

RATIONALITY, MEANING, AND THE ANALYSIS OF DELUSION

JOHN CAMPBELL



KEYWORDS: Capgras, Cotard, schizophrenia, radical interpretation, recognition, the first person

ON WHAT I will call a rationalist approach to delusion, delusion is a matter of top-down disturbance in some fundamental beliefs of the subject, which may consequently affect experiences and actions. On an empiricist approach, in contrast, delusion is a rational response to highly unusual experiences that the subject has, perhaps as a result of organic damage. Ellis and Young (1990) recently provided an empiricist analysis of the Capgras and Cotard delusions. I want to begin with some remarks on just why it is important to the empiricist approach that it should acknowledge the rationality of the subject's delusional responses to unusual stimuli. We will see that when the rationale for the rationality constraint is fully set out, it is questionable whether Ellis and Young's approach actually succeeds in giving its place to the rationality constraint. In conclusion, I will look briefly at the prospects for a rationalist approach and at what other approaches might be possible.

I. RATIONALITY AND KNOWLEDGE OF REFERENCE

It is often said that rationality on the part of the subject is a precondition of the ascription of

propositional attitudes such as beliefs and desires. One simple reason for thinking that rationality is critical here is that unless you assume the other person is rational, it does not seem possible to say what the significance is of ascribing any particular propositional state to the subject. If you tell me that someone rational thinks that it is raining, then given that the person is rational and does not want to get wet, I know what kinds of behavior to expect. If, however, the person is not at all rational, then saying they have the belief has no implications at all for how they will behave.

This, though, is not the fundamental reason why rationality is important for ascriptions of belief and desire. This consideration only tells us that there must be some architecture to someone's propositional states, without telling us which architecture we have to assign. The real reason for thinking that rationality is important has to do with the relation between belief and meaning. Quine long ago ridiculed the notion of a "pre-logical people" (Quine 1960). Suppose someone claims to have discovered some savages so benighted that they have not yet managed logical inference. They understand the word "alpha," let us suppose, just as we understand the word "and," but they have not yet sorted out which inferences to make using it. From a complex sentence of the form, "A alpha B," for example,

they do not infer A, and they do not infer B. Rather, in their benighted condition, they will wait to have as premises both “A alpha B” and A before they infer B. Moreover, they do not even understand from what premises they may infer “A alpha B.” They mistakenly suppose that if they have managed to prove B from A as premise (perhaps together with some further assumptions), they can conclude from that to “A alpha B” (depending on those further assumptions alone). Quine’s point was that in this kind of situation, we cannot sustain the charge of irrationality. We have to conclude that we have mistranslated “alpha,” and that the foreigners are in fact making quite sensible use of a notational variant of “if . . . then” The meaning of “alpha” dictates what constitutes a rational way of using the term. It is not really possible for someone to have grasped the meaning of a term but to be using it in a quite irrational way. The finding of irrationality can always be traded for a finding of mistranslation. And we should always translate so as to find the subject rational in the use of a term by the lights of the subject’s own understanding of the term.

This point does not apply only to logical constants. Suppose someone says to you, “The Statue of Liberty has a rather crowded location in the middle of Trafalgar Square in London, but anyone would have to admire the lions at its base; such a statue could only be English.” And so on. You might suppose that this is someone with wildly irrational beliefs about the Statue of Liberty. But the compelling diagnosis is that she does not attach the same meaning to the phrase “the Statue of Liberty” as you do. When she says, “the Statue of Liberty,” she means, “Nelson’s Column.” Then it all falls into place (cf. Davidson 1980). The problem with ascribing irrationality is that if she really knows the meanings of the words she is using, how can she be reasoning so wrongly? Surely a grasp of meaning does provide you with a capacity for correct reasoning using the term.

Let us see how this point plays out in the case of the Capgras delusion. Someone who experiences the Capgras delusion is convinced that someone, usually a near relative or family mem-

ber—generally the person’s spouse—has been replaced by an impostor (Capgras and Lachaux 1923). Over time, the number of people who seem to have been replaced by impostors may increase. The patient will claim to be able simply to see well enough that the person is an impostor, though she will typically have some difficulty in explaining what this visible difference comes to. What exactly is the right way to formulate the basic delusion? What is the content of the basic delusional belief? I do not think that the basic delusional belief is properly characterized as, for example, “That woman is not my wife,” because this would be a suitable way to react to the news that there had been something invalid about the original ceremony of marriage. But someone who reacts to this news by saying, “This woman is not my wife,” is not giving expression to a Capgras delusion, even if he is making a mistake and the original ceremony had been perfectly in order. I think a better proposal is that the singular term to use in formulating the basic delusion here is a kind of memory demonstrative: “that [remembered] woman,” referring to the woman of whom you have a number of conscious recollections. Those recollections will be what provide your knowledge of the reference of the term and which control the use that you make of the term. On this interpretation of the situation, when the patient says, “That [currently perceived] woman is not my wife,” the underlying delusion is “That [currently perceived] woman is not that [remembered] woman,” and he also has the (no doubt, accurate) belief, “That [remembered] woman is my wife.” So why is it that the patient has the false belief, “That [perceived] woman is not that [remembered] woman”? How would you go about verifying such a judgment? You would have to check that the woman you currently perceive is indeed the one of whom you have all those memories. The canonical way to do this would be to find out whether you have shared memories of the events in which you both took part. And the canonical way to do that would be to discuss those past events. It is not that your memories have to coincide at all points or even that they have to be correct memories, but that they recognizably derive from the same

episodes. Since the patient does not use this way of checking who it is that is before him, he seems to have lost his grasp of the meaning of the word.

A similar point can be made about the Cotard delusion. In the 1880s, Cotard described a syndrome he called "le delire de negation," in which everything seems so unreal that the patient thinks that she or he has died (Cotard 1882). Sometimes the patient will say that she is disembodied. The patient may claim to be dead despite being able to walk and talk. Indeed, the patient may say that she is dead even though she realizes that no one else would accept this claim. The trouble is, how can the patients really be said to be holding on to knowledge of the meanings of their remarks when they are using the words in such a deviant way?

You might propose that what we have here is not, strictly speaking, belief at all, but "empty speech" masquerading as belief. The trouble with this diagnosis is that these are perfectly sincere assertions made by people who seem to understand what they are saying, who may indeed act on the basis of what they are saying.

The basic philosophical problem raised by delusions is, then, this: since we have to ascribe meaning in such a way as to make the subject rational, we end up having no way in which to formulate the content of the subject's delusions. The empiricist solution is to say that the patient is, after all, broadly rational and in command of the meaning of the words being used; it is just that we are dealing with a broadly rational reaction to some very unusual experiences. Despite its attractiveness, though, we will see that serious difficulties face this strategy. The rationalist takes the more difficult route of saying that there is a top-down disturbance in the subject's beliefs and thus faces the problem of explaining how the subject can retain a stable grasp of meaning through this disturbance.

2. EMPIRICISM

Ellis and Young's proposal is that despite everything, the beliefs held by Capgras patients are rational. They are rationally intelligible responses to highly unusual experiences to which the

patients are subject as the result of brain damage. The key to Ellis and Young's account is the idea that there are two visual systems involved in the recognition of faces. This idea appeared first in connection with prosopagnosia—cases of people who simply cannot recognize familiar faces. In prosopagnosia, patients cannot tell who they are looking at, but there are still some non-verbal forms of response to familiar faces that are intact. When shown a familiar face, the prosopagnosic will have no idea who it is, but some affective response will still be shown. This is usually tested by measuring skin conductance responses. The idea is that one visual system, used for overt recognition of faces, has been damaged, but another visual system, used for emotional arousal in response to seen faces, is intact. So the patient may be quite unable to say at whom she is looking but have a strong sense of familiarity or affection for the seen face. Ellis and Young's proposal is that the Capgras delusion involves a kind of mirror image of prosopagnosia. The Capgras patient has intact the visual system used for overt recognition of faces, but the pathway used for emotional arousal in response to seen faces has been damaged. So what happens is that the Capgras patient looks at a familiar face, sees quite clearly its detailed similarity to the face she knows well, but finds no affective response in her own reaction to the face. This generates a sense that something is seriously wrong, and the hypothesis that the other person is an impostor is generated to explain this anomalous situation.

Young and his colleagues (Wright et. al. 1993; Young 2000) suggest that it is exactly the same kind of damage to that visual pathway that generates the Cotard delusion—the belief that one is dead. On this analysis, what happens is that the Cotard patient experiences exactly the same lack of affective response in her perceptions of other people as does the Capgras patient. The difference is that being severely depressed, the Cotard patient attributes this anomaly to a deficiency in herself—that she is emotionally dead, or simply dead, or disembodied. The Capgras patient, on the other hand, being of a more suspicious turn of mind, and perhaps inclined to feel persecuted,

attributes the anomaly not to herself but to the presence of impostors. Young says, "The closeness of this link became really convincing for us when we studied a patient who experienced both the Cotard and Capgras delusions in sequence This curious association of two unusual delusions has been reported in other cases too, and the key factor seems to be the patient's moods—when in a suspicious mood, they think other people are impostors, when depressed, they think they are dead. The underlying basis of both delusions could therefore be quite similar; a delusional interpretation of altered perception (especially loss of affective familiarity) . . . whilst the persecutory delusions and suspiciousness that are often noted in Capgras cases contribute to the patients mistaking a change in themselves for a change in others ('they are impostors'); people who are depressed might exaggerate the negative effects of a similar change whilst correctly attributing it to themselves ('I am dead')" (Young 2000, 65).

For present purposes, I want to accept the idea that this damage to the autonomic system is implicated in the Capgras and Cotard delusions and ask how it helps with the fundamental problem. Suppose we consider the Capgras delusion. The easiest way to use the Ellis and Young analysis would be to suppose that the patient thinks of the person supposedly replaced by the impostor as "whoever it is that produces that affective response in me." Then given the damage to the autonomic system, it would be perfectly sensible for the patient to judge that the person confronting her is not producing that affective response so cannot be the person who does produce that response. But that is not, actually, how we ordinarily think of our spouses. We do, in fact, tend to think in terms of our memories of them. The puzzle is to understand how it can be a rational reaction to the lack of affective response to conclude that the person confronting you is not the person of whom you have memories. As Ellis and Young present their proposal, the idea is that the patient attempts to explain her anomalous experience by supposing that the person of whom she has memories has been replaced by an impostor. But it is actually very hard to see how this could

be a rational reaction to her experience. The experience does indeed call out for explanation, but that hypothesis could not be sustained. For that hypothesis has to be checked by asking whether the observed person is indeed the one of which you have memories; all of these further enquiries show that this is not in fact an impostor.

So far I have been taking it that the content of the Capgras delusion is given by something like, "This person is not [that remembered] person, but an impostor." However, you might dispute this analysis of the content of the basic delusion. You might think that the content of the basic delusion is not given by a memory demonstrative at all. In *The Varieties of Reference*, Evans (1982) talked about the possibility of "recognition-based" ways of thinking of particular objects or people, which would not be the same as recollection-based ways of identifying them. The idea was that your ability to recognize an object would of itself provide you with knowledge of the reference of a term referring to that thing. So the content of the basic delusion could be given by "This [currently perceived] woman is not that [recognizable thus-and-so] woman."

If we suppose that the content of the "[recognizable thus-and-so]" clause is filled out by the specific layout of the target's appearance, then we still have our puzzle, because the Capgras patient's experience represents the target as having exactly that layout, and this is apparent to the patient, who may even comment on the resemblance. So we will still have no understanding of how the patient could be ascribing that meaning to the term yet come to the conclusion that he does. If, however, the "[recognizable thus-and-so]" clause is filled out by conditions that include the patient's own affective responses, then we can immediately see why the breakdown in the affective pathway should result in the patient judging that this is an impostor.

There are, however, a number of problems with this proposal. Most immediately, it is not obvious that there is such a thing as a recognition-based way of knowing the reference of a singular term. Recognition is a way of verifying judgments about a particular individual. It does

not seem to be a way of knowing the reference of a term. It seems to be quite unlike the case in which your experience of an object provides you with knowledge of which thing is being demonstratively referred to. Moreover, if we do manage to separate off this recognition-based way of thinking from a memory-demonstrative way of thinking, we will be left without any way of understanding why the Capgras patient comes to the conclusion that the person before him is not the person of whom he has all those memories.

Finally, some remarks on how these points bear on the analysis of the Cotard delusion. We can make sense of the idea that vision could give canonical verifications of some first-person judgments, such as those involved in visual kinesthesia: "I am moving towards a building," and so on. And you can imagine that affective perception in social interaction might give knowledge of a kind of social kinesthesia, as when you verify a proposition like, "No one here thinks my jokes are funny," and so on. But it is very hard to see how vision could be giving you knowledge of your own existence. By the same token, it is hard to see how vision could be giving you knowledge of your own lack of existence. In that case, the Cotard delusion does seem to be a top-down imposition on experience.

3. USE VERSUS KNOWLEDGE OF REFERENCE

In a recent article, Stone and Young (1997) try to give a more explicit analysis of the mistake that the deluded subject is making. They point to a tension between two opposed principles governing belief revision. One is what they call a principle of conservatism. The other is a principle of observational adequacy. I remember at high school my chemistry teacher demonstrating an experiment before the class. At the end she held aloft a test tube full of a vivid blue liquid. She said, "As you can see, the solution has turned yellow," and we all dutifully wrote that down. This teacher was keeping to a principle of conservatism. She was rejecting a hypothesis that was inconsistent with what she already believed. Although this is something of a special case, this

evidently is in general a constraint on belief revision. When revising your beliefs, you should not make the most radical or far-reaching changes you can think of. You should always aim to minimize the change in your beliefs. Pressed too hard, of course, this principle would mean that you never changed your mind about anything at all. You would always simply reject the deliverances of experience and stand by your original views. So this principle of conservativeness has to be balanced by a principle demanding that your beliefs should be revised so as to be consistent with the observed data. So when you see something surprising, sometimes it is, after all, right to abandon the belief you held previously and conclude that butter really does remove stains from velvet, or whatever it might be. The principles of conservativeness and observational adequacy are obviously in tension with one another, and they have to be kept in balance. There may be no general algorithm for doing this. But what we can say about the deluded subject, Stone and Young suggest, is that she gives too much emphasis to observational adequacy at the expense of conservativeness. In concluding that you are faced by an impostor, or that you are dead, you may be reporting what experience seems to show, but the cost, in violation of your previously held beliefs, is so high that the need to set off observational adequacy against conservativeness has been forgotten.

Straight off, this way of putting things makes it sound as though the mistake of the deluded subject is a relatively technical one. The "failure of rationality," of which the subject is accused, is a matter of not quite having the balance right between two principles, which are anyway in tension and which we have no firm rules for adjudicating between. It would certainly not be enough to make the thinking of the subject "un-understandable." We can perfectly well understand someone who makes this kind of mistake, just as we can understand people in general even if they deviate somewhat from ideal rationality.

In their discussion of this proposal, Davies and Coltheart distinguish between two readings of the observational adequacy principle. On one way of reading it—the first construal—"the ob-

servational data to which belief revision should be adequate concern the external world . . . rather than my experiences. On the second construal, the data to which belief revision should be adequate are data about my experiences.” (Davies and Coltheart 2000, 18). Davies and Coltheart propose that Ellis and Young go wrong in locating the source of the delusion in the patient’s attempt to explain an anomalous experience (using the second construal). After all, they say, the patient is quite right to suppose that the experience needs explaining. The problem is to understand why the patient gives such weight to the delusional hypothesis. Davies and Coltheart suggest that the source of the delusion lies rather in the subject’s taking her experience at face value. They suggest we should be focusing on the first rather than the second construal of observational adequacy. In our present terms, their proposal is that the patient’s perceptual experience has the content, “This person is not [that remembered] person, but an impostor.” In general, it is, of course, perfectly sensible to take your perceptions at face value, otherwise perception would be of no use in finding your way around. The reason why the patient has a perception with this content may be in part the damage to the affective visual pathway described by Ellis and Young, together with reasoning biases of the patient. On the Davies and Coltheart proposal, the suggestion is, we can still see the Capgras patient as broadly rational. The patient is simply taking experience at face value, which in general is perfectly rational. What is unusual about this case is that taking experience at face value leads to radical revisions of the patient’s prior beliefs. But the patient’s willingness to tolerate those revisions may in turn be explained by reasoning biases of the subject, which do not of themselves remove the patient from the sphere of the broadly rational.

Stone and Young are certainly right that there is some principle of conservativeness in belief revision. But if we represent the problem as being simply that the deluded patient violates conservativeness, we lose sight of the point that makes it so compelling that there is a constraint of rationality on our ascriptions of beliefs and

desires. The point was that there has to be a connection between the way in which the subject forms and finds the implications of her propositional states and the subject’s knowledge of the meanings of the terms she is using. Davies and Coltheart are, I will suggest, onto something when they insist that constraint on belief formation has to do with the way things are in the external world. But they miss the critical point by not connecting the problem in the subject’s use of propositional concepts to the subject’s grasp of the meanings of those concepts.

To see the basic point here, suppose we go back to the line of argument that originally made it seem compelling that there must be some constraint of rationality on the ascription of propositional attitudes. This was Quine’s point, that we can make nothing of the idea of a pre-logical people, a people who, for example, understand conjunction perfectly well, but because of deficiencies of rationality, use it in the same way as we would use “If . . . then” Suppose we just stay with Quine’s point for a moment and try to sort through just why it is correct.

The simplest reading of Quine’s point would be that our grasp of the meaning of a propositional constant is simply defined by the rules of inference we associate with the sign. Now it certainly does seem to be possible to introduce a sign merely by laying down rules for its use in inference. A sign introduced in this way is simply a syntactic trick, a technical device to facilitate inference. But that is not how we think of the ordinary propositional constants. We think of there being a semantic foundation for the use that we make of them, a semantic foundation provided by our knowledge of the truth-tables for those signs. Of course it is controversial whether ordinary speakers, as opposed to those who have had some training in logic, can be said to have knowledge of truth-tables (Dummett 1991). But for the moment, let us assume that we are dealing with a speaker who does have knowledge of the truth-table for a particular logical constant. The point of doing so is to develop the parallel between knowledge of the truth-table for a particular constant and knowledge of the reference of a singular term.

How does your knowledge of the truth-table for a logical constant relate to the pattern of use that you make of the constant? One easy answer is that the subject can by reflective reasoning establish that the standard introduction and elimination rules are correct. You can establish that the rules will be truth-preserving; if the inputs are true, then, you can read off the truth-table, the outputs will be true. Moreover, the introduction rule demands as little as possible, consistently with its being truth-preserving, and the elimination rule allows you to extract as much as possible from the proposition, consistently with the elimination rule's being truth-preserving. The trouble is that the subject engaged in this kind of reflective derivation of the rules from the truth-table will have to be using deductive reasoning in the derivation (cf. Quine 1976). The sense in which the speaker grasps the justification of the practice need not, though, involve the use of reasoning at all. The point is rather this: which particular introduction and elimination rules the speaker employs must be systematically causally dependent on which truth-table he associates with the sign. So the use of inference rules in connection with a logical constant is systematically dependent on which truth-table you associate with the constant. Change one line of the truth-table that the subject associates with the sign, and there is a corresponding change in the inference rules the subject uses in connection with the sign. Moreover, this grasp of the truth-table acquaints you with the validation of your practice. Accordingly, the reason we have to question the myth of the pre-logical people is not that the meaning of the logical constant is fixed directly by the rules of inference, but that the use they are making of the constant makes it questionable whether they are indeed interpreting the sign by means of the truth-table for "and."

How does this point apply to singular terms? How does it apply, for example, to perceptual demonstratives, terms such as "this hand" or "that hand" referring to currently perceived objects? We can have exactly the same picture here. Knowledge of the reference of the term is provided by your experience of the object. And it is your experience of the object that causes, and

justifies, the use you make of the term—the way in which you verify propositions involving it, and the implications you draw from them. Experience of the object is what provides you with your knowledge of how to form beliefs involving the demonstrative and what the implications of those beliefs are. The interpretation of the demonstrative provides a ground for the use that one makes of it in forming beliefs about those objects.

The really key question about the deluded subject is how the use that she makes of the terms in which she frames her delusion relates to her knowledge of the meaning of the terms. I think that this key question is simply missed in the way in which Stone and Young and Davies and Coltheart set up the issues. The key question does not have to do with observational adequacy (on whatever construal) versus conservativeness. The key question is whether the deluded subject can really be said to be holding on to the ordinary meanings of the terms used.

A simple way to bring this out is to reflect on one of the classic examples of delusion: a patient who looked at a row of empty marble tables in a café and became convinced that the world was coming to an end. The problem here is not that the subject is quite reasonably taking his experience at face value and going wrong only in not balancing the reasonable inclination to take experience at face value against the radical implications of doing so for the remainder of his beliefs. The problem is to understand how any experience at all, let alone an experience of marble tables, could be relevant to the verification of the proposition, "The world is ending." What we cannot find is any relation between the patient's use of the words and his knowledge of their references. We cannot take it that any transition, from a perception with a certain content to a judgment with that very same content, is *prima facie* rational, so that the only question is whether the subject ought really to have allowed that *prima facie* rational transition to be overturned by considerations of conservativeness in belief revision. Everything depends on the particular content in question and whether the transition can be seen as justified by the subject's knowl-

edge of which thing is in question. In the present case, it is hard to avoid the sense that the patient's experience could have that content ("The world is ending") only as a result of top-down loading of the experience by the patient. The reason for that is that the meanings of the terms used, "The world is ending," are not themselves sensory. It is not as if the patient has been taken in by a simple illusion of color or motion, for example.

Similarly, even if we accept that the affective visual pathway has been damaged and that the Capgras patient has an experience with the content, "This person is not that [remembered] person, but an impostor," this is not enough to establish that there has been a rational transition from the perception to the judgment. Sameness of content is not alone enough to establish that the transition is *prima facie* rational. The whole thrust of the empiricist approach is to show that the patient is thinking rationally in moving from perception to judgment. The problem comes principally in the disordered perception. But the lack of affect in the perception of the present woman simply does not, of itself, license the move to the judgment, "That [perceived] woman is not that [remembered] woman." The mere lack of affect does not itself constitute the perception's having a particular content. And it is hard to see how the proposition, "That [perceived] woman is not that [remembered] woman," could be the content of the perception, except by the kind of top-down loading of the perception that we have in the case of the marble tables. Lack of affect in the perception does not of itself amount to the perception having the content, "That [perceived] woman is not that [remembered] woman," and this lack of affect does not of itself provide verification of the proposition. So on this analysis of the content of the delusional proposition, we have to abandon empiricism and accept that at best, there is some top-down loading of the perception by the delusional content and that this loading of the perception may indeed be instrumental in the maintenance of the delusion, but that the top-down loading cannot be regarded as itself the source of the delusion.

4. RATIONALISM

Young's analysis is what I am calling an empiricist analysis of delusion. To set out the contrast between this and a rationalist analysis, I want to bring in here the notion of a framework principle. The rationalist draws a parallel between the status that the deluded subject assigns to the propositions that express the delusion and the status that Wittgenstein said attaches to what he called "framework propositions." In *On Certainty*, Wittgenstein (1969) talked about the epistemological status of propositions like "There are a lot of objects in the world," "The world has existed for quite a long time," "There are some chairs and tables in this room," "This is one hand and this is another," and so on. Wittgenstein said that beliefs expressed by such propositions are not ordinary factual beliefs, but rather form the background needed by any inquiry into truth or falsity. As he put it, "All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments: no, it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life." (Wittgenstein 1969, 105). That is, it is only when we have the framework propositions in play, assumed to be correct, that it makes any sense to try to establish whether any proposition agrees with reality or not; we need the framework propositions in order to have any methods of testing at all. In these terms, an obvious question to raise about delusions is whether the delusional beliefs do not have, for the subject, the epistemological status of Wittgenstein's framework propositions. The kind of status that we ordinarily assign to propositions like "The world has existed for quite a long time" or "This is one hand and this is another" is assigned by the deluded subject to propositions like "I am dead" or "My neighbor has been replaced by an impostor." That is, they are treated as the background assumptions needed for there to be any testing of the correctness of propositions at all. They are not themselves, in any ordinary way,

subject to empirical scrutiny. Of course, as some of Wittgenstein's examples bring out, framework propositions can be found either at quite general levels, as with the proposition that the world has existed for a long time, or in quite specific domains of enquiry, as with the proposition that these are tables and chairs. The delusions of the Capgras or Cotard patients may be relatively circumscribed in their significance, but within the appropriate domain of enquiry, they may nonetheless be functioning as framework propositions.

If you had this kind of analysis of what is going on in the Cotard and Capgras cases, you would expect there to be differences in the affective aspects of the patient's perceptions of other people. You would think that as a result of the patient's having formed the view that she is dead or that her neighbor has been replaced by an impostor, that there would be differences in the affect with which the patient perceived other people generally—after all, if you are dead, you will not interact with other people as you would if you were alive—or differences in the affect with which you perceive your neighbor—after all, you would not perceive an impostor with the same affect as you would your neighbor. But the direction of explanation here would be the opposite of the direction of explanation proposed by Young and colleagues. According to Young and his colleagues, it is because of the difference in perceptual affect that the patient comes to think that she is dead or that her neighbor has been replaced by an impostor. This is what I am calling an empiricist analysis. On a rationalist analysis, however, the difference in framework proposition comes first, and that is what explains the difference in perceptual affect. The difference in perceptual affect is, on the rationalist analysis, a consequence of a more fundamental difference in which framework propositions are being maintained.

Young's empiricist analysis of the Capgras case depends on supposing that you recognize someone else, in part, by the affect in your perception of them. But this is a surprising view. We would not ordinarily think that seeing someone else (a) produces a particular affect in you—for exam-

ple, a sense of welcome—and that finding yourself with that sense of welcome, you conclude (b) it must be Sally! We would rather have thought that what happens is (a) you recognize Sally and, consequently, (b) you experience a rush of welcome. Of course this commonsense view is shaken by the findings on skin-conductance responses in prosopagnosics, which do seem to show that there can be some affective response in the absence of explicit recognition of the other person. But these findings cannot be pressed so hard as to imply that there is no top-down component in perceptual affect. Suppose, for example, that you and I are friends but that we have a fight. Later I see you, having forgotten the fight, and I have my usual positive affect on seeing you. Then I remember the fight, and now the perceptual affect is not so good. Since there can be top-down determinants of perceptual affect, we can hardly conclude immediately that the lack of perceptual affect in Capgras patients' responses is not the result of top-down factors, such as a prior belief that this person is an impostor. Similarly, in the Cotard case, the general lack of affect in my response to other people may be the result of my having determined that I am dead rather than being the cause of that belief.

One reason for preferring an empiricist to a rationalist analysis of delusion seems to be a sense that it is easier to explain a change in experience as the result of organic malfunction than it is to explain a change in belief as a result of organic malfunction. As long as unusual experiences can in principle be the result of organic malfunction, and changes in belief rational consequences of those unusual experiences, the empiricist program may seem in principle capable of being implemented. And so as long as it seems impossible that change in belief could be a result of organic malfunction, there will seem to be a difficulty of principle for the rationalist program. But why should we think there to be a difficulty of principle about explaining change in belief as a direct result of organic malfunction? It is sometimes pointed out that a change in one belief can, in principle, affect endlessly many further beliefs held by the subject. For example, if I find out that I was adopted, that may affect endlessly

many further beliefs that I have. How, you might ask, could there be a single organic malfunction capable of producing these endless revisions of belief? But in fact it is hard to see a difficulty of principle here. Suppose I am simply told, directly and authoritatively, that I was adopted and that the people I lived with all those years were not, as I had supposed, my biological parents. Then there will indeed be endless reverberations of this revision all through my system of beliefs. But the situation is not radically changed if we suppose that it is as the result of an organic malfunction that I come to believe that I was adopted. It may be the very same causal mechanisms that operate in both cases to bring about the endless reverberations of that single change in belief. If there can be a change in experience as the direct result of organic malfunction, it is very hard to see why there could not be change in a single belief as the direct result of organic malfunction.

I have been arguing that the difficulty for the empiricist program is a difficulty not of principle but of detail: When we work through just what is supposed to be happening, in the Cotard and Capgras delusions, on the empiricist analysis, we find it impossible to explain how the subject is supposed to be retaining a grasp of the meanings of the singular terms involved in expressing the content of the delusion, consistently with making the kind of use of those terms that the patient must be making to form and maintain that delusion. What happens, on the rationalist analysis, to the subject's grasp of meaning? Wittgenstein's notion of a framework proposition was never worked out in great detail. But it is certainly part of the picture here that a change in framework principles would bring with it a change in the meanings of the terms used. The subject who moves from one set of framework principles to another destabilizes the meanings of the terms used. The new meanings were not even commensurable with the old meanings; there is no translation from one framework into another. So on a rationalist analysis, the deluded subject saying, "That [perceived] woman is not that [remembered] woman," is not here using the memory demonstrative, "that [remembered] woman," in a way that is commensurable with the way in

which the subject used the term before the onset of the delusion. The situation is rather as Kuhn (1970) described the relations between the terms used in scientific theory before and after a revolutionary change in the key principles of the discipline. According to Kuhn, the meaning of the term "mass," for example, in classical mechanics is simply unrelated, except as historical antecedent, to the meaning that the term "mass" bears in relativistic physics. Just so, on this rationalist analysis, the meaning that the memory demonstrative bears before the onset of the delusion is unrelated, except as historical antecedent, to the meaning that the subject assigns to it after the onset of the delusion. Because the delusions of the Cotard or Capgras subject are relatively circumscribed in domain, there may be no global change in meaning to remark, but it is illuminating to bear in mind here the remark of one schizophrenic subject that when he speaks, the words bear two meanings: the meanings that they ordinarily bear and the meanings that he is trying to use them express.

5. ANOMALOUS EXPERIENCE AND DELUSION AS HAVING A COMMON CAUSE

I want finally to look briefly at an alternative account of the Cotard delusion that is given by Philip Gerrans (1999) in his paper in the *Monist*. Gerrans stresses the severe depression, the flattening of affect, which is characteristically experienced by the Cotard patient. In effect, he rejects both the empiricist and rationalist approaches. He argues that severe depression is a common cause of both the perceptual anomalies, the lack of affect in vision experienced by the patient, and the patient's belief that she is dead. In a striking passage, Gerrans writes:

Cases of Cotard's have been reported (Enoch and Trethowan 1991) in which the subject proceeds beyond reporting her rotting flesh or her death to the state of describing the world as an inert cosmos whose processes she merely registers without using the first-person pronoun. In this type of case the patient conceives of herself as nothing more than a locus, not of experience, because, due to the complete suppression of affect, her perceptions and cognitions are not an-

nexed to her body, but of the registration of the passage of events. She has effectively effaced herself from the universe: nothing which occurs is of any significance to her and, hence she describes the world without implicating herself in that description. Few such cases are reported nowadays because that level of depression is recognized as preceding suicide and attracts pharmacological intervention, or, in extreme cases, ECT, at an earlier stage. However, their occurrence does strongly suggest that, in the absence of affective processing, perception and cognition have no bodily consequences and thus are not "felt" at the phenomenal level to belong to the agent.

. . . My conclusion is that, if affect is completely flattened, the taken-for-granted background of stable selfhood will be disrupted because information-processing derived from perceptual or cognitive channels has no bodily consequences. Thus it is natural that the extreme cases of psychotic depression whose physiological result seems to be to shut down global affective processing, are experienced as disembodiment. The patient does not recognize experiences as significant for her because, due to the global suppression of affect, she has no bodily responses to the acquisition of even the most emotionally significant information (Gerrans 1999, 604).

On one view, the deviant framework proposition, "I am dead," is what causes the loss of affect in perception of other people. That is what I have been calling "rationalism." In a rival view, the loss of affect in perception of other people is what causes acceptance of the deviant framework proposition, "I am dead." That is what I have been calling "empiricism." Gerrans, in effect, rejects both views. In his analysis, it is depression, a global flattening of affect, that is responsible for both the loss of affect in perception and acceptance of the framework proposition, "I am dead."

This is an engaging analysis, and I want finally to remark that the fact that this view seems even comprehensible brings out something quite special about the first person. It is certainly not in general true that we can perceive only objects concerning which we have some affect. For example, Sirius B is markedly perceptible even if you feel no emotion in connection with it at all. And the Capgras patient who has affectless perceptions of other people does not, therefore, find that those other people are simply invisible. They may be impostors, but they do still exist.

Gerrans is bringing out something quite special about the self: that there is some sense in which it can be regarded as defined as a locus of affect. Just how and why this is so would take quite a deep enquiry to determine. But there is considerable initial appeal to his point. The proposal is that if affect is entirely flattened, this object does indeed disappear from view. We can conceive of the self only as a locus of affect. It is that point that makes possible his rejection of both rationalism and empiricism, so that depression can function as a common cause of the delusional belief and the affectless perception in the Cotard patient.

REFERENCES

- Capgras, J., and Lachaux, J. 1923. L'illusion des "Sossies" dans un Délire Systématisé Chronique. *Bulletin de Société Clinique de Médecine Mentale* 11:6–16.
- Cotard, J. 1882. Du Délire des Négations. *Archives de Neurologie* 4:152–170, 282–95.
- Davidson, D. 1980. On saying that. In *Inquiries into truth and interpretation*, 93–108. Oxford: Oxford University Press.
- Davies, M., and M. Coltheart. 2000. Introduction: Pathologies of belief. In *Pathologies of belief*, ed. M. Coltheart and M. Davies, 1–46. Oxford: Blackwells.
- Dummett, M. 1991. *The logical basis of metaphysics*. London: Duckworth.
- Ellis, A. W., and A. W. Young. 1990. Accounting for delusional misidentifications. *British Journal of Psychiatry* 157:239–48.
- Enoch, M. D., and W. H. Trethowan. 1991. *Uncommon psychiatric symptoms*, 3d ed. Oxford: Butterworth-Heinemann.
- Evans, G. 1982. *The varieties of reference*. Oxford: Oxford University Press.
- Gerrans, P. 1999. Delusional misidentification as sub-personal disintegration. *Monist* 82:590–608.
- Kuhn, T. S. 1970. *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Quine, W. V. O. 1960. *Word and object*. Cambridge: MIT Press.
- . 1976. Truth by convention. In *The ways of paradox and other essays*. Cambridge: Harvard University Press.
- Stone, T., and A. W. Young. 1997. Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language* 12:327–64.

Wittgenstein, L. 1969. *On certainty*, ed. G. E. M. Anscombe and G. H. von Wright, trans. D. Paul and G. E. M. Anscombe. Oxford: Blackwell.

Wright, S., A. W. Young, and D. J. Hellowell. 1993. Sequential Cotard and Capgras delusions. *British Journal of Clinical Psychology* 32:345-49.

Young, A. W. 2000. The neuropsychology of abnormal belief. In *Pathologies of belief*, ed. M. Coltheart and M. Davies, 47-73. Oxford: Blackwells.