British Journal of Psychiatry*, 163, 69–76.
- Young, A.W. 1999: Delusions. *The Monist*. 82/4, 571–589.
- Young, A. and Leafhead, K. 1996: Betwixt life and death: Case studies of the Cotard delusion. In P. Halligan and J. Marshall (eds.), *Method in Madness: Case Studies in Cognitive Neuropsychiatry*, Hove, UK: Psychology Press.