

A ONE-STAGE EXPLANATION OF THE COTARD DELUSION

PHILIP GERRANS



ABSTRACT: Cognitive neuropsychiatry (CN) is the explanation of psychiatric disorder by the methods of cognitive neuropsychology. Within CN there are, broadly speaking, two approaches to delusion. The first uses a one-stage model, in which delusions are explained as rationalizations of anomalous experiences via reasoning strategies that are not, in themselves, abnormal. Two-stage models invoke additional hypotheses about abnormalities of reasoning. In this paper, I examine what appears to be a very strong argument, developed within CN, in favor of a two-stage explanation of the difference in content between the Capgras and Cotard delusions. That explanation treats them as alternative rationalizations of essentially the same phenomenology. I show, however, that once we distinguish the phenomenology (and the neuroetiology), a one-stage model is adequate. In the final section I make some more general remarks on the one- and two-stage models.

Keywords Cotard delusion, Capgras delusion, irrationality, cognitive neuropsychology, cognitive neuropsychiatry, psychopathology, face processing.

IT IS COMMON GROUND among theories of delusion developed within cognitive neuropsychiatry (CN) that delusions are rationalizations of anomalous experiences. (Maher [1999] gives a comprehensive list of the type and origins of disturbing experience implicated in delusion, including affective disturbances, feelings of derealization or depersonalization, auditory or other

hallucinations, disturbances of volitional control, or disturbed somatic phenomenology). Because CN operates within the theoretical framework of cognitive neuropsychology, both the anomalous experience (stage one of delusion formation) and its rationalization (stage 2 of delusion formation) are understood as cognitive functions implemented in the neurophysiology of the brain. Thus, the task of CN is to construct a cognitive theory of normal function at each stage and then show how malfunctions produce characteristic psychopathology.

For example, neuropsychological investigation of schizophrenic subjects with delusions of alien control (Georgieff and Jeannerod 1988; Dapriati et al. 1996) reveals sensorimotor deficits that disrupt the subject's sense of agency, making it difficult for them to judge whether their bodily movements are produced by their own will. Delusions of alien control arise when a subject who does not experience volitional control of her own bodily movements adopts the belief that her actions are the result of another's intentions to explain the anomalous experience resulting from the sensorimotor deficit. The sensorimotor deficit is the first stage of delusion formation and the delusional explanation of the resultant disturbance in phenomenology is the second stage.

Typically, CN explanations focus on the first stage because neuropsychological techniques are

best adapted to the explanation of perceptual or quasiperceptual deficits (Fodor 1987; Shallice 1988). Some, such as Brendan Maher (Maher 1999; Oltmanns and Maher 1988), have claimed that the explanation of delusion is exhausted by the explanation of the first stage. That is, that there is no need to appeal to abnormalities in reasoning to explain delusional belief. Such theories are fortified by evidence:

1. that the formal reasoning of deluded subjects is not worse than that of normal subjects when measured on standard tests of inferential competence (Huq et al. 1988)
2. that reasoning styles (introjective versus extrojective), biases (jumping to conclusions or reluctance to abandon them), and strategies (e.g., protecting a self-image or minimizing distress) that depart from ideal rationality are found among delusional and nondelusional subjects (Bentall et al. 1991; Garety 1991; Garety and Freeman 1999; Garety and Hemsley 1994; Kaney and Bentall 1989; Kemp et al. 1997; Kinderman and Bentall 1997). So the fact that, for example, paranoid delusional subjects exhibit a tendency to jump to a conclusion on minimal evidence and to explain their experience in terms of a hostile external world is not necessarily evidence of pathological irrationality. Rather, perhaps, their tendency to do so, which falls within the normal range, issues in a delusional belief when they have the bizarre and disturbing phenomenology characteristic of delusional disorders.

Thus, one-stage accounts should not be thought of as claiming that a delusional subject is rational, where rationality is understood as revising ones belief according to idealized norms of deductive or probabilistic reasoning (Cherniak 1986; Stanovich 1999; Stein 1996). Clearly, it is irrational, measured against canons of inferential consistency, to believe a proposition for which you have conclusive falsifying evidence (for example, to believe that you are dead, as the Cotard subject claims). Rather, the one-stage theorist should be understood as claiming that the actual psychology of belief formation, which departs considerably from ideal rationality, functions in the same way in normal and delusional subjects. (For a fuller account of the issues see Garety and Freeman 1999; Gerrans, forthcoming)

In this paper, I defend a one-stage explanation of the Cotard delusion against a two-stage ac-

count developed by Young and Leafhead (1996) of both the Capgras and Cotard delusions. There are two ways to defend the one-stage account: The first is to show that the delusion can be explained in terms of anomalous experience alone (i.e. at the first stage). The second, slightly more complicated, is to show that the two-stage account, which invokes reasoning abnormalities, actually describes a rationalization process which is within the normal range. To take the second way is to engage with the general issue of whether one- or two-stage accounts are the best theory of delusion, which also requires one to take a stance on the general issue of the relationship between delusional and normal reasoning and the nature of rationality itself.

In this paper, I concentrate on understanding the origin of the anomalous experience that gives rise to the Cotard delusion. This is because, if I am correct about this matter, the Young/Leafhead explanation requires revision at the first stage, irrespective of their claims about the second stage of delusion formation. I will, however, make some remarks about the second stage and general theories of delusion in the concluding section.

The reason for agnosticism on the second stage of delusion formation is because of the structure of Young and Leafhead's argument in their 1996 paper. They claim there that the Capgras and Cotard delusions are different explanations of *essentially the same anomalous experience that arises following localized brain malfunction* (the *Capgras delusion* is the delusion that familiars have been replicated; the *Cotard delusion* that the subject has died or has no bodily existence). Thus they propose that two persons (or one person at different times) with the same localized neural deficit adopt different beliefs to explain the resultant anomalous experience. The Capgras patient adopts the belief that aspects of his environment (typically familiar people) have been replicated, and the Cotard subject explains the experience by adopting the belief that she has died. If this really is the case, then the argument for a two-stage account seems irresistible. The essential difference between the two delusional subjects is not their experience, but in the way they rationalize it.

Against this view I argue that the delusions can be distinguished at the first stage. The phenomenology of the Cotard delusion is, at core, disembodiment, whereas the phenomenology of the Capgras delusion is one of estrangement or derealization. Furthermore, this distinction in phenomenology reflects a difference in neuroetiology. To see why, we need to look more closely at the CN explanation of the Capgras delusion.

A NEUROPSYCHOLOGICAL THEORY OF THE CAPGRAS DELUSION

A classic instance of CN explanation (described by Ellis as “an exemplary vindication” [1998]) is the explanation of the Capgras delusion that explains it as the rationalization of a deficit in face processing. It builds on an elegant neuropsychological theory developed over the last fifteen years whose key insight is that face recognition has a covert affective (emotional) component and an cognitive component, which enables overt recognition (Bauer 1984; Breen et al. in press; Ellis and Young 1992; Ellis et al. 1997; Hirstein and Ramachandran 1997; Schweinberger et al., 1995; Sergent and Villemure 1989; Wacholtz 1996; Young et al. 1992).

A survey and critical discussion of this literature is contained in Farah et al. (1993) and Breen et al. (in press). It is part of the theory that the two components can dissociate because they have independent neuroanatomic substrates that can be selectively lesioned. Initially (Bauer 1984), these separate neuroanatomic substrates were hypothesized to lie in separate dorsal and ventral processing pathways, but the most persuasive hypothesis is that of Breen et al.: that face recognition is performed in the ventral visual processing pathway, whereas covert recognition effects result from the activation of pathways *connecting* the ventral pathway and the limbic system. It is damage to these latter pathways that gives rise to the Capgras delusion.

In our proposed cognitive model there is a single [face recognition unit]FRU module, from which there are independent pathways leading to the [personal identity nodes] PINs and to the affective response to previously known faces. Differential damage along these two pathways explains the dissociations be-

tween overt and covert face recognition in prosopagnosia and the Capgras delusion. (Breen et al. in press)

The Capgras delusion arises in cases where, following a neurologic accident, the pathways connecting the FRU to the limbic system are damaged leaving ventral pathways and hence overt recognition unimpaired. Thus the subject sees someone who appears, in all respects, identical to the familiar person but the subject does not experience the normal affective response. Consequently, she experiences an uncanny sensation based on the absence of an affect whose normal production is automatic and instantaneous. The Capgras delusion is an attempt to explain that uncanny feeling. The subject adopts the hypothesis that the familiar has been replaced by an identical replica, an inference that would explain why everything looks the same but feels strange.

They [delusional subjects] receive a veridical image of the person they are looking at, which stimulates all the appropriate overt semantic data held about that person but they lack another, possibly confirming, set of information which, as Lewis (1987) and Bauer (1986) have suggested, may carry some sort of affective tone. When patients find themselves in such a conflict (that is receiving some sort of information which indicates that the face in front of them belongs to X, but not receiving confirmation of this) they may adopt some sort of rationalisation strategy in which the individual before them is deemed to be an impostor, a dummy a robot, or whatever extant technology may suggest. (Ellis and Young 1992, 244)

The one-stage model here seems adequate to explain how it is that this belief might arise. The hypothesis of replication explains why a person could appear identical but the experience of seeing her produces feelings of estrangement, due to the lack of activation in pathways leading to the affective system. In this respect, the explanation of the delusion offered by CN is quite similar to classic Freudian accounts, which also focused on the disturbing feeling produced by the absence of an affect of familiarity (Capgras and Reboul-Lachaux 1924). The difference between Freudian and CN accounts is the postulated cause of the affective deficit. The CN account considered here is developed to explain instances of the Capgras delusion that arise following localized brain damage, whereas Freudian accounts provide psychogenic explanations of the loss of relevant affect.

THE COTARD DELUSION

However, Young and Leafhead's (1996) explanation of the Cotard delusion, in which the patient claims to be dead or have no bodily existence, complicates matters considerably, *because the organic basis of the two delusions is hypothesized to be the same*:

[A]lthough the Capgras and Cotard delusions are phenomenally distinct we thus think that they represent the patients attempt to make sense of fundamentally similar experiences. (p. 168)

Thus, something like a two-stage model seems required to account for the difference in content of the delusions. The anomalous experience does not exhaust the explanation of delusion because, *ex hypothesi*, the anomalous experience is the same in the two cases. Thus, the difference in content must be accounted for by *differences in reasoning* between the two subjects.

It turns out that a crucial factor differentiating Capgras and Cotard patients is that the Cotard patient is typically extremely (often psychotically) depressed. Noting this, Young and Leafhead pursue a theory (Beck 1989; Candido and Romney 1990) about introjective attributional styles in depressive patients to the following conclusion:

[W]hile depressed patients tend to attribute negative events to internal causes, people with persecutory delusions tend to attribute them to external causes . . . the persecutory delusions and suspiciousness often noted in Capgras cases may therefore contribute to the patents mistaking a change in themselves for a change in others ("they must be imposters"), whereas people who are depressed exaggerate the effect of a similar change whilst correctly attributing it to themselves ("I must be dead"). (Young and Leafhead 1996, 167).

Clearly, this account fits best with the two-stage model. The Capgras and Cotard patients form different delusional beliefs, not because their delusions have different causes but because their reasoning processes are different. In other words, the delusions are the result of divergence at the second stage.

DEPRESSION AND THE COTARD DELUSION

However, when we look more closely at the role of depression in the Cotard delusion, the

idea that the anomalous experiences of the Capgras and Cotard patient arise from "essentially the same deficit" looks less plausible. Cotard (1882) initially thought of the *delire de negation* as a subtype of melancholia, or severe depression that, in extreme forms, may produce a feeling of disembodiment. The *delire de negation* implies lack of material existence, which is the rationalization of this phenomenal experience (Berrios and Luque 1995). I suggest that the Cotard delusion is best accounted for as a rationalization of that distinctive phenomenology. Furthermore, the distinctive phenomenologies of the Capgras and Cotard have distinct neuroetiologies.

Put most simply, in the Capgras delusion the FRU alone is disconnected from the limbic areas, whereas in the Cotard delusion *all* perceptual processing is disconnected.

In Cotard all the sensory areas are disconnected from the limbic system . . . I would predict that Cotard patients will have a complete loss of GSR for all external stimuli—not just faces— . . . (Ramachandran and Blakeslee 1998, 167)

If Ramachandran is correct, then the Cotard subject has a lack of affective response to all perceptual inputs, not just the perception of familiar faces. Consequently, nothing that occurs to her evokes the normal emotional response. The Cotard patient experiences her perceptions and cognitions, not as changes in *herself*, but changes in the states of the universe, one component of which is her body, which now feels like an inanimate physical substance, first decomposing and finally disappearing (Gerrans 2000). The lack of affective experience effectively renders the physical world, including their own body, inert, at the same time as it produces a feeling of insubstantiality. Enoch and Trethowan (1991) report cases in which subjects describe themselves as mere points in space observing, but uninvolved in, the events which take place around them and another patient has spoken of herself as a "dead star" overseeing an inert cosmos. "Negation is total. Nothing exists any longer, not even themselves" (Cotard 1882, 354).

The best, integrated, explanation of the connection between depression, the Cotard delu-

sion, and the neurophysiology of affect is this: Affective mechanisms evolved to put the body into an appropriate state for action. Whether subtle and pervasive as in moods (“feeling down”) or florid and instantaneous as in fright or rage, their basis is the neurophysiologic mechanisms that regulate body state (Le Doux 1999). The phenomenology of affect is thus a way of consciously experiencing the operation of mechanisms whose functioning is automatic and inaccessible to consciousness. The reason for this inaccessibility is the role that phenomenology plays in enabling action. We evolved to experience ourselves as unified agents confronting a physically and temporally unified mind-independent world. We are never conscious of the mechanisms on which this experience depends. We see objects and colors, not the construction of transient models of our environment via the computation of reflectance properties and wavelengths; we experience sadness, fear, or elation as states of a continuing self, not the precariously integrated endocrine and dopaminergic regulation of a physical substance. There is never a straightforward connection between phenomenology, including the phenomenology of self-awareness, and neurophysiology. Yet, when something goes wrong at the level of neurophysiology, the phenomenological counterpart is characteristic, even if it is not such as to make the cause transparent (Metzinger 2000).

According to Young and Leafhead (1996), “we think that they [the Cotard and the Capgras delusion] make sense of fundamentally similar experiences” (p. 168). I have argued that the experiences are fundamentally dissimilar because the lack of affective response is different in each case. In the Cotard case it is global; in the Capgras case it is restricted to the FRU.

Of course, the neuroetiology might well overlap, with the disconnection in the Capgras case being a subset of the global disconnection in the Cotard case. This might explain the phenomenon discussed by Young and Leafhead of alternating delusions in the same subject (Cotard 1882; Enoch and Trethowan 1991; Young and Leafhead 1996). They take this phenomenon as further evidence for a two-stage explanation: that is that the same deficit rationalized in different

ways according to whether the patient’s cognitive style is depressed or paranoid, depending on fluctuations in mood.

However, we need to distinguish between cases where the Capgras delusion results from neuroanatomic damage restricted to FRU and cases where affective disconnection results from the operation of neurochemical mechanisms whose operation is diffuse, although often transient and susceptible to neuroleptic medication. These latter produce the global deficits in affect that cause the Cotard delusion. The former are more likely to occur in cases of localized damage.

Some of the alternating cases (e.g., JK and KH) described by Young and Leafhead seem to overlap, which suggests that both etiologies are involved. They point out (1996, p. 167) “that for both JK and KH, delusional misidentifications occurred after they were no longer convinced that they were dead.” This suggests that after the global affective suppression, whose basis was neurochemical, resolved, the more durable local deficit remained. Although the deficit was more global, the Cotard delusion, characteristically associated with depression (and its introjective mindset), was manifest. When the depression resolved, the more durable (because it is neuroanatomic) misconnection implicated in the Capgras delusion produced its characteristic, restricted, affective deficit. The important point to note is that the differences in neuroetiology have corresponding phenomenologies that are rationalized in different ways.

CONCLUSION

Within the framework of the two-stage conception of delusion, Young and Leafhead can be seen as arguing that the Capgras and Cotard delusion share a common first stage and thus must be distinguished in terms of differences at the second stage. I have distinguished them at the first stage and argued that this distinction is sufficient to account for the difference in content between the delusions.

However, one of the appeals of the two-stage account is that it allows us to explain, not merely why a belief might seem to plausibly account for experience, but why it is maintained

in the face of accumulating counterevidence and incongruity with the rest of the agent's knowledge. The one-stage account seems stuck with the counterintuitive idea that the delusional patients differ from us only in their experience of the world, not in the way they explain that experience.

However, as we saw, showing that a delusion depends on a systematic bias in reasoning that falls short of ideal standards of rationality is not sufficient to substantiate the two-stage account. The same type of reasoning biases are present in the general population. One might claim that delusions occur when those biases are exaggerated or introduced by intractable anomalous experiences, but that claim is of course congenial to one-stage accounts.

So, for example, Young and collaborators might be right that delusions occur in the context of paranoid or introjective mindsets, but that alone would not suffice to determine the issue between one- and two-stage theorists (Stone and Young 1987). A one-stage theorist might claim that the delusion results from an anomalous experience rationalized by a mind whose divergence from ideal rationality is within the normal range of human psychology.

Regarding the more general issue of whether a one- or two-stage account of delusion should be adopted, we can note that, within the framework of CN, the argument for two-stage accounts would be unequivocally supported by cases where two subjects have identical phenomenology and neuroetiology, but only one forms a delusional belief. Few, if any such cases have been reported. (This is not surprising however: Subjects who do not manifest abnormalities do not attract clinical attention!). In such cases, we might reasonably conclude that the delusional subject was reasoning differently from the normal and, hence, that a two-stage account was mandatory.

REFERENCES

- Bauer, R. M. 1984. Autonomic recognition of names and faces. A neuropsychological application of the Guilty Knowledge Test. *Neuropsychologica* 22:457–69
- Beck, A. T. 1989. *Cognitive therapy and the emotional disorders*. London: Harmondsworth.
- Bentall, R. 1994. Cognitive biases and abnormal beliefs. In *Neuropsychology of schizophrenia*, ed. A. S. David and J. Cutting. Hove, UK: Erlbaum.
- Bentall, R., Kaney, S., and Dewey, M. 1991. Persecutory delusions: An attribution theory analysis. *British Journal of Clinical Psychology* 30: 13–23.
- Berrios, G. A., and Luque, R. 1995. Cotard's syndrome. Analysis of 100 cases. *Acta Psychiatrica Scandinavia* 91:185–9.
- Breen, N., Coltheart, M., and Caine, D. In press. Models of face recognition and delusional misidentification: A critical review. *Cognitive Neuropsychology*.
- Candido, C. L., and Romney, D. M. 1990. Attributional style in paranoid vs depressed patients. *British Journal of Medical Psychology* 63:355–63.
- Capgras, J., and Raboul-Lachaux, J. 1924. L'illusion des 'sosies' et complexe d'Oedipe. *Annales Médico-Psychologique* 12:46–8.
- Cherniak, C. 1986. *Minimal rationality*. Cambridge, Mass: MIT Press.
- Cotard, J. 1882. Du delire des negations. *Archives de Neurologie* 4:152–70.
- Damasio, A. 1997. Towards a neuropathology of emotion and mood. *Nature* 386:769–70.
- Dapriati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., and Jeannerod, M. 1996. Looking for the agent: An investigation into the consciousness of action and self consciousness in schizophrenic patients. *Cognition* 65:71–86.
- Ellis, H. 1998. Cognitive neuropsychiatry and delusional misidentification syndromes: An exemplary vindication of the new discipline. *Cognitive Neuropsychiatry* 3:81–90.
- Ellis, D., and Young, A. 1992. Accounting for delusional misidentifications. *British Journal of Psychiatry* 157:239–48.
- Ellis, H., Young, A., Quayle, A., and De Pauw, K. 1997. Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society, London (B)* 264:1085–92.
- Enoch, M. D., and Trethowan, W. H. 1991. *Uncommon psychiatric symptoms*. Oxford: Butterworth-Heinemann.
- Farah, M. J., O'Reilly, R. C., and Vecera, S. P. 1993. Dissociated overt and covert recognition as an emergent property of a lesioned neural network. *Psychological Review* 100:571–88.
- Fodor, J. 1987. *The modularity of mind*. Cambridge, Mass: MIT Press.
- Garety, P. 1991. Reasoning and delusions. *British Journal of Psychiatry* 159(suppl 14):14–18.
- Garety, P., and Freeman, D. 1999. Cognitive approaches to delusions: A critical review of theories and

- evidence. *British Journal of Clinical Psychology* 38:113–54.
- Garety, P., and Hemsley, D. 1994. *Delusions. Investigations into the psychology of delusional reasoning*. Oxford: Oxford University Press.
- Georgieff, N., and Jeannerod, M. 1998. Beyond consciousness of external reality. A “who” system for consciousness of action and self consciousness. *Consciousness and Cognition* 7:465–77.
- Gerrans, P. 2000. Refining the explanation of the Cotard delusion. *Mind and Language* 15:111–22.
- Gerrans, P. Forthcoming. Delusions as performance failures. *Cognitive Neuropsychiatry*.
- Hirstein, W., and Ramachandran, V. S. 1997. Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society, London (B)* 264:437–44.
- Huq, S., Garety, P., and Hemsley, D. 1988. Probabilistic judgements in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology* 40A:801–12.
- Kaney, S., and Bentall, R. 1989. Persecutory delusions and attributional style. *British Journal of Medical Psychology* 62:191–8.
- Kemp, R., Chua, S., McKenna, P., and David, A. 1997. Reasoning and delusions. *British Journal of Psychiatry* 170:398–405.
- Kinderman, P., and Bentall, R. 1997. Causal attributions in paranoia and depression. Internal, personal and situational attributions for negative events. *Journal of Abnormal Psychology* 105:106–13.
- Le Doux, J. 1999. *The emotional brain*. London: Phoenix.
- Maher, B. 1999. Anomalous experience in everyday life: Its significance for psychopathology. *The Monist* 82:547–70.
- Metzinger, T. 2000. The subjectivity of subjective experience: A representationalist analysis of first person experience. In *Neural correlates of consciousness. Empirical and conceptual questions*, ed. T. Metzinger. Cambridge, Mass: MIT Press.
- Oltmanns, T., and Maher, B., eds. 1988. *Delusional Beliefs*. New York: Wiley. Interscience.
- Ramachandran, V. S., and Blakeslee, S. 1998. *Phantoms in the brain. Human nature and the architecture of the mind*. London: Fourth Estate.
- Schweinberger, S., Klos, T., and Sommer, W. 1995. Covert face recognition in prosopagnosia: A dissociable function? *Cortex* 31:517–29.
- Sergent, J., and Villemure, J. 1989. Prosopagnosia in a right hemispherectomized patient. *Brain* 112:975–95.
- Shallice, T. 1988. *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
- Stanovich, K. 1999. *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
- Stein, E. 1996. *Without good reason: The rationality debate in philosophy and cognitive science*. Cambridge, Mass: MIT Press.
- Stone, T., and Young, A. 1997. Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language* 12:327–64.
- Wacholtz, E. 1996. Can we learn from the clinically significant face processing deficits, prosopagnosia and Capgras delusion? *Neuropsychology Review* 6:203–57.
- Young, A., and Leafhead, K. 1996. Betwixt life and death: Case studies of the Cotard delusion. In *Method in madness. Case studies in cognitive neuropsychiatry*, ed. P. Halligan and J. Marshall. Sussex, UK: Psychology Press.
- Young, A., Reid, I., Wright, S., and Halliwell, D. 1993. Face processing impairments and the Capgras delusion. *British Journal of Psychiatry* 162:695–8.
- Young, A., Robertson, I., Helliwell, D., de Pauw, K. W., and Pentland, B. 1992. Cotard delusion after brain injury. *Psychological Medicine* 22:799–804.