PERCEPTION, EMOTIONS AND DELUSIONS:

REVISITING THE CAPGRAS' DELUSION[1]

Elisabeth Pacherie
Institut Jean Nicod UMR 8129
Département d'Etudes Cognitives
Ecole Normale Supérieure
29, rue d'Ulm
75005 Paris – France
pacherie@ens.fr

This is a DRAFT of a paper that is forthcoming in T. Bayne & J. Fernandez (eds.), *Delusions and Self-Deception: Affective Influences on Belief Formation*. Psychology Press.

---

1. Introduction

The study of delusions has important implications for understanding the role played by affective processes on the road from experience to belief. It can also shed light on the forms of modularity these processes manifest. There are at least two different ways in which emotional processes may be relevant to the etiology of delusional beliefs. First, current models of delusion converge in proposing that such beliefs are based on unusual experiences of various kinds. These unusual experiences are thought to include affective or emotional experiences. For example, it is argued that Capgras' delusion (the belief that a known person has been replaced by an impostor) is triggered by an abnormal affective experience in response to seeing a known person (Ellis & Young, 1990).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 et al, 2001). In

delusions of persecution, the experiential component could be an over-sensitivity to other people's disingenuous expressions of emotions (LaRusso, 1978; Davis & Gibson, 2000). Experience-based proposals have been provided for a number of other delusions (Stone & Young, 1997; Maher, 1988; Davies *et al.*, 2001; Langdon & Coltheart, 2000).

There is also a second way in which emotional processes may be relevant to the aetiology of delusional beliefs, for one must also explain why these abnormal experiences give rise to delusional beliefs and why these delusional beliefs are "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). 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. I call these the endorsement and explanationist models respectively (see Bayne & Pacherie, 2004a, 2004b). Emotional factors may also contribute to such an explanation.

In the present paper, I will focus on Capgras' delusion. Three reasons motivate this choice. First, central to this delusion is an emotion or rather a pair of emotions not so often discussed in philosophical circles, namely the feelings of familiarity and unfamiliarity.[2] Second, there now exist in the literature several proposals as to how the feeling of familiarity one normally experiences when encountering known people is generated and what would cause the anomalous experience in Capgras' patients. These proposals strongly suggest that the system underlying the feelings of familiarity and unfamiliarity is modular. Third, at least one of these proposals also suggest a way of fleshing out an endorsement account of Capgras' delusion that

exhibits an important explanatory link between the modularity of the underlying emotional system and the persistence of the delusional belief.

I will start by saying a little more on the distinction between endorsement and explanationist models of delusions. I will then discuss some recent models of visual face processing and the way they account for the generation of affective responses to familiar faces. I will argue that despite disagreeing on how exactly emotional responses to faces are generated, they all support the view that the system that generates them is modular. I will also argue that at least one of these accounts provides strong support for an endorsement account of Capgras' delusion. Finally, I will discuss the role affective factors may play in explaining why the delusional belief once formed is maintained and will argue that there is an important link between the modularity of this emotional system and the persistence of the delusional belief.

2. Two experiential routes to delusion

Let us consider the distinction between endorsement and explanationist models in more detail.[3] According to endorsement models, the experience comprises the very content of the delusion, such that the delusional patient simply believes —that is, doxastically endorses—the content of their experiential state, or at least something very much like the content of their experiential state; whereas on explanationist accounts the delusion is adopted in an attempt to explain an unusual experience.[4] Experience-based accounts of delusions involve (at least) two components: (i) an explanation of the delusional patients *experiential* state; and (ii) an explanation of the delusional patient's *doxastic* state. 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 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 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?[5]. 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. I 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 and/or affective: 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).

Returning now to Capgras' delusion, we can see that an endorsement account of this delusion would hold, for example, that the 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). The difficulty for such an account would be to explain how experience can represent the person in front of you not just as unfamiliar but as an impostor of your spouse. In contrast, according to the explanationist, the 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. The main difficulty the explanationist confronts lies in explaining why the person forms such an implausible explanation of their unusual experience. In addition, both accounts face the problem of explaining why the delusional belief is maintained. General knowledge tells us that impostor stories are unlikely in most instances. Why would someone want to impersonate your relatives? The testimony of others, whether family members, friends or doctors, goes against the impostor theory. Why don't Capgras' patients listen to them?

Before turning to this latter question, let us discuss recent cognitive models of the experiential factor in Capgras' delusion and see whether they support an explanationist or an endorsement account of the delusion.

3. The experiental factor in Capgras' delusion

The best-known model of Capgras' delusion is the "two-route model of face recognition" first proposed by Bauer to account for propopagnosia and adopted by Ellis and Young (1990). It is a robust finding (Tranel, Fowles and Damasio, 1985; Ellis, Quayle & Young, 1999) that when shown both familiar and unfamiliar faces, normal subjects produce larger autonomic arousal as measured by skin conductance responses (SCRs) to familiar faces than to unfamiliar faces. This autonomic response has been interpreted as a form of covert recognition. Bauer (1984,

1986) discovered that prosopagnosic patients, despite being unable to consciously recognize previously known faces, still produced a larger SCR to them than to previously unfamiliar faces. To account for this finding, he proposed a two-route model of face recognition. On this model, face recognition involves two information-processing pathways: a ventral visuo-semantic pathway, that constructs a visual image encoding semantic information about facial features and is responsible for overt recognition, and a dorsal visuo-affective pathway responsible for covert autonomic recognition and for the specific affective response to familiar faces (the feeling of familiarity). In prosopagnosia, the visuo-semantic pathway would be damaged, which would account for the patient's inability to recognize faces, while the visuo-affective pathway would remain intact, which would explain why they retain a covert autonomic recognition of familiar faces. Ellis and Young proposed that Capgras' syndrome might be a mirror image of prosopagnosia, with the affective pathway damaged but the visuo-semantic pathway intact.[6] They predicted that Capgras' patients would fail to produce the normal, higher SCR to familiar compared with unfamiliar faces. This prediction was borne out in two studies by independent groups (Ellis & al., 1997; Hirstein & Ramachandran, 1997).

When looking at familiar faces, Capgras' patients would have highly unusual experiences. For instance, when looking at their spouse's face, the spouse would be recognized as looking like one's spouse, but the normal feeling of familiarity would be absent (and, indeed, a feeling of unfamiliarity would be present). The fact that Capgras' delusion is usually restricted to close relatives can be explained if we assume, quite plausibly, that the affective response to close relatives is normally higher than to familiar but emotionally neutral persons such as one's grocer or mailman. The perceptual-affective dissonance resulting from the absence of the expected emotional response would thus be much greater for close relatives.

This original two-route model was proposed as both a neuroanatomical model and a cognitive model. But these two aspects of the model can be evaluated separately. Indeed, the plausibility of Bauer's neuro-anatomical conception has been questioned (Breen, Caine & Coltheart, 2000; Hirstein & Ramachandran, 1997; Tranel, Damasio, & Damasio, 1995), but the idea of a dissociation between overt recognition and covert affective discrimination has generally been retained.

What remain unclear in this original two-route model, however, is at what stage the two cognitive pathways bifurcate. Bruce and Young (1986) developed a single route model of face processing involving a series of sequential stages. In the first stage of their model, what they call "structural encoding", the seen face is encoded using "descriptions" that are viewer-centred. These structural descriptions can then be analysed independently for expression, facial speech, information about sex, age, and race, and identification. In the second stage of their model, the seen face, if it is familiar, will contact with its stored representation in the Face Recognition Units (FRU). Familiar faces will then activate information held at the third stage of the model, the Person Identity Node (PIN) that contains semantic and biographical information about the person and can be accessed by ways other than face recognition. At the fourth and final stage, the person's name, which is stored independently of their biographical details will be retrieved. In a two-route model of face recognition, the pathway described in Bruce and Young's model would correspond to the explicit recognition route. The question then is at what stage the autonomic recognition pathway bifurcates from this explicit recognition pathway. The original neuro-anatomical interpretation of the dual-route model seemed to require a very early bifurcation given the early anatomical separation of the dorsal and ventral pathways. However, as Breen *et al.* (2000) point out in their critical discussion of

this model, it is implicit in the arguments of both Bauer and Ellis and Young that the affective response must be attached to a particular face and hence that the face must have been at least implicitly recognized. A very early bifurcation would then require a reduplication of the face recognition stage. Breen *et al.* (2000) see this reduplication as unparsimonious and the anatomical arguments for it as problematic. Instead, they propose a modified dual route model. They argue that there is only a single face recognition stage but they posit two pathways subsequent to recognition, one leading to the processing of semantic and biographical information about the person, the other to the system responsible for generating affective responses to familiar faces.

In a recent paper, Ellis and Lewis (2001) endorse Breen *et al.*'s modified dual route model, but they introduce an important modification of their own. In Breen's *et al.* model, the person identity nodes and the affective response module are directly connected. Ellis and Lewis argue against such a direct connection; they point out that if this connection existed damage to either the pathway from the face recognition units to the person identity nodes or the pathway from the face recognition units to the affective response module could be circumvented, in which case the double dissociation between explicit recognition and implicit autonomic discrimination seen between prosopagnosics and Capgras' patients would not be explained.

Instead, Ellis and Lewis argue that the two modules are not directly connected but that their outputs each feed into an integrative device where they are recombined to yield a single percept. This would provide the necessary data for the person to be identified by comparing the joint information representing recognition and the affective response against a stored and therefore expected representation. In Capgras' patients, where the affective response module is impaired, this device would detect a mismatch between the expected and the actual

affective response, yielding a misidentification of, say, the spouse as someone else, someone looking like her and yet unfamiliar. Another possible motivation for positing such an integrative device would be to account for the transition from the unconscious autonomic response to the conscious feeling of familiarity. One may think that to be conscious of an affective response, one must bind it to a specific object. The fact that typically patients with prosopagnosia show a normal heightened SCR to previously known faces but fail to experience a conscious feeling of familiarity may be taken as evidence for this idea.

Although this modified dual-route model of face perception provides an account of the experiential anomaly in Capgras' delusion, it is unclear whether it supports an endorsement over an explanationist account of the delusion. This would seem to depend on how exactly the integrative device compares incoming with stored information and how it interprets discrepancies. This would seem also to depend on whether this comparison is integral to the face processing system or is carried out in a separate center possibly subject to top-down influences. But as Ellis & Lewis acknowledge these are issues on which work still needs to be done.

There is however a very recent proposal by Hirstein (2005) that would seem to more clearly tip the balance in favor of an endorsement account. Hirstein's is also a dual-route account but it builds on a different model of face perception developed by Haxby and colleagues (Hoffman & Haxby, 2000; Haxby *et al.*, 2000). This model is both a neuro-anatomical and a cognitive model of face processing. Working primarily from functional brain imaging studies, Haxby *et al.* (2000) found that the temporal lobe contains two face-processing streams, a medial temporal pathway involving the fusiform gyrus and a lateral temporal pathway involving the superior temporal sulcus. They hypothesized different functional specializations

for the two pathways and suggested that the medial pathway specializes in representing the invariant aspects of faces that underlie identity while the lateral pathway produces representations of the changeable aspects of faces. Their model distinguishes a core face processing system and an extended system. The core system is comprised of the inferior occipital gyri, the lateral fusiform gyrus and the superior temporal sulcus. The inferior occipital gyri would be involved in the early visual analysis of facial features and would provide input to both the lateral fusiform gyrus specializing in the representation of the invariant aspects of faces and the superior temporal sulcus specializing in the representation of changing aspects of faces. These representations would then be used by the extended system in a variety of tasks. Representations of invariant aspects of faces would underlie explicit recognition of unique identity, while representations of variable aspects of faces would provide input to various systems specialized in the processing of eye gaze direction, speech-related movements, or the facial expression of emotions.

O'Toole *et al.* (2002) propose an interesting modification of Haxby's model to accommodate psychological evidence that information for identifying a human face can be found both in the invariant structure of features and in idiosyncratic movements and gestures. More precisely, psychological evidence indicates that dynamic information contributes more to recognition under non-optimal viewing conditions — such as poor illumination or low image resolution — than invariant information does, even when the latter is available. Moreover, the contribution of dynamic information increases as a viewer's experience with a face increases. In particular, familiarity with a face allows one to extract its dynamic signature, i.e. the characteristic or idiosyncratic movements a particular face makes such as the distinctive smile or the way of expressing surprise a person may have. Of course the better one knows a person, the more reliable this dynamic signature becomes for identification. O'Toole *et al.*

therefore propose that the STS could be involved in the identification of dynamic facial signatures and that this information could be used, together with representations of invariant features, in the identification of familiar faces.

In Hirstein's interpretation, Haxby's model provides the basis for a mind-reading theory of Capgras' delusion and of delusions of misidentification more generally. According to his proposal, the medial temporal pathway produces 'external representations', i.e. representations of the outward appearance of a person's face. In contrast, the lateral temporal pathway would yield information relevant to 'internal representations' of a person, representations of what their mind is like. Perception of changeable aspects of faces provides information about another's person current state of mind. In particular, eye-gaze can inform us about what a person is attending to, what her current interests and intentions are, while facial expression can inform us about the person's current emotional state. Hirstein further suggests that Haxby's two routes could correspond to the processes that are doubly dissociated in prosopagnosia and Capgras' syndrome. Thus, in proposopagnosia, the medial temporal pathway would be damaged, and the patient would not be able to produce a representation of the outward appearance of the seen face. Conversely, in Capgras' delusion the lateral temporal pathway would be dysfunctional and would either fail to produce an internal representation or produce one that is not the same one that the patient has always used.

Consistent with the hypothesis that in Capgras' syndrome, the normal functioning of the STS would be impaired is the fact that the patient DS examined by Hirstein & Ramachandran (1997) was defective in processing gaze and unable to tell whether or not a face in a photograph was looking toward him. In contrast, DS was not impaired in the recognition of facial expressions of emotions. However, Hirstein and Ramachandran tested him only with

digitised images of models posing basic emotions such as fear, anger or happiness. One is left to wonder whether DS would have performed as accurately if shown dynamic displays of emotional expressions or less prototypical facial expressions of emotions. Another study by Breen *et al.* (2002) investigating patient MF with a delusion of misidentification resembling Capgras' delusion showed that MF was unable to identify the facial expressions of anger, disgust and fear.[7] Interestingly, MF was also unable to match faces across expressions. In other words, if shown two pictures of either the same or two different people each having a different expression, he was almost at chance at telling whether or not they were pictures of the same person. To explain this result, one may speculate that when the system specialized in representing the variable aspects of faces is damaged, the other system specialized in the representation of invariant aspects would tend to overshoot and to treat changeable aspects as invariant.

Further evidence in favor of Hirstein's view comes from the fact that some comments of patients with Capgras' syndrome refer straightforwardly to psychological differences between the original and impostor. Thus, one patient "became convinced that her husband's personality had changed" (Frazer & Roberts, 1994). Another claimed that there were two doctors looking after him: the first consultant (who he called John Smith) was 'a nice Bloke', whereas the second (a Dr. J. Smith) was someone who was distant and aloof'" (Young, 1998: 39).

Often, the supposed change of personality is for the worse. Adolphs (1999) suggests that when confronted with ambiguous expressions of emotions or complex blends of emotions in an other person's face, people would judge their emotional state from their facial expression by reconstructing in their own brain a simulation of what the other person might be feeling; in other words they ask themselves how they would feel if they were making the same facial

expression. Capgras' patients often exhibit paranoid tendencies and a suspicious cast of mind. One may speculate that when they are confronted with facial expressions that are for them 'ambiguous' because of their impairment in the emotional processing of faces, Capgras' patients would use their own emotional system in simulation to understand others and would project their own negative states of mind on those surrounding them. This might explain why, in contrast to patients with amygdala damage who show a positive bias in judging faces (Adolphs, 1999), Capgras' patients tend to give negative ratings to faces. This might also explain why they tend to see people around them as ill-intentioned. Thus, one patient described by Butler (2000) accused the nursing staff of having murdered members of his family. When he interacted with his father, this patient "minutely examined [his father's] face before accusing him of being a criminal double who had taken his father's place" (Butler, 2000, p.685).

It is as yet unclear what evidential support Hirstein's interpretation of Haxby's model has as many of the predictions it yields remain untested. For instance, there are no systematic data as yet regarding possible impairments of patients with Capgras' syndrome in the processing of emotional expressions in faces. It is also somewhat unclear how exactly this model would account for the presence of normal SCRs to faces in patients with prosopagnosia and their absence in patients with Capgras' syndrome. Relying on evidence that both the medial and the lateral temporal pathway connect with the amygdala (Carmichael & Price, 1995), Hirstein suggests that both routes may be able to produce SCRs, where the fusiform gyrus would be involved in producing an SCR to the sight of a familiar face and the superior temporal sulcus an SCR to emotional facial expressions. But the presence of normal SCRs in prosopagnosic patients and their absence in Capgras patients suggests that the lateral temporal pathway contributes more to the production of specific SCRs to familiar faces.[8] Building on O'Toole et

al.'s idea that the lateral temporal pathway is involved in the identification of dynamic facial signatures, one may speculate that the production of SCRs depends at least in part on the identification of these signatures.

This is important because of an objection to Hirstein's interpretation of Haxby's model that may naturally come to mind.[9] One may agree that an impairment of the lateral temporal pathway would lead one to miscontrue facial expressions of emotions. But why should mistaking one's father expression of concern for an expression of anger lead one to form the belief that the person in front of you is not your father but an impostor rather than the less farfetched belief that your father is angry at you or in a bad mood, perhaps for some reason you can't fathom? There are two complementary lines of answer to this objection one may think of. The first is that although indeed a single or a few experiences of discrepancy between one's usual representation of the personality of someone and one's immediate experience of their present state of mind might not lead one to form the Capgras delusion, systematic discrepancies experienced over a period of time may well lead one to such a belief. The second line of answer is that impairment to the lateral temporal pathway would disrupt not just the correct reading of expressions of emotions but also the identification of the dynamic signature of the face of the person. Someone with such an impairment would not just mistake his father's expression of concern for one of anger, but would also see this expression of anger as different in its dynamics from his father's ordinary way of facially expressing anger. More generally, the way this person animates his face would appear discrepant with the way your father animates his face when experiencing various emotions.

Thus, one may tentatively conclude that Hirstein's story about the experiential basis of Capgras' delusion, if true, would enhance the plausibility of the endorsement account. As

Hirstein points out, according to this view, "the Capgras' patient is looking at someone who visually resembles his father, but who appears to have a different mind, a different personality, with different dispositions to do different things. This is exactly what an impostor is, and this is exactly the experience one would have looking at an impostor" (Hirstein, 2005: 133). If the content of the patient's experience is as Hirstein describes it—an experience of the visually presented person as unfamiliar—and not just an experience as of a person that looks like one's father but lacks the feeling of familiarity that normally accompanies this visual experience, then the impostor belief, far from being a fantastic explanation of the abnormal experience, would be a direct reading of it.

4. The modularity of familiarity

If we take as our guide the set of criteria proposed by Fodor (1983) for modularity, it seems pretty obvious that the processes through which feelings of facial familiarity are generated qualify as modular. To convince the sceptical reader, let us examine each of these criteria in turn.

Dedicated neural architecture

Although it is still debated what the exact neuro-anatomical pathways involved in the two routes to face recognition are, all the models described in the previous section agree that there are dedicated neural pathways for explicit recognition and for affective processing of faces.

Specific breakdowns

Capgras' delusion is a clear example of a specific breakdown and its double dissociation with prosopagnosia is a further sign of its specificity. One may add that although Capgras' delusion

often occurs in a psychiatric setting, most typically in subjects diagnosed as suffering from paranoid schizophrenia, over one third of the documented cases of Capgras' syndrome have occurred in conjunction with traumatic brain damage, with lesions predominantly in the temporal cortex, which suggests that the syndrome has an organic basis (Signer 1994).

Mandatory operation

When seeing a known face, the feeling of familiarity is automatically generated. Indeed, if it were not, there would be no reason why it would be disturbing to see the face of a well-known person without at the same time experiencing a feeling of familiarity.

Fast operation

The feeling of familiarity is experienced quickly. It is typically simultaneous with the conscious recognition of the face and may even precede it.

Shallow output

This is somewhat more controversial. On two-route models such as Ellis and Young's or Breen *et al.*'s, the immediate output of the affective processing of faces is indeed shallow and takes the form of a 'glow' of arousal. On Hirstein's view, the output would be something more complex, namely an 'internal representation', a representation of the way a faces is animated and of what this reveals about the personality of the person one sees together with a sense of familiarity (or lack thereof).[10]

Inaccessibility

We have no conscious access to the stages through which the feeling of familiarity is generated. Indeed, Capgras' patients who admit that the person in front of them looks, say,

just like their son but deny he is are typically at a loss to explain what makes them think this person is not their son. If pushed, they might point to some minor detail such as the way the "impostor" ties his shoelaces, the size of his eyes or the texture of his skin. For instance, one patient remarked, "there's someone like my son's double which isn't my son. I can tell my son because my son's different but you have to be quick to notice it." (Young *et al.*, 1993: 696; see also Merrin and Silberfarb, 1976).

Informational encapsulation

In the same way that measuring the two arrows in the Müller-Lyer illusion won't make you see them as of equal length, being told by someone you trust that the person in front of you is someone you know (or don't know) won't restore a feeling of familiarity or sense of their personality if you do not experience it in the first place (or won't suppress it if you experience it). Indeed, Capgras' patients seem quite impervious to all the evidence they may be given that the person they take to be an impostor of, say, their father is actually their father. Patient DS, studied by Hirstein and Ramachandran (1997), provides an intriguing illustration of this point. To try to get him rid of his delusional belief, his father thought of the following trick. One day he walked into his son's room and announced that he had sent away the impostor to China and was his real father. DS's delusion seemed to abate slightly as the result of this unorthodox procedure, but, as his father himself acknowledged, although DS seemed to have accepted him as his father intellectually, he had not done so emotionally.

Domain-specificity

Here things get a bit tricky. Of course, other things than just faces can produce feelings of familiarity. Animals, especially pets, and various kinds of inanimate objects (one's worn-out philosophical armchair, one's favorite sweater), can also give rise to feelings of familiarity.

Even if we restrict ourselves to people, it's not just the sight of their face that can produce feelings of familiarity, the sound of their voice can as well. Indeed, although the most common form of Capgras' delusion is for people, there are also documented cases of Capgras' delusion for animals and inanimate objects that may or may not coexist with Capgras' delusion for persons (see Berson (1983) and Edelstyn & Oyebode (1999) for reviews). Similarly, although Capgras' delusion is usually visual, there at least three documented cases of patients, who although blind, suffered from Capgras-type delusions (Reid *et al.*, 1993; Rojo *et al.*, 1991, Signer *et al.*, 1990) suggesting that there could be an auditory form of the delusion.

In light of this, it would be improper to say that the affective system that generates the sense of familiarity is domain-specific in the sense that it only takes as input visual stimuli from faces. If we temporarily restrict ourselves to Capgras' delusion in the visual modality, one thing to point out is that it is perhaps unduly restrictive to call the dual route models discussed in the previous section models of face recognition. A recent brain-imaging study shows that in humans both the fusiform gyrus and the superior temporal sulcus respond similarly to faces and animals (Chao *et al.*, 1999). Another study (Gauthier *et al.*, 1999) also indicates that the brain areas thought to be critical for face perception are also specifically activated by non-face objects for expert subjects, i.e. subjects — such as bird-watchers or car experts — who can categorize such objects at the individual level rather then at the more general family level. Thus, an alternative characterization of the so-called face recognition system would be as a system specialised in the recognition of objects at very specific level – typically the level of individuals. Faces are recognized most often at this very specific level (Jules vs. Jim), whereas objects and animals are typically recognized in a less specific manner, as a cat vs. a dog, as a chair vs. a table. Pets and personal belongings may, however, be exceptions. A cat

owner will recognize his cat not just as a cat but as this very specific cat. A carpenter will recognize his tools not just as tools or as hammers and saws and screwdrivers but as this particular hammer, saw or screwdriver. Interestingly, subjects with Capgras' delusion for animals or objects typically have the delusion for objects that are significant for them and that they can recognize at the individual level, for instance a favorite cow in the case of a farmer or his tools for another subject. The same remarks apply to the auditory form of Capgras' delusion. Known cases are only for voices that can be recognized at the individual level, not for other types of auditory objects that we typically categorize at a less fine-grained level. Thus, although this may go beyond the sense in which Fodor intended the notion of domain-specificity, the affective system that generates the sense of familiarity may still be considered domain-specific insofar as it takes as its inputs specific types of descriptors, such as face recognition units, voice recognition units and other very fine-grained recognition units yielded by earlier perceptual analysis processes.

Why insist that familiarity be modular? In the previous section, I presented models of face-processing that suggest that there is more to the feeling of familiarity (or unfamiliarity) than simply a glow of arousal (or lack thereof). In particular, according to Hirstein's view, the feeling of familiarity comes attached to an object that is less the face understood simply as an external representation than the face understood as a window into the personality of someone, what Hirstein calls an 'internal representation'. In other words, the feeling of familiarity does not so much attach to a face simply conceived as a particular configuration of physical features than to the person behind the face. I suggested that this way of conceiving of the experiential basis of Capgras' delusion, if correct, would enhance the plausibility of the endorsement account.

But a further condition on the plausibility of this endorsement account is that the processes through which the feeling of familiarity — understood in the substantive sense just delineated — is generated be to some extent modular. If, as Fodor insists, modularity is what demarcates perception from cognition or, as I would say, experience from belief, an endorsement account needs to secure the modularity of familiarity. Otherwise, an explanationist could well argue that the so-called feelings of familiarity or unfamiliarity, when they are taken to involve more than a glow of arousal or lack thereof, are not actually experiences but rather interpretations or explanations of more primitive experiences. To put it differently, the modularity of feelings of familiarity or unfamiliarity is consistent with the explanationist position as long as these feelings are thought to involve no more than a glow or arousal or lack thereof. But when more is built into these feelings, their modularity does not sit so well with explanationist accounts. The *raison d'être* of explanationist accounts is to fill the gap between the contents of the experience and the contents of the delusional belief. If there is no gap to be filled, they become superfluous.

As we will see in the next section, there is still another reason one may want to secure the modularity of feelings of familiarity, for this modularity may contribute to explaining why delusional beliefs are not just formed but also firmly maintained.

5. From delusional experience to delusional belief

Even if we accept an endorsement account of Capgras' delusion and think the delusional belief inherits its content from the delusional experience, we still have to explain why the delusional belief is maintained, why Capgras' patients fail to take account of their background beliefs and of the testimony of others. General knowledge tells us that impostor stories are

unlikely in most instances. Why would someone want to impersonate your relatives? Family members, friends, doctors insist that this person is your wife and not an impostor. Why don't Capgras' patients listen to them?

A number of proposals have been made to explain why the belief once formed is maintained tenaciously in spite of contrary evidence. Some of these proposals postulate biases of various kinds, biases in probabilistic reasoning — a tendency to jump to conclusions — and attributional biases —a tendency to explain their experience in terms of external rather than internal causes — (Kaney & Bentall, 1989) or alternatively an observational bias (Stone & Young, 1997). Others suggest a failure of inhibition of prepotent doxastic response, i.e. an impairment in reality-testing, (Davies *et al.*, 2001) or an impairment in global consistency-checking procedures (Hirstein and Ramachandran, 1997).

The problem with these explanations is that they make unwanted predictions. They imply that Capgras' patients would develop delusional beliefs whenever they have any kind of unusual experiences, such as visual illusions. But Capgras' delusion, like other monothematic delusions, tends to be relatively circumscribed. In domains other than that of their delusions, the reasoning skills and cognitive behavior of Capgras' patients appear by and large normal. What needs explaining is therefore not just why subjects fail to appropriately check their delusional belief, but why the failure is localized.

There are three kinds of checking procedures one may use to decide whether or not an observational belief should be accepted. One could check the belief: (1) by enlarging one's set of observations; (2) by using background knowledge and general encyclopaedic knowledge;

and (3) by relying on the testimony of others. Why is it that the delusional belief is not refuted using these procedures?

Further observations.

In Capgras' delusion, the delusional belief is grounded in the unusual experiences the subject has when looking at his or her relatives. Further observation would mean to keep looking to see if the feeling of familiarity is restored. The problem though is that if the affective route to face recognition is damaged, the checking procedure will keep giving the same negative result. As Bermúdez (2001) and Hohwy and Rosenberg (2005) point out, the recurrence of the experience will result in a reinforcement of the belief rather than its rejection. To this it may be objected that if the damage is to a visual pathway, using another modality would restore the feeling of familiarity. Indeed, the patient DS studied by Hirstein and Ramachandran (1997) who regarded his father as an impostor when in his presence never treated him as such when talking to him on the phone. The problem though is that in humans the visual modality tends to dominate over other sensory modalities. Thus, when talking to his father face to face, the conflict between the visual and the auditory modality would typically be resolved in favor of the visual modality. As Hohwy and Rosenberg (2005) argue, when the experience occurs in sensory modalities or at processing stages that keep giving the same results and when further inter-modal testing cannot be performed (or, if performed cannot outweigh the results of the dominant modality), it will be taken as veridical. If the experience is generated in a modular way and the module is damaged, this first checking procedure is useless or rather, instead of helping falsify the experience-based belief, it will bring only further confirmation of it.

Background knowledge

Here we should distinguish two kinds of background knowledge. First, there is biographical knowledge concerning the relative supposedly replaced by an impostor. If the patient's wife, say, the date they were married, the place where they spent their honeymoon and various other episodes of their common life. Second, there also is general knowledge about the world, such as the fact that impostor stories are implausible in the first place. If we consider the first type of background knowledge, a confrontation with the purported impostor may not yield incontrovertible evidence that your belief is wrong. The purported impostor knows when you were married, knows that you spent your honeymoon in Hawaii, that the two of you have regular fights over the education of the children, and so on. But is it proof that she is who she says she is or is it rather proof that she is a clever impostor? An impostor is not just someone who happens to look like your wife, a 'sosie', but someone who pretends to be your wife and wants to make you believe she is. In addition, if this discussion takes place face to face, the Capgras' patient will experience a disturbing feeling of unfamiliarity, together perhaps with the impression that they are ill-intentioned while talking to the person and that this may suffice to bias his evaluation of the biographical evidence that is being laid out for him.

If we consider general background knowledge, Capgras' patients may be able to appreciate the implausibility of impostor stories, but this consideration alone may not carry enough weight. Implausible is not synonymous with impossible. Consider the following well-known exchange:[11]

E: Isn’t that [two families] unusual?

S: It was unbelievable.

E: How do you account for it?

S: I don’t know. I have tried to understand it myself and it was virtually impossible.

S: What if I told you I don't believe it?

E: That's perfectly understandable. In fact, when I tell the story, I feel that I'm concocting a story … it's not quite right, something is wrong.

E: If someone told you the story what would you think?

S: I would find it extremely hard to believe. I should be defending myself.

(Alexander, Stuss, & Benson, 1979: 335)

Thus, checking procedures that appeal to background knowledge would not yield unequivocal results. Use of biographical knowledge could be taken as confirmation that the person who looks like your wife is trying to pass for her, hence is an impostor. Use of general knowledge could be taken as a confirmation that the situation the subject confronts is indeed weird and in need of explanation rather than as an indication that the situation is not what the subject thinks it is.

The testimony of others

The testimony of others is part of the social division of epistemic labor. In the same way that, for language, we rely on 'expert' speakers to know the exact meaning of certain words, for beliefs we rely on experts to tell us whether we should accept a belief or reject it. But, of course, who counts as an expert depends on what the belief is about. The experts you would rely on to check your beliefs about mathematics need not be those you would consult about gardening or politics. Perfect strangers are not qualified to tell you who your wife is. Presumably, you're one of the top experts in this field. Of course, other relatives and friends may qualify as experts too, so why don't we listen to them? One problem though is that typically Capgras' delusion is initially about one close relative but, as time goes, tends to spread to other relatives. You start by thinking that your wife has been replaced by an

impostor and you end up thinking that your whole family has been replaced. Thus, it may well be that those who would be the natural experts to turn to are actually people about whom the subject already harbours nagging suspicions. Even if they have no doubt about the identity of their other relatives, Capgras' patients may, as Hirstein suggest, be impaired at reading their expressions of emotion and misinterpret their expressions of concern, sadness, etc. for negative intentions (e.g., they are out to get me, it's a plot, they are trying to drive me crazy). On either scenario, the testimony of these potential experts would be discredited, and the subject would have to rely on his sole expertise, an expertise that tells him the person in front of him is not his wife.

In a nutshell then, the main lines of the story I told here are as follows. The failure of a modular affective process involved in the recognition of emotional expression, the identification of dynamic signatures and the generation of autonomic responses and feelings of familiarity to known faces accounts for the delusional experiences of Capgras' patients. Their delusional beliefs inherit their content from their delusional experience. The particular nature of the beliefs determines what the appropriate checking procedures are. The reason why the Capgras' patients fail to dismiss their delusional beliefs is not that they fail to use these checking procedures. Rather, it happens that these procedures fail to yield disconfirming evidence. For them to give solid grounds to reject the belief, the damaged module would have to be intact. The Capgras' patient is not epistemically incompetent; he is, in a way, the victim of a vicious epistemic circle. Fortunately though, this vicious circle is limited to beliefs with a specific type of content and etiology, hence the circumscribed nature of the delusion.

Notes

[1] This paper stems in large part from work done in collaboration with Tim Bayne over the last three years. An early version of this material was presented jointly by the two of us at the Conference on "Delusion, self-deception, and affective influences on belief-formation", organized by the Macquarie Centre for Cognitive Science and the Department of Philosophy of Macquarie University in Sydney in November 2004. I also presented it at the Conference on "The modularity of emotions" organized by Université de Montréal and Université du Québec à Montreal in May 2005. I am grateful to the participants at both conferences for their comments and suggestions. Special thanks to Renée Bilodeau, my commentator at the Montreal conference, and to Tim Bayne for many insightful comments and discussions.

[2] In the present paper, I'll use emotions and feelings interchangeably.

[3] The material in this section draws heavily on Pacherie, Green & Bayne (2006).

[4] 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 (Fleminger, 1992).

[5] 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 & Rosenberg (2005).

[6] Davies & Coltheart (2000) also make this point. Note though that prosopagnosia is not quite the mirror image of Capgras' syndrome, since although prosopagnosics retain an autonomic response to familiar faces, they have lost the conscious (overt) feeling of familiarity towards them.

[7] Instead of the typical Capgras delusion — the false belief that someone has been replaced by an almost identical impostor whose actual identity is unknown to the patient — patient MF misidentified his wife as former business partner. This is the main reason why Breen *et al.* (2002) report his delusion as resembling a Capgras delusion rather than as a Capgras delusion in the strict sense. The particulars of the case are important. MF's former business partner, JY, bore a certain physical resemblance to his wife and the two women had similar names. But whereas MF had a very close positive emotional attachment to his wife, he intensely disliked JY on a personal level. Breen *et al.* speculate that these factors together with MF's difficulty in discriminating some facial expressions and in recognizing face identity when the face showed an expression were likely contributing to his misidentification of his wife as his former business partner JY. In Hirstein's terms, MF's face processing impairments would have led him to form an incorrect 'internal representation' of his wife when seeing her, a representation that happened to match his stored internal representation of his former business. The fact that the 'internal representation' yielded by faulty face processing would match a stored internal representation of JY together with the fact that the two women had

similar names and physical appearances would then account for the unusual features of the case.

[8] Unfortunately, most studies investigating autonomic responses and feelings of familiarity to faces lump together in the category of familiar faces both faces of celebrities and faces of people personally known to the subjects, such as relatives and friends. Yet it may well be that automatic responses and feelings of familiarity are not generated in exactly the same way for these two types of faces. In particular, identification of the dynamic signature of a face might play a more important role for people we interact with on a regular basis.

[9] Thanks to Renée Bilodeau for pointing out this objection.

[10] Note though that having shallow output is probably one of the less central features of modularity and that Fodor himself seems to have a rather generous notion of shallow output in mind. For instance, in his discussion of this feature Fodor (1983) considers that the output of the peripheral visual system does not just encode information about color and shape but provides basic-level categorizations, *à la* Rosch, of the objects seen.

[11] The patient who takes part in the exchange is presented by Alexander et al. (1979) as suffering from Capgras' syndrome. However, the case presents some unusual features. In particular, the patient claimed that he had two families of identical composition and described positive feelings toward "both wives". This suggests that his delusion may perhaps be better classified as a form of reduplicative paramnesia. How best to taxonomize the various

misidentification syndromes is, however, a vexed issue on which there is at present no consensus.

References

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Science*, 3, 12, 469-479.

Alexander, M.P., Stuss, D.T., & Benson, D.F. (1979). Capgras' syndrome: A reduplicative phenomenon. *Neurology*, 29, 334-339.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, Text Revision*. Fourth Edition. Washington, D.C.: American Psychiatric Association.

Bauer, R.M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the Guilty Knowledge Test. *Neuropsychologia*, 22, 457-469.

Bauer, R.M. (1986). The cognitive psychophysiology of prosopagnosia. In H. Ellis, M. Jeeves, F. Newcombe & A.W. Young (Eds.) *Aspects of face processing*. Dortrecht: Nijhoff.

Bayne, T., & Pacherie, E. (2004a). Bottom-up or top-down?: Campbell's rationalist account of monothematic delusions, *Philosophy, Psychiatry, & Psychology*, 11/1, 1-11.

Bayne, T., & Pacherie, E. (2004b). Experience, belief and the interpretive fold, *Philosophy, Psychiatry, & Psychology*, 11/1, 81-86.

Bermúdez, J.L. (2001). Normativity and rationality in delusional psychiatric disorders. *Mind & Language*, 16, 5, 457-493.

Berson, R.J. (1983). Capgras' syndrome. *American Journal of Psychiatry*, 140, 8, 969-978.

Breen, N., Caine, D., & Coltheart, M. (2000). Models of face recognition and delusional misidentification: A critical review. *Cognitive Neuropsychology,* 17, 1/2/3, 55-71.

Breen, N., Caine, D., & Coltheart, M. (2002). The role of affect and reasoning in a patient with a delusion of misidentification. *Cognitive Neuropsychiatry*, 7, 2, 113-137.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305-327.

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, 684-87.

Carmichael, S.T., & Price, J.L. (1995). Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology*, 363, 642-664.

Chao, L.L., Martin, A., & Haxby, J.V. (1999). Are face-responsive regions selective only for faces? *Neuroreport*, 10, 2945-2950.

Davies, M., & Coltheart, M. (2000). Introduction. In M. Coltheart & M. Davies (Eds.) *Pathologies of Belief* (pp. 1-46). Oxford: Blackwell Publishers.

Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic Delusions: Towards a Two-Factor Account. *Philosophy, Psychiatry, and Psychology*, 8/ 2,3, 133-58.

Davis, P.J., & Gibson, M.G. (2000). Recognition of posed and genuine facial expressions of emotion in paranoid and nonparanoid schizophrenia. *Journal of Abnormal Psychology*, 109, 445-50.

Edelstyn, N.M., & Oyebode, F. (1999). A review of the phenomenology and cognitive neurospychological origins of the Capgras' syndrome. *International Journal of Geriatric Psychiatry*, 14, 48-59.

Ellis, H.D., & Lewis, M.B. (2001). Capgras' delusion: a window on face recognition. *Trends in Cognitive Science*, 5, 4, 149-156.

Ellis, H.D., & Young, A.W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157, 239-248.

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

Ellis, H.D., Quayle, A.H., & Young, A.W. (1999). The emotional impact of faces (but not names): face specific changes in skin conductance responses to familiar and unfamiliar people. *Current Psychology*, 18, 88-97.

Fleminger, S. (1992). Seeing is believing: The role of preconscious perceptual processing in delusional misidentification. *British Journal of Psychiatry*, 160, 293-303.

Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge MA: MIT Press.

Frazer, S.J., & Roberts, J.M. 1994. Three cases of Capgras' syndrome. *British Journal of Psychiatry*, 164, 557-559.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2, 6, 568-53.

Haxby, J.V., Hoffman, E.A., & Gobbini M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, 4, 6, 223-233.

Hirstein, W., & Ramachandran, V.S. (1997). Capgras' syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society of London B.*, 264, 437-444.

Hirstein, W. (2005). *Brain Fiction*. Cambridge, MA: MIT Press.

Hoffman, E.A., & Haxby, J. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3, 1, 80-84.

Hohwy, J., & Rosenberg, R. (2005). Unusual experiences, reality testing, and delusions of alien control. *Mind and Language*, 20/2, 141-62.

Kaney, S., & Bentall, R.P. (1989). Persecutory delusions and attributional style. *British Journal of Medical Psychology* 62, 191—198.

Langdon, R., & Coltheart, M. (2000). The cognitive neuropsychololgy of delusions. In M. Coltheart & M. Davies (Eds.) *Pathologies of Belief* (pp. 183-216). Oxford: Blackwell Publishers.

LaRusso, L. (1978). Sensitivity of paranoid patients to nonverbal cues. *Journal of Abnormal Psychology*, 87, 463-71.

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.

Merrin, E.L., & Silberfarb, P.M. 1976. The Capgras' phenomenon. *Archives of General Psychiatry* 33, 965-968.

O'Toole, A., Roark, D.A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences*, 6, 6, 261-266.

Pacherie, E., Green, M., & Bayne, T. (2006). Phenomenology and delusions: Who put the 'alien' in alien control? *Consciousness and Cognition*, in press.

Reid, I., Young, A.W., & Hellawell, D. J. (1993). Voice recognition impairment in a blind Capgras patient. *Behavioral Neurology*, 6, 225-228.

Rojo, V.I., Caballero, L., Iruela, L.M., & Baca, E. (1991). Capgras' syndrome in a blind patient. *American Journal of Psychiatry*, 148, 1272.

Signer, S.F. (1994). Localization and lateralization in the delusion of substitution. *Psychopathology*, 27, 168-176.

Signer, S.F., Van Ness, P.C., & Davis, R.J. (1990). Capgras' syndrome associated with sensory loss. *Western Journal of Medicine*, 152, 719-20.

Stone, T. & Young, A. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language*, 12, 327-64.

Tranel, D. Fowles, D.C., & Damasio, A.R. (1985). Electrodermal discrimination of familiar and unfamiliar faces: A methodology. *Psychophysiology*, 22, 4, 403-408.

Tranel, D., Damasio, H. & Damasio, A.R. (1995). Double dissociation between overt and covert face recognition. *Journal of Cognitive Neuroscience*, 7, 4, 425-432.

Young, A. (1998). *Face and Mind*. Oxford: Oxford University Press.

Young, A.W., Reid, I., Wright, S., & Hellawell, D.J. (1993). Face processing impairments in the Capgras' Delusion. *British Journal of Psychiatry* 162, 695-98.