

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Consciousness and Cognition xxx (2006) xxx-xxx

**Consciousness
and
Cognition**www.elsevier.com/locate/concog

Phenomenology and delusions: Who put the 'alien' in alien control?

Elisabeth Pacherie^{a,*}, Melissa Green^b, Tim Bayne^c^a *Institut Jean Nicod UMR 8129, CNRS-EHESS-ENS, 1 bis, avenue de Lowendal, 75007 Paris, France*^b *Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia*^c *Department of Philosophy, Macquarie University, Sydney, NSW 2109, Australia*

Received 25 April 2005

Abstract

Although current models of delusion converge in proposing that delusions are based on unusual experiences, they differ in the role that they accord experience in the formation of delusions. On some accounts, the experience comprises the very content of the delusion, whereas on other accounts the delusion is adopted in an attempt to explain an unusual experience. We call these the *endorsement* and *explanationist* models, respectively. We examine the debate between endorsement and explanationist models with respect to the 'alien control' delusion. People with delusions of alien control believe that their actions and/or thoughts are being controlled by an external agent. Some accounts of alien control (e.g., Frith, Blakemore, & Wolpert, 2000a) are best thought of in explanationist terms; other accounts (e.g., Jeannerod, 1999) seem more suited to an endorsement approach. We argue that recent cognitive and neurophysiological evidence favours an endorsement model of the delusion of alien control.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Delusions; Alien control; Perception; Monitoring; Agency; Intentions; Willed action; Simulation

1. Introduction

Current models of delusion converge in proposing that delusional beliefs are based on unusual experiences of various kinds. For example, it is argued that the Capgras delusion (the belief that a known person has been replaced by an impostor) is triggered by an abnormal affective experience in response to seeing a known person; loss of the affective response to a familiar person's face may lead to the belief that the person has been replaced by an impostor (Ellis & Young, 1990). Similarly, the Cotard delusion (which involves the belief that one is dead or unreal in some way) may stem from a general flattening of affective responses to external stimuli (Ellis & Young, 1990), while the seed of the Frégoli delusion (the belief that one is being followed by known people who are in disguise) may lie in *heightened* affective responses to unfamiliar faces (Davies, Coltheart,

* Corresponding author. Fax: +33 1 53 59 32 99.

E-mail address: Elisabeth.Pacherie@ens.fr (E. Pacherie).

Langdon, & Breen, 2001). Experience-based proposals have been provided for a number of other delusions (Breen, Caine, Coltheart, Hendy, & Roberts, 2000; Breen, Caine, & Coltheart, 2001; Davies et al., 2001; Davies, Aimola Davies, & Coltheart, 2005; Langdon & Coltheart, 2000; Maher, 1988; Stone & Young, 1997).

But behind this broad agreement lies an important controversy about the precise role that experience plays in the formation of delusions. On some accounts the experience comprises the very content of the delusion, such that the delusional patient simply believes what they experience; the delusional belief encodes the content of the perceptual experience in linguistic form. We will call such accounts *endorsement accounts*, on the grounds that the person believes—that is, doxastically endorses—the content of their perceptual state, or at least something very much like the content of their perceptual state.¹ An endorsement account of the Capgras delusion, for example, would hold that the Capgras patient sees the woman he is looking at (who is his wife) as an imposter (that is, as someone who merely looks like his wife).

Other experience-based accounts of delusion construe the relationship between delusional experience and delusional belief in *explanationist* terms. The patient adopts the delusion in an attempt to explain, or make sense of, an unusual experience. According to the explanationist, the Capgras patient does not perceive his wife as an impostor, rather, he simply fails to have the expected experience of familiarity when looking at his wife. He forms the belief that the woman he is looking at is not his wife in an attempt to explain his lack of affect.²

In this paper, we employ the distinction between endorsement and explanationist models to evaluate accounts of the ‘alien control’ delusion. People with delusions of alien control believe that their actions and/or thoughts are being controlled by an external agent. Some accounts of alien control (e.g., Frith et al., 2000a) are best thought of in explanationist terms; other accounts (e.g., Jeannerod, 1999) seem more suited to an endorsement approach. We argue that recent cognitive and neurophysiological evidence favours an endorsement model of the delusion of alien control.

2. Two experiential routes to delusion

Let us consider the distinction between endorsement and explanationist models in more detail. First, it should be noted that it is possible that a comprehensive account of delusions will contain both endorsement and explanationist elements. Perhaps some delusions should be accounted for in endorsement terms and others in explanationist terms. It is also possible that in some instances patients adopt delusional beliefs in an attempt to explain their unusual experience, but as a result of having adopted the delusional belief their experiences come to inherit the content of the delusion itself. For example, someone might form the Capgras delusion in an attempt to account for their strange experience of lack of affect, but having formed the delusion may come to see their wife as an imposter (see Fleming, 1992).

Experience-based accounts of delusions involve (at least) two components: (i) an explanation of the delusional patient’s *experiential* state; and (ii) an explanation of the delusional patient’s *doxastic* state (his belief). Endorsement and explanationist models face distinct challenges in providing these explanations. Explanationist models appear to have an easier job of (i) than endorsement models: the less one packs into the content of the perceptual experience, the easier it is to explain how the experiential state acquires its content. Very primitive explanationist models, according to which the delusion in question is generated by nothing more than an absence of certain kinds of affect, would seem to have rather little work to do here.

But what explanationist models gain with respect to (i) they lose with respect to (ii). The explanationist holds that delusional beliefs are adopted in an attempt to explain unusual experiences. The problem with this suggestion is that delusional beliefs are typically very poor explanations of the events that they are supposedly intended to explain. More plausible explanations of their strange experiences are available to the patients, some of which might be actively recommended to them by family and medical staff. Furthermore, delusional patients do not appear to hold their delusions in the tentative and provisional manner with which explanations are usually held. Explanationists are well-positioned to account for the content of the patient’s experiential

¹ The “something very much like” clause is intended to handle the worry that while the delusional belief has conceptual content, the perceptual state might have only non-conceptual content.

² For discussions of the possible contents of the abnormal experience in Capgras delusion see Bayne and Pacherie (2004a, 2004b); Pacherie (forthcoming).

state, but they face problems in explaining why the patient refuses to acknowledge the implausibility of the delusional beliefs they adopt in response to those experiences.

By contrast, endorsement models would seem to have a more plausible story to tell about how delusional patients move from experiences to belief. Perhaps, as Davies et al. (2001) suggest, delusional individuals might have difficulties inhibiting the pre-potent doxastic response to their experiences. Seeing is certainly not believing, but the transition from perceiving ‘that P’ to believing ‘that P’ is a familiar and attractive one. Of course, things are not completely plain sailing for the endorsement theorist. For one thing, we would need to know why delusional patients fail to take account of their background beliefs; why do they fail to inhibit the pre-potent doxastic response in the way that a ‘healthy’ person presumably would, if faced with the same bizarre and implausible sensory experience?³ But on the face of things the endorsement account looks to have a more plausible account of why, given the experiences that the account ascribes to the patients, they go on to form the beliefs that they do. Where the endorsement account would appear to be weakest is in explaining how delusional patients could have the experiences that the account says they do. We return to this point below.

How does the distinction between endorsement and explanationist models map on to the better-known distinction between one-deficit and two-deficit accounts of delusions? One-deficit accounts, such as Maher’s (Maher, 1974), hold that the only impairments delusional patients have are perceptual: their mechanisms of belief-fixation operate within the normal range (although they might be biased in some way). Two-deficit accounts, by contrast, hold that delusional patients have belief-fixation processes that are outside the normal range. The distinction between one- and two-deficit accounts is *orthogonal* to the distinction between explanationist and endorsement accounts (Davies et al., 2001). Both endorsement and explanationist models can be developed in either one-deficit or two-deficit terms. Consider first the endorsement account. As the Müller-Lyer illusion demonstrates, normal individuals do not always believe ‘that P’ when confronted with the perception ‘that P.’ And although the explanationist model of delusions might be thought to suggest a two-deficit view, it can be developed in one-deficit terms. Whether or not the explanationist will need to invoke a belief-formation abnormality depends on whether a normal individual would form (and maintain) the sorts of explanations of their unusual experiences that delusional patients do (Bayne & Pacherie, 2004a, 2004b).

These distinctions allow us to notice that one way one might be tempted to argue for a two-deficit account is fallacious. It is sometimes suggested that the discovery of two individuals who share the same experiential abnormality, but only one of which was delusional, would weigh decisively in favour of a two-deficit account of delusions. The logic behind this claim is that we would need to appeal to a second (belief-fixation) *deficit* to explain why only the delusional individual adopted the delusional belief in response to the unusual experience. But this inference is fallacious: for all we know, a vast range of belief-fixation processes fall *within* the normal range, and it is quite possible that there will be individuals who share exactly the same phenomenology, and whose belief-forming processes are within the normal range, but only some of which go on to form delusional beliefs. Two individuals, S1 and S2, could reason from exactly the same types of experiential states, via *different but normal* belief-fixation procedures (doxastic styles), to quite different doxastic states; S1 might put a higher premium on theoretical simplicity than S2, while S2 might put a higher premium on mechanistic explanations than S1. The dissociation argument would show that belief-forming processes must play a role in the formation of delusional belief, but it would not show that delusional individuals have a beliefs-forming *deficit*.⁴

³ Or would they? It might be argued that by the very nature of the aberrant experience, even a ‘healthy’ individual may not have the capacity to override the pre-potent doxastic response. See Hohwy and Rosenberg (2005).

⁴ Of course, if one thinks of a “two-factor” account as any account which appeals to factors about belief-fixation processes to explain the formation of delusional belief, irrespective of whether those factors place delusional patients within the normal range or not, then the argument presented above goes through—with the caveat that the second ‘factor’ need not entail a ‘deficit’ in belief-fixating processes to distinguish the deluded from non-deluded person (when beliefs were formed on the basis of identical sensory input). For this reason, we prefer to view empiricist models of delusion in terms of two *factors*, given that very little is known about ‘normal’ belief-fixation processes, and since there appears to be no definitive evidence to suggest that these must be deficient to account for delusion. However, with the use of the ‘two-factor’ terminology there remains the issue of whether an account which says that delusional patients’ belief-evaluation processes lie within the normal spectrum is representative of a “one-stage” (i.e., one-deficit) model. Proponents of a one-stage view are committed to the idea that a model can entail two-factors only if the second process falls outside of the normal spectrum. But both accounts agree that an initial neuropsychological *deficit* will account for the experiential component, and both agree that a second factor is necessary to move the sensory experience from the status of perception to belief.

Although recent accounts of delusions have generally been quite vague about both the content of the abnormal experiences they posit and the precise way in which such experiences generate delusional beliefs, most theorists seem inclined towards explanationism. Young and Leafhead (1996) suggest that Cotard and Capgras patients arrive at different delusional states because they adopt different explanatory strategies towards the same abnormal experience of loss of affect: Cotard patients are depressed, and as a result they explain their loss of affect in terms of a change to themselves, while Capgras patients are suspicious, and as a result they explain their loss of affect in terms of changes to their environment. One could have reason to challenge this attractive suggestion if, as Gerrans argues, there is reason to think that the Cotard and Capgras delusions are grounded in distinct phenomenal states (Gerrans, 2002).⁵ Gerrans himself seems to adopt an explanationist account of the Capgras delusion. He argues that “The Capgras person does not perceive the other person as a double. Rather she perceives the other person and, while doing so, has a very atypical affective experience. Because this experience occurs in a context in which, normally, perception is coupled with a recognition judgment, she infers that the person she is seeing is a double” (2002, p. 67).

One reason for the widespread sympathy with explanationist models may be the view that the relationship between perception and belief is *typically* explanatory. Some theorists think of perception in general, and emotional and affective states in particular, as non-representational. On this view, perceptual beliefs are adopted in the attempt to explain our perceptual states: I believe that I am looking at a cat in an attempt to explain certain sensations I am currently having.

Such explanationist approaches to the perception-belief interface should be rejected. There are two central respects in which they fall short. First, the explanationist needs to explain how the adoption of perceptual beliefs (such as “this is a cat”) could explain the sensations in question. Exactly how this explanation might go is anything but clear. Second, the proposal flies in the face of phenomenology. Our experience of the world is shot through with representational content. This is clear in the case of visual perception, as the much discussed Müller–Lyer illusion demonstrates: the two lines appear to be of different lengths, even when one believes that they are the same length. But it is worthwhile pausing to consider the degree to which other facets of experience also have representational content. Think of the patient with phantom limbs, who experiences her phantom limb as reaching for a door, even though she knows that she is performing no such action. Think of what it is like to watch Heider’s visual stimuli (Heider & Simmel, 1944), where one sees the geometrical stimuli as intentional entities (“the big square is chasing the small triangle”). In all of these cases, one has perceptual experiences that naturally give rise to beliefs with the same content unless (slow, conscious) processes of doxastic inhibition intervene. Given that our experience of the world is rich with representational content, it is not implausible to suppose that the dominant experience-based route to belief takes an endorsement form.

3. Delusions of alien control

We propose to examine the debate between endorsement and explanationist accounts in the context of delusions of alien control. The delusion of alien control involves the belief that some other agent—another person, a supernatural entity (e.g., God), a collective of others (e.g., the government), or a non-human device such as a satellite or computer—is controlling some of one’s actions. Patients with alien control will report:

“My fingers pick up the pen, but I don’t control them. What they do is nothing to do with me.”

“The force moved my lips. I began to speak. The words were made for me.” (From Mellors, 1970, p. 18)

“I felt like an automaton, guided by a female spirit who had entered me during it [an arm movement].”

“I thought you [the experimenter] were varying the movements with your thoughts.”

“I could feel God guiding me [during an arm movement].” (From Spence et al., 1997).

There are four main components to the content of the delusional belief. First, the patients report a sense of *passivity* vis-à-vis the movements they produce. The second component is *externality*: the movements are reported as controlled by an external force; they are not just experienced as involuntary movements in the

⁵ Cases of concurrent Capgras and Cotard delusions would also be problematic for the Young–Leafhead suggestion—see Butler (2000); Joseph (1986); Wolfe and McKenzie (1994).

way that some motor reflexes or twitches are. Third, the belief involves reference to *agency*: the external force controlling the movements is thought of as an agent, not merely a physical force as would be the case for instance if we felt that a strong gust of wind is making us stumble. The fourth is *particularity*: the alien agent is identified by the patient as a particular individual or collective agent (God, the CIA, the experimenter, etc.)

Which of these aspects of the content of the delusional belief are already parts of the patient's experience? It is this question that is at the heart of the debate between endorsement and explanationist accounts of alien control.

According to endorsement accounts of alien control, the patient experiences their actions and/or their thoughts as being under the control of someone else. The representational content of the patient's experience would be roughly, "so and so is making me do X," or "So and so is doing X (where X involves my body)." A slightly weaker account would incorporate otherness into the experience, without tying the action to any particular agent.

The explanationist might respond to the endorsement model in two ways. On the one hand, she might argue that the contents of the delusion of alien control cannot be perceptually encoded. Alternatively, she might allow that the contents of the alien control belief can enter into perception, but only on the condition that the person in question already has the delusional belief: experiencing one's action as controlled by an alien agent is possible only on the condition that one already believes that an alien agent is controlling one's actions. We will focus on the stronger and more straightforward objection here.

Could the experience of someone else controlling one's actions be loaded into perception, especially when the person is not perceptually salient? In the remainder of this paper, we will discuss two recent models of alien control and examine to what extent they support an endorsement approach.

4. The central-monitoring account

Frith's original account of alien control is most naturally thought of in explanationist terms (Frith, 1987, 1992). The main components of his central account were a distinction between two kinds of intentions, stimulus intentions (i.e., unconscious intentions) automatically triggered by a stimulus and willed intentions (i.e., conscious intentions based on internal plans and goals), together with a distinction between two levels of monitoring. At the lower level, action-monitoring involved using efference-copying mechanisms to distinguish between changes due to our actions and changes due to external factors. At the higher level, intention-monitoring made possible the distinction between stimulus-induced actions and spontaneous actions resulting from willed intentions. Frith's model of alien control posited the existence of a deficit in intention-monitoring resulting in the *loss of awareness of 'willed' intentions* to act. The loss of such awareness was equated with an experience of *lack of sense of agency* over one's actions, and was grounded in the assumption that we usually feel a sense of effort with respect to our 'willed' actions.

Lack of a sense of agency over one's action is still a far cry from the presence of a sense of alien control and in that respect the model is clearly explanationist. Moreover, it is not even clear that an impairment in intention-monitoring could account for a sense of passivity vis-à-vis one's actions. The phenomenology resulting from any such impairment would not differ from that of stimulus-induced actions, and it seems that, in normal subjects at least, a minimal sense of agency—rather than a sense of complete passivity—attaches to stimulus-induced actions. Impaired action-monitoring combined with impaired intention-monitoring may lead to a blurring of the distinction between what one does and what happens to one, but this does not yet amount to an experience of alien control.

Independently of the explanationist/endorsement debate, there are several respects in which this original account was questionable. First, as has been pointed out by several authors (Campbell, 1999; Gallagher, 2000; Pacherie, 2001; Spence, 2001), it would seem to have difficulty accounting for thought-insertion, given that most of the time we do not have any conscious feeling of effort or intention to think when thinking a certain thought. Although there may be some sense of *effort* involved in keeping one's attention focused for the purpose of thinking through an issue, many of our other undirected thoughts have no such quality. A second criticism, voiced by Spence (2001), concerns the paradoxical nature of the model. Alien control is supposed to result from defective monitoring of willed intentions: the patient is unaware of his willed intention. But on Frith's view, one defining feature of willed intentions is their conscious character. The model

therefore seems to require that in delusions of alien control willed intentions be both conscious and unconscious. As Spence points out “this apparent paradox might be resolved if the patient were said to be conscious of his intention as one that is ‘alien,’ but then the patient would no longer be unaware of his intention, and so his ‘unawareness’ of it could no longer form the basis of its ‘alien-ness’” (2001, p. 167). Third, as Frith himself later acknowledged, the idea that experiences of alien control arise through a lack of awareness of intended actions “is inconsistent with the patients’ ability to follow the commands of the experimenter, to avoid showing utilization behaviour, and to correct errors on the basis of sensory feedback about limb positions (which requires comparisons of intended actions and their consequences)” (Frith et al., 2000a, 1784).

Frith, Blakemore, and Wolpert’s revised account of delusions of alien control (Blakemore & Frith, 2003; Blakemore, Wolpert, & Frith, 2002, 2003; Frith et al., 2000a, Frith, Blakemore, & Wolpert, 2000b) is based on a more detailed model of action control. According to this model, the motor control system makes use of two kinds of internal models, controllers and predictors, together with a number of comparators. The controllers, also called inverse models, compute the motor commands necessary to achieve a certain outcome given the actual state of the system and of its environment. The predictors (or forward models) are fed a copy of these motor commands and they compute estimates of the sensory consequences of the ensuing movement. These predictions can be used in several ways. First, they can be used to anticipate and compensate for the sensory effects of movements. Second, they can also be used to filter sensory information and to attenuate the component that is due to self-movement.⁶ Third, they can be used to maintain accurate performance in the presence of feedback delays. The internal feedback of the predicted state of the system is available before the actual sensory feedback and can be compared with the desired state to determine performance error and trigger corrections.

Besides its role in the control of actions, the motor system also has a role to play in the awareness of action. According to Frith and his colleagues, in normal circumstances when an agent is performing an action, she is aware of (i) her goal, (ii) her intention to move, (iii) her movement having occurred, and (iv) her having initiated her movement. In contrast, a patient with delusions of control has normal awareness of (i)–(iii) but not of (iv). According to the revised model, awareness of initiating a movement depends on awareness of the predicted sensory consequences of the movement. This view is based on evidence that awareness of initiating a movement in healthy subject is reported by the agent between 80 and 200 ms before the movement actually occurs (Haggard & Magno, 1999; Libet, Gleason, Wright, & Pearl, 1983; Libet, 1985). It therefore seems that our awareness of intending to move may rest upon the *internally predicted* sensory consequences of movement, available prior to the actual execution of the action. In delusions of control, the *prediction* mechanism would be faulty and the patient would therefore be unaware of having initiated the movement.

Yet, it is somewhat unclear what Frith and his co-workers think is wrong with the predictors. In Frith et al. (2000a) they accept Jeannerod’s point (Jeannerod, 1999) that generation and control of a movement require one kind of representation, while conscious judgements about a movement require another kind of representation. The predictors are therefore thought to compute two different kinds of representations, but where exactly do they go wrong? According to Jeannerod (1999), reaching for a target, for instance, requires that the spatial coordinates of the target be transformed into a set of commands coded in a body-centred reference frame. Motor control relies on the comparison of predictions and outcomes within the motor system, and in order for the comparator to use the predictions they must be coded in the same body-centred reference frame. Conscious judgements about movement, by contrast, rely on comparisons between the internal model of the goal and (typically visual) perceptions of the environment. It follows that such judgements are likely to be made on the basis of *central* representations coded in a set of coordinates used for perception rather than the coordinates used in the body-centred reference frame. These central representations will also be employed by other agents or observers attending the same visual scene. Borrowing Barresi and Moore’s terminology (Barresi & Moore, 1996), Jeannerod refers to the representations used for motor control as ‘private’ representations (because they encode first-person information), and to the representations used for judgements about actions as ‘public’ (because they encode third-person information). Furthermore, private representations are

⁶ Evidence for this claim comes from a series of investigations (see Blakemore, Wolpert, & Frith, 2000) showing that healthy people are unable to tickle themselves because the sensory consequences are attenuated due to expectancies generated by the forward model.

not accessible to consciousness, whereas public representations are consciously accessible. To be used for motor control, predictions should therefore be represented in a ‘first-person’ or ‘private’ format, while conscious judgements about movements would require predictions to be represented in a ‘third person’ or ‘public’ format.

The idea, then, is that the predictors proceed in two stages. They start by computing first-person representations of the predicted consequences of movements that are used for (non-conscious) motor control. These first-person representations are then translated into third-person representations that are needed for conscious awareness of predicted sensory consequences and conscious monitoring of the motor control system.

The abnormality could be found at the first stage, yielding faulty or inaccurate predictions of the sensory consequences of the action, or there could be something wrong with the mechanism that translates the first-person representations computed at the first stage into third-person representations. Frith et al. (2000a) reject the first option, arguing that there is nothing obviously abnormal in the motor control of patients with delusions of control. They suggest instead that the problem may lie with the mechanism that translates the first-person representations into third-person ones. But there are several different things that may be wrong with this translation mechanism. First, it may be that it yields inaccurate third-person translations of the predictions made at the first stage. But then the problem would not be one of lack of awareness of predictions, but one of awareness of *inaccurate* predictions. This would be sufficient to explain why schizophrenic patients, as opposed to normal controls, do not show perceptual attenuation of self-produced sensory stimulation—for instance, they *can* tickle themselves (Blakemore et al., 2000)⁷—but it would not explain why patients with delusions of control are not aware of initiating an action. If, as the model postulates, one’s awareness of initiating an action rests on the forward model’s prediction about the sensory consequences of an action, awareness of initiating an action should occur whether the prediction is accurate or not.

Alternatively, it could be that although the predictions are accurately translated, for some reason they are prevented from reaching consciousness. This would explain why patients with alien control are not *aware* of initiating actions. However, the lack of sensory self-attenuation gives us reason to reject this possibility. It may be that sensory self-attenuation requires predictions to be translated into a third-person format. Yet, it is arguable that these predictions need not be conscious; non-conscious, subpersonal signals would appear sufficient to do the job. So if the translation mechanisms yielded accurate although non-conscious third-person representations of the sensory consequences of a movement, sensory attenuation of self-produced movements should be normal. Thus, it seems that to explain both the lack of sensory self-attenuation and the lack of awareness of initiating an action, the abnormality in the predictors should result in a lack of awareness of inaccurate third-person predictions of the consequences of an action. One would then have to explain why faulty third-person predictions and lack of awareness co-occur. Of course, one radical possibility would be to claim that the translation mechanism is not just abnormal but completely knocked out: no predictions, hence no sensory attenuation; no predictions, hence nothing to be aware of!

However these aspects of the account are developed, it seems to account for two components of the phenomenology of alien control: the sense of *passivity* is seen to result from a lack of awareness of having initiated the action, and the sense of *externality* (the agent feels that some external force caused his actions) is seen to result from a lack of sensory self-attenuation. Yet, the model does not explain why this external force is thought of (or experienced) as an agentive force rather than simply a physical force, nor does it explain why the patient experiences the action as having a particular author. At this point, Frith and colleagues take an explanationist line, for they attribute these features to a (faulty) *belief* system. The following nicely summarizes the view:

We suggest that, in delusions of control, the prediction mechanism is faulty and as a consequence self-generated movements are not attenuated and are wrongly classified as externally generated. The patient is not aware

⁷ It has also been suggested (Dierks et al., 1999; McGuire et al., 1996) that a failure of sensory attenuation could be responsible for the verbal hallucinations of those schizophrenic patients who perceive their inner speech as coming from external sources. During verbal hallucinations, the auditory temporal areas remain active, which suggests that the nervous system in these patients behaves as if it were actually processing the speech of an external speaker, since self-generated inner speech is normally accompanied by a mechanism that decreases the responsiveness of primary auditory cortex.

of the predicted consequences of a movement and is therefore not aware of initiating a movement. In parallel, the patient's belief system is faulty so that he interprets this abnormal sensation in an irrational way. (Blakemore et al., 2002, p. 240).

Yet, at least some of the reports seem to suggest that the alien agency aspect of delusions of control is part and parcel of their phenomenology, not merely the result of a further layer of interpretation.

Is it possible to do justice to these reports and to offer an endorsement model that encompasses not just the passivity and externality aspects of the phenomenology of delusions of control but also its alien agency aspect? The simulationist account, to which we now turn, suggests a positive answer.

5. The simulationist account

The simulation account of action-monitoring developed by Jeannerod and colleagues (Daprati et al., 1997; Georgieff & Jeannerod, 1998; Jeannerod, 1999, 2003; Jeannerod & Pacherie, 2004) has a lot in common with the central-monitoring account. Both accounts exploit the idea that the motor control system makes use of internal models, including inverse and predictive models, and comparators. Both accounts also agree that action-control mechanisms and action-awareness mechanisms are importantly connected. What distinguishes the simulation account from the central-monitoring account is its emphasis on the fact that the motor system serves not just to represent one's own actions but also to represent the actions of others.

According to the simulation account, the motor system with its sets of predictors and controllers serves as a simulation engine that constructs motor representations not just of actions the agent is preparing to execute, but also of actions he or she observes someone else performing or simply imagines in either a first-person perspective (imagining oneself acting) or a third-person perspective (imagining someone else acting). Action preparation, action observation, and imagination of action share representations. The evidence for the existence of such shared representations ranges from single-cell recording studies in monkeys, where mirror neurons were discovered that fire both during goal-directed action and observation of actions performed by another individual (see Fogassi & Gallese, 2002, for a review), to functional neuroimaging experiments in humans (see Blakemore & Decety, 2001; Grèzes & Decety, 2001, for reviews), which demonstrate that the neural circuits involved in action execution, action observation, and action imagination overlap.

According to Jeannerod and colleagues (Georgieff & Jeannerod, 1998; Jeannerod, 1999) this shared motor representations mechanism provides a functional bridge between first-person information and third-person information, and hence a foundation for intersubjectivity. At the same time, representations of one's own actions and representations of the actions of others must be disentangled, as must representations of overt actions, whether self-performed or observed, and representations of purely covert actions—i.e., imagined actions in the first or third-person. At the neural level, the overlap between regions activated in these different conditions is only partial. Action preparation, action observation, first-person action-imagining, and third-person action-imagining should be conceived as different modes of simulation, sharing a common core, the shared representations, but also engaging mode-specific processes. For instance, when observing someone else acting, one should inhibit motor output but not, of course, when one prepares to execute an action. Similarly, predictions of the sensory consequences of an action should be used for perceptual attenuation when one prepares to act, but not—or at least not in the same way—when one observes someone else acting. Each mode of sensory simulation has a proprietary set of inhibitory mechanisms for shaping the network involved in the production of motor output and in the analysis of actual and predicted consequences of both overt and covert actions. Activity in non-overlapping regions as well as differences in intensity of activation in the overlapping regions are associated with differences in simulation modes and would provide signals usable for action attribution.

Finally, an action-attribution system monitoring signals from non-overlapping parts of the networks involved in the various simulation modes would be in charge of determining whether a given motor representation refers to a self-produced action or to an action performed by someone else and thus of attributing actions to their source, oneself or another agent, whether it is actually performed or merely imagined. In other words, action attribution would be based on the monitoring of the mode of simulation. Existing neurobiological evidence suggests that the right inferior parietal cortex in conjunction with prefrontal areas may play a crucial role in mode monitoring and self- vs. other-attribution (see Jackson & Decety, 2004, for a review).

In a nutshell then, the simulationist account contains three components that differentiate it from the central-monitoring account:

- (1) Shared representations: prepared actions, observed actions, and imagined actions share motor representations yielded by a common simulation engine.
- (2) Modes of simulation: although they make use of shared representations, the various modes of simulation differ in the way they shape the networks involved in the analysis of the actual and predicted consequences of overt and covert action.
- (3) A ‘who’ system: this system attributes actions to either the self or another agent by monitoring signals specific to the different simulation modes.

Instead of a solipsistic action-monitoring system, simply yielding a “Me/Not-Me” type of answer, we have an inherently intersubjective system yielding a “Me/Another agent” type of answer. The implicit solipsistic assumption of the central-monitoring account is that the predictors are typically engaged when one prepares to act, and therefore that a mismatch between predictions and incoming sensory signals yields an interpretation of these signals as caused externally (not me). In contrast, the simulationist account insists that predictors in the motor system are engaged *both* when one prepares to execute an action *and* when one observes an action. Note also that in standard cases of passive or involuntary movements the predictors within the motor system are *not* engaged. The activation of the predictors therefore yields a presumption of agency, although not necessarily one’s own. The default options are thus Me/Another agent rather than Me/Not Me. To decide between these default options is precisely the job of the ‘Who’ system. The possibility of a non-agentive external physical force becomes a live one only when both of these default options have been eliminated.

When the answer yielded by this system is ambiguous, either because the signals themselves are ambiguous or because the subject is not attending to them, further information may be taken into account to yield a more definite answer. For example, the subject might use information about: (i) the presence or absence of a conscious goal or desired state (intentionality), (ii) the degree of match between the desired state and the actual state (satisfaction), and (iii) the situational salience of other agents (potential source of action).⁸

We propose that the phenomenology of alien control might result from impairments to the mechanisms controlling and/or monitoring the different modes of simulation involved in the ‘Who’ system. Jeannerod (2003) suggests that these impairments could be a consequence of the hypoactivity of the prefrontal cortex known to exist in many schizophrenic patients. Prefrontal areas normally exert an inhibitory control on other areas involved in various aspects of motor and sensory processing. As we have seen, the simulation model assumes that each mode of simulation involves its own set of inhibitory mechanisms for shaping the network involved in the control of motor output and in the analysis of the actual and predicted consequences of both covert and overt actions. The default setting of the mode of simulation would be affected by an alteration of the inhibitory control exerted by the prefrontal cortex, resulting in abnormal activation patterns. In other words, either the shape of the networks corresponding to different representations, and/or the relative intensity of activation in the areas composing these networks, would be altered. As a result, the signals used by the ‘who’ system would be inaccurate and this would give rise to attribution errors.

Through lack of inhibition, some regions may become over-activated. It is notable that an increased activity of the right posterior parietal lobe has been observed in patients with delusions of influence, either at rest (Franck, O’Leary, Flaum, Hichwa, & Andreasen, 2002) or during an action recognition task (Spence et al., 1997). Prefrontal hypoactivity may also result in a loss of distinctiveness of the networks involved in the various simulation modes. The degree of overlap between the representations would increase in such a way that the representations would become indistinguishable. Depending on the way the patterns of activation are altered, the signals used by the ‘who’ system might either be biased toward either self- or other-attribution, or they could simply be ambiguous.

⁸ Pace Wegner (2002), we do not think of this as the primary process of action attribution, but rather as a backup procedure, used when the “who” system does not provide a clear answer.

In the latter case, other cues would have to be used to disambiguate between self- and other-agency. For instance, in the experiment by Daprati et al. (1997), where subjects had to decide whether a hand they saw executing a movement was theirs or not, only one hand was shown at a time. The experimental situation therefore privileged self-attribution responses because it always referred to the patient as the putative agent of the action. Indeed, as long as the movement performed by the hand shown was of the same type as the movement performed by the subject's hand, schizophrenic patients with first-rank symptoms tended to systematically self-attribute the hand they saw. In contrast, in a later experiment (van den Bos & Jeannerod, 2002), where the subject's hand was shown along with another hand, schizophrenic patients with first-rank symptoms tended to misattribute the hand more frequently to the other than to themselves. Thus, in an ambiguous situation, the match between visually perceived movement and intended movement functions as a cue for self-attribution, but the situational salience of another agent can override this cue.

In contrast to the central-monitoring model, the simulation model accounts not just for the elements of passivity and externality in the phenomenology of alien control, but also for the sense of alien agency. The motor system represents the actions of others to the same extent that it represents one's own. The role of the action-attribution mechanism is therefore to disentangle situations in which the activation of the system corresponds to the representation of one's own actions from situations in which it represents the actions of others. This is done by monitoring signals specific to each condition. When the signals are unambiguous (whether they are accurate or not), one experiences either a sense of self-agency or a sense of alien agency for the action. But even when the signals are ambiguous, the ambiguity is between self-agency and alien agency.

6. Conclusion

We began this paper with the distinction between endorsements and explanationist accounts of delusion: endorsement theorists hold that the content of the delusion in question is encoded in the patient's perceptual experience, explanationists hold that although the delusion is grounded in an unusual experience of some kind, the content of the delusion results from the patient's attempt to explain this unusual experience. Our goal in this paper has been to develop an endorsement-based account of the delusion of alien control. We distinguished four aspects of the content of alien control delusions: passivity, externality, agency, and particularity. We saw that Frith's central-monitoring account gives us a way to understand how it is that a person could experience their willed actions as passive and external. But Frith's account does not take an endorsement approach to either the agency or the particularity components of alien control. To develop an endorsement account of alien agency we turned to Jeannerod's simulationist account of action monitoring, arguing that the inherent inter-subjectivity of his model gives us a way in which a person could experience their own actions as the actions of someone else.

We finish with some outstanding questions. First, we still need to account for particularity: why do patients with alien control delusions believe that particular agents are controlling their actions? Is this also encoded in their experience, or do we have to appeal to explanationist principles at this point? Second, there is what Gallagher (2004) calls the problem of specificity. Why do patients with alien control regard only some of their actions as under the control of other agents? Because this issue is orthogonal to the debate between endorsement theorists and explanationist theorists we have left it to one side here, but it is clearly a pressing one for accounts of the delusion of alien control, no matter what form they take.

References

- Barresi, J. U., & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, *1119*, 107–154.
- Bayne, T. J., & Pacherie, E. (2004a). Bottom-up or top-down?: Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry, and Psychology*, *11*(1), 1–11.
- Bayne, T. J., & Pacherie, E. (2004b). Experience, belief and the interpretive fold. *Philosophy, Psychiatry, and Psychology*, *11*(1), 81–86.
- Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.
- Blakemore, S.-J., & Frith, C. D. (2003). Self-awareness and action. *Current Opinion in Neurobiology*, *13*, 219–224.
- Blakemore, S.-J., Oakley, D. A., & Frith, C. D. (2003). Delusions of alien control in the normal brain. *Neuropsychologia*, *41*, 1058–1067.
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2000). Why can't we tickle ourselves? *NeuroReport*, *11*, 11–16.

- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Science*, 6(6), 237–242.
- Breen, N., Caine, D., & Coltheart, M. (2001). Mirrored-self misidentification: Two cases of focal onset dementia. *Neurocase*, 7, 239–254.
- Breen, N., Caine, D., Coltheart, M., Hendy, J., & Roberts, C. (2000). Towards an understanding of delusions of misidentification: Four case studies. *Mind and Language*, 15, 75–110.
- Butler, P. V. (2000). Diurnal variation in Cotard's syndrome (copresent with Capgras delusion) following traumatic brain injury. *Australian and New Zealand Journal of Psychiatry*, 34, 684–687.
- Campbell, J. (1999). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, 82, 609–625.
- Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., et al. (1997). Looking for the agent. An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71–86.
- Davies, M., Aimola Davies, A., & Coltheart, M. (2005). Anosognosia and the two-factor theory of delusions. *Mind and Language*, 20(2), 209–236.
- Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, and Psychology*, 8, 133–158.
- Dierks, T., Linden, D. E. J., Jandl, M., Formisano, E., Goebel, R., Lanferman, H., et al. (1999). Activation of the Heschl's gyrus during auditory hallucinations. *Neuron*, 22, 615–621.
- Ellis, H. D., & Young, A. W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157, 239–248.
- Fleminger, S. (1992). Seeing is believing: The role of preconscious perceptual processing in delusional misidentification. *British Journal of Psychiatry*, 160, 293–303.
- Fogassi, L., & Gallese, V. (2002). The neural correlates of action understanding in non-human primates. In M. I. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 13–35). Amsterdam: John Benjamins.
- Franck, N., O'Leary, D. S., Flaum, M., Hichwa, R. D., & Andreasen, N. C. (2002). Cerebral blood flow changes associated with Schneiderian first-rank symptoms in schizophrenia. *Journal of Psychiatry and of Clinical Neuroscience*, 14, 277–282.
- Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*, 17, 631–648.
- Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove E. Sussex: Lawrence Erlbaum Associates.
- Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000a). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London B*, 355, 1771–1788.
- Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000b). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, 31, 357–363.
- Gallagher, S. (2000). Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. In D. Zahavi (Ed.), *Exploring the self: Philosophical and psychopathological perspectives on self-experience* (pp. 203–239). Amsterdam and Philadelphia: John Benjamins.
- Gallagher, S. (2004). Neurocognitive models of schizophrenia: A neurophenomenological critique. *Psychopathology*, 37, 8–19.
- Georgieff, N., & Jeannerod, M. (1998). Beyond consciousness of external reality. A 'Who' system for consciousness of action and self-consciousness. *Consciousness and Cognition*, 7, 465–477.
- Gerrans, P. (2002). Multiple paths to delusion. *Philosophy, Psychiatry, and Psychology*, 9(1), 65–72.
- Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation and verb generation of actions: A meta-analysis. *Human Brain Mapping*, 12, 1–19.
- Haggard, P., & Magno, E. (1999). Localising awareness of action with transcranial magnetic stimulation. *Experimental Brain Research*, 127, 102–107.
- Heider, F., & Simmel, F. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243–259.
- Hohwy, J., & Rosenberg, R. (2005). Unusual experiences, reality testing, and delusions of alien control. *Mind and Language*, 20(2), 141–162.
- Jackson, P. L., & Decety, J. (2004). Motor Cognition: A new paradigm to investigate self-other interactions. *Current Opinion in Neurobiology*, 14, 259–263.
- Jeannerod, M. (1999). To act or not to act: Perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology*, 52A, 1–29.
- Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain Research*, 142, 1–15.
- Jeannerod, M., & Pacherie, E. (2004). Agency, simulation, and self-identification. *Mind and Language*, 19(2), 113–146.
- Joseph, A. B. (1986). Cotard's syndrome in a patient with coexistent Capgras' syndrome, syndrome of subjective doubles, and palinopsia. *Journal of Clinical Psychiatry*, 47, 605–606.
- Langdon, R., & Coltheart, M. (2000). The cognitive neuropsychology of delusions. In M. Coltheart & M. Davies (Eds.), *Pathologies of belief* (pp. 183–216). Oxford: Blackwell Publishers.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8, 529–566.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential). The unconscious initiation of a freely voluntary act. *Brain*, 102, 623–642.
- Maher, B. (1974). Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30, 98–113.
- Maher, B. (1988). Anomalous experience and delusional thinking: The logic of explanations. In T. F. Oltmans & B. A. Maher (Eds.), *Delusional beliefs* (pp. 15–33). New York: Wiley.
- McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine*, 26, 29–38.

- Mellors, C. S. (1970). First-rank symptoms of schizophrenia. *British Journal of Psychiatry*, *117*, 15–23.
- Pacherie, E. (2001). Agency lost and found. *Philosophy, Psychiatry, and Psychology*, *8*(2–3), 173–176.
- Pacherie, E. (forthcoming). Perception, emotions and delusions: The case of Capgras' delusion. In T. Bayne & J. Fernández (Eds.) *Delusion and self-deception: Affective influences on belief-formation*.
- Spence, S. A. (2001). Alien control: From phenomenology to cognitive neurobiology. *Philosophy, Psychiatry, and Psychology*, *8*(2/3), 163–172.
- Spence, S. A., Brooks, D. J., Hirsh, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain*, *120*, 1997–2011.
- Stone, T., & Young, A. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language*, *12*, 327–364.
- van den Bos, E., & Jeannerod, M. (2002). Sense of body and sense of action both contribute to self recognition. *Cognition*, *85*, 177–187.
- Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wolfe, G., & McKenzie, K. (1994). Capgras, Frégoli and Cotard's syndromes and Koro in folie a deux. *British Journal of Psychiatry*, *165*, 842.
- Young, A. W., & Leafhead, K. M. (1996). Betwixt life and death: Case studies of the Cotard delusion. In P. W. Halligan & J. C. Marshall (Eds.), *Method in madness: Case studies in cognitive neuropsychiatry*. Hove: Psychology Press.