

Chapter 14

Top-down and bottom up in delusion formation

(Originally published in *Philosophy, Psychiatry and Psychology* 11(1): 65-70, 2004)

Preview

Both bottom-up and top-down cognitive mechanisms may be involved in delusion formation. Many accounts of delusion formation prioritise one of these mechanisms at the expense of the other. I argue that such accounts are less promising than accounts that allow interaction between top-down and bottom-up mechanisms. I speculate that a promising unifying candidate for the understanding of delusion formation may be the fundamental cognitive processes associated with predictive coding.

Chapter 14

Top-down and bottom up in delusion formation *

Some delusions may arise as responses to unusual experiences (Maher 1974; Davies, Coltheart et al. 2001). The implication is that delusion-formation in some cases involves some kind of bottom-up mechanism—roughly, from perception to belief. Delusion-formation may also involve some kind of top-down mechanism. This could be in the shape of a patient's flawed background beliefs, or biases, which somehow modulate perception, or ensure delusional interpretations of experience. There is little agreement about what kind of mechanism is primary in delusion formation, or, indeed, what the particular mechanisms and their neural substrates might be.

In the evaluation of top-down and bottom up mechanisms in delusion formation a number of conceptual questions should be addressed (Davies and Coltheart 2000). For example, if there is an experiential component, why is a hypothesis with a delusional explanation of the unusual experience seen as relevant? Why do the subjects come to believe it tenaciously? How come some unusual experiences do not develop into delusions?

A number of empirical findings also need to be taken into account. For example, the damage to the autonomic system in Capgras delusion (Ellis, Young et al. 1997); the decreased ability for fast error correction in schizophrenia (Frith and Done 1989);

* I am grateful to Chris Frith for very valuable discussions and suggestions.

modulation of activity in parietal cortex in delusions of alien control and other passivity experiences (Blakemore, Oakley et al. 2003); in schizophrenia, the increased sensitivity to self-produced stimuli, such as tickling oneself (Blakemore, Smith et al. 2000); and the sound of one's own voice (Ford, Mathalon et al. 2001).

Theory choice can also be guided by philosophical doctrines about, for example, the nature of belief, the constitutive role of rationality, and intersubjectivity and the nature of judgements about the external world. It is natural, however, to let beliefs about philosophical doctrine be secondary to empirical findings and conceptual issues.

The distinction between bottom-up and top-down mechanisms is particularly sharp in John Campbell's account of delusions (2001) and in Tim Bayne and Elizabeth Pacherie's (this Journal) discussion of Campbell. In this discussion, top-down approaches are associated with the general philosophical doctrine of rationalism, and bottom-up approaches with the philosophical doctrine of empiricism. Campbell's account is top-down because he posits delusional framework beliefs that arise directly as a result of brain damage or malfunction, and that work in some top-down fashion on the interpretation of experience. Bayne and Pacherie convincingly defend empiricism against Campbell's rationalist account. Among their many other objections, they briefly puzzle about how a delusional framework belief (e.g., that one's spouse is an impostor) in a top-down fashion could cause the damage to the autonomic system seen in Capgras and Cotard (ms p21). I think this is a very important objection because the skin conductance data is part of the evidence for the bottom-up hypothesis concerning Capgras. We could add to this, for example, that it is mysterious how belief in a top-down fashion could cause people with schizophrenia to be worse at fast error correction than normal controls,

to be better at tickling themselves than normal controls, and to have an increased response to the sound of their own voice.

On the other hand, though the SCR evidence suggests a bottom-up approach to Capgras, it doesn't explain the character of the subjective experience that is hypothesised to give rise to the delusional belief: reduced SCR is not part of the experience itself (thanks to Ian Gold for discussion of this). Something more is required to give a full account of the type of experience that gives rise to the delusion. Further, as they stand, most bottom-up accounts are unable to account for why reality testing fails for beliefs based on these experiences.

Neither rationalism nor empiricism on their own seem able to address the conceptual questions while paying heed to the empirical findings. One moral to draw from this is that it is not fruitful to think of bottom-up and top-down mechanisms as mutually exclusive, philosophically inspired doctrines.

Bayne and Pacherie also object that the strong top-down view cannot account for the relatively monothematic nature of delusions such as Capgras. Why should a belief induced directly by brain damage give rise to beliefs with only certain sorts of content? It is instructive to see that Robert Klee (this Journal) argues, on the other hand, that an empiricist bottom-up model cannot explain the thematic content of delusions. Why should the attempt to explain an anomalous experience (e.g., of an intention not labelled 'mine') predominantly give rise to delusions that elaborate the theme of alien control (ms p.4)? This echoes some of the conceptual questions mentioned above, though here it is used to set aside bottom-up approaches.

Klee then proposes that delusions instead should be elucidated in top-down fashion in terms of whether delusional beliefs are rationally explicable or not. What he calls ‘pedestrian’ delusions (e.g., beliefs about persecution by governmental authorities) are explainable as cognitive mistakes (ms p10), whereas occurrences of ‘stark’ delusions (e.g., delusions of alien control) necessarily are inexplicable, even though the contents of the delusions themselves can be understood (ms p15-16). The discussion is driven partly by the insight that clinicians in fact do understand the contents of delusions even though they are very bizarre, and the distinction itself is motivated through a critical discussion of philosophical considerations about the individuation of belief content: the view that content of belief is determined qua rational relations to other beliefs. However, this approach seems less satisfactory.

First, what decides whether a delusion is pedestrian or stark? According to Klee, some persecutory delusions concerning governmental authorities are pedestrian, but what should we then say about Frégoli delusions that can have a persecutory element too (Courbon and Fail (1927) described a theatre enthusiast who believed that the actresses Robine and Sarah Bernhardt were disguising themselves so as to persecute her)? It seems best to group them with other monothematic, circumscribed delusions of misidentification, which presumably are stark (e.g., Capgras, Cotard). But, on Klee’s account, it seems they could just as well belong with other, pedestrian, persecutory delusions.

Second, Klee argues that stark delusions deny beliefs that stand fast for the rest of us, and that, since explanation begins from what stands fast for us, those delusions are necessarily beyond explanation. This presupposes a foundationalist conception of

explanation that may well be inadequate. What is needed in explanation is not a firm final foundation: when someone explains that they believe in UFOs because they saw a UFO, you can appreciate the explanation without understanding why they saw a UFO (Lipton 1991).

It would be easier to say something about these two issues if the bottom-up conception of delusions had not been set aside. Perhaps the difference between Frégoli and persecutory delusions is that the former has a particular experiential component associated with increased activity in the autonomic system (for discussion, see Davies, Coltheart et al. 2001). And, once we allow the possibility of delusional experiences, we can at least begin to explain delusional belief with reference to these experiences. The fact that such explanations cannot fully account for the thematic content of the delusions doesn't go far enough to entail that the delusions are necessarily inexplicable. Similarly, it may be correct that, beginning solely from the patient's prior beliefs, it is not possible to predict stark delusions (ms p17), but once we take the patient's experiences into consideration, it may indeed be possible to predict their delusions.

Again, though our understanding of delusions is enhanced by closer attention to, say, philosophical top-down notions, it is more fruitful to allow a more integrated perspective on bottom-up and top-down mechanisms.

Klee's distinction between stark and pedestrian delusions was made operable in terms of non-deluded subjects' ability to understand and explain the delusional beliefs. Eugenie Georgaca (this Journal) appeals to social constructivist notions of discursive factualization devices to elucidate the central question about our judgements of the veridicality of delusions. Georgaca illustrates how such judgements are negotiated, and

how they may depend on social and institutional factors. However, she focuses on purely formal discursive devices and the broader rationalizations of patients (here, in terms of religious beliefs), at the expense of considerations about the quality and nature of the patient's experiential evidence for the belief (e.g., that he very firmly reports seeing and hearing little devils, extract 2, lines 98-116, ms p 10). Quite apart from constructivist notions of factualization, surely the fact that the patient reports his experiences and says "there was definitely someone there" when nobody else saw anybody, is a factor in establishing the falsity of his belief. Again, a less satisfactory account is arrived at when bottom-up accounts are set aside.

The upshot is that our understanding of delusion formation is best served by considering both top-down and bottom-up mechanisms in tandem, not by prioritising one at the expense of the other. Here is a brief speculative example of how more integrated accounts may go.

A case of delusion of alien control in schizophrenia. Top-down: hypofrontality (Spence, Hirsch et al. 1998) associated with posterior hyperactivity modulates the experience of self-initiated movement so that it is experienced the way one experiences externally generated movement (Blakemore, Oakley et al. 2003). Bottom-up: there is no inhibition of the pre-potent doxastic response (i.e. our tendency to believe what we experience) (Davies, Coltheart et al. 2001), and the content of the experience is adopted as belief. Top-down performance (Gerrans 2001) failure: the beliefs based on experience in some sensory modalities or at some processing stages are inaccessible to reality-testing, they cannot then be revised, and are subsequently explained in terms of supernatural hypotheses. The supernatural theme is prioritized because what needs to be

explained is the belief that the movement is externally generated, not the belief that it is as if the movement is externally generated, in which case themes concerning the patient's possible mental illness would be more likely to present themselves (Hohwy and Rosenberg submitted).

I have mentioned several methodological considerations in the evaluation of delusion formation: conceptual issues, empirical results, and philosophical doctrines. But there are also broader cognitive considerations. Schizophrenia may have multiple aetiological factors (DNA, gene expression, viruses, birth injury, etc.), and it also has multiple combinations of symptoms (hallucinations, delusions, negative symptoms, disorganised speech and behaviour) involving a variety of cognitive functions (attention, memory, language, executive function, emotion). It is reasonable to suppose, however, that schizophrenia arises because those multiple aetiological factors disturb a fundamental cognitive mechanism, which, in turn, plays a role in the functioning of the various second order cognitive mechanisms (Andreasen 1999). Given that a variety of top-down and bottom-up mechanisms may play a role in delusion formation we can speculate that the nature of this fundamental cognitive mechanism concerns top-down and bottom-up connectivity.

The move towards more fundamental top-down/bottom-up cognitive processes is illustrated in the development of what is in my view the most convincing account of delusion formation: the forward modelling account developed by Blakemore, Wolpert and Frith (2000) to explain delusions of alien control. This account began (in Frith 1992) as a monitoring account in terms of relatively higher order cognitive functions concerning intentions. In its current version, it builds on the notion that the motor control system is

centrally a predictive system where efference copies of motor commands allow fast on-line modelling of the movements that are about to be performed and the sensory consequences of those movements. When the movement happens as predicted, posterior activity is attenuated and the predicted afferent signal is cancelled. Malfunctions in this system lead to less attenuation and the effect is that one's movements are experienced the way one experiences passive movement.

There is reason to believe that prediction and attenuation mechanisms occur throughout the brain (for theory and review, see Friston 2002). The general picture is that representation is a matter of bottom-up driving and top-down prediction, with attenuation of activity when the predictions are matched by sensory input. A core reason for thinking that representation involves this kind of predictive coding is that it helps avoid the inverse problem (of trying to use only the effect to determine the causes). The idea is that top-down predictions are based on hypotheses about what, in varying contexts, best explains a particular bottom-up stimulus. The best predictive model determines the representational content, and activity in the lower levels can be decreased. Predictions are generated on the basis of prior knowledge, but this may well be a dynamic relation in the sense that priors may be affected by feedforward error signals on an on-going basis. This highly interactive system can be contrasted with more traditional feedforward feature-detecting models of perception. Apart from the motor control system, predictive coding may be found in perceptual integration, e.g., shape perception (Murray, Kersten et al. 2002), and, perhaps, in social interaction (Wolpert, Doya et al. 2003). Interconnectivity of this sort may be relatively local (e.g., between striate and extrastriate areas), or relatively global (e.g., between frontal and posterior areas); it may also be highly hierarchical, such

that ever higher levels predict and modulate the representations of lower predictive systems.

If the top-down/bottom-up processes associated with the fundamental cognitive mechanism of predictive coding occur for all representational systems, from something as basic as shape perception, to social interaction, over motor control (because they are all concerned with representing causes of driving signals), and, further, if core symptoms of schizophrenia are best characterised (conceptually, empirically, philosophically) in terms of various types of interacting top-down/bottom-up mechanisms, then types of disturbances to predictive coding transpire as an interesting candidate in the explanation of schizophrenic symptoms.

We may, at a purely speculative level, consider how disturbances to predictive coding may explain aspects of schizophrenia.

The predictive system is highly dynamical, interconnected and hierarchical, therefore it seems plausible that relatively small differences in the disturbances to the system (perhaps at slightly different developmental stages) can give rise to relatively large differences at the level of manifested symptoms, as is seen in schizophrenia.

Predictive coding is associated with different patterns of increase and decrease in activity in lower areas. This could relate to negative and positive symptoms in schizophrenia. For example, lack of attenuation of activity in areas of the parietal cortex is associated with delusions of alien control (Spence, Brooks et al. 1997; Blakemore, Oakley et al. 2003); and it could be—though I again stress this is speculative—that the reverse, namely relative lack of activity in these areas, contributes to negative symptoms, such as avolition. This may happen if no error signal is generated even though the

observed action is not predicted (if my experiences of other people's action resemble my experience of my own, then, even though reality testing may be intact, there will be less incentive for me to respond to these actions).

In addition to processing of error signals, predictive coding involves bottom-up driving and generation of predictions. This could map onto different types of symptoms. Biased selection of predictions could lead to faulty, perhaps persecutory, interpretations of stimuli; malfunctioning dynamic up-date of priors could lead to tenacious maintaining of interpretations once arrived at; deficient and poorly integrated bottom-up driving could be associated with hallucinations (see Frith and Dolan 1997).

Predictive coding may be given a role in the account of monothematic delusions like Capgras too. If activity in the autonomic system affects the priors in the visual processing of a loved one's face, then lack of such activity may modulate experience of the loved one's face: perhaps early activity is in the normal case decreased more for loved one's faces than for other faces.

If something like this story is correct, then we get a much more nuanced picture of top-down and bottom-up mechanisms than much of the literature would suggest. It would not be true that one or the other is primary: the content of experience is modulated by prior knowledge at many different levels; and the content of prior knowledge is driven and up-dated in a highly dynamic way by incoming stimuli. Furthermore, though family structures and social and institutional factors may play a role in the development of delusions, it may be that we cannot understand the role of these factors until we grasp the fundamental cognitive systems they interact with.

References

- Andreasen, N. (1999). "A unitary model of schizophrenia." Archives of General Psychiatry **56**: 781-787.
- Blakemore, S.-J., D. A. Oakley, et al. (2003). "Delusions of alien control in the normal brain." Neuropsychologia **41**: 1058-1067.
- Blakemore, S.-J., J. Smith, et al. (2000). "The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring." Psychological Medicine **30**: 1131-1139.
- Campbell, J. (2001). "Rationality, meaning and the analysis of delusion." Philosophy, Psychiatry, and Psychology **8**: 89-100.
- Courbon, P. and G. Fail (1927). "Syndrome "d'illusion de Fregoli" et schizophrénie." Annals of Medical Psychology **85**: 289-290.
- Davies, M. and M. Coltheart (2000). "Introduction: pathologies of belief." Mind and Language **15**(1): 1-46.
- Davies, M., M. Coltheart, et al. (2001). "Monothematic delusions: Toward a two-factor account." Philosophy, Psychiatry, Psychology **8**(2-3): 133-158.
- Ellis, H. D., A. W. Young, et al. (1997). "Reduced autonomic responses to faces in Capgras delusion." Proceedings of the Royal Society: Biological Sciences B **264**: 1085-1092.
- Ford, J. M., D. H. Mathalon, et al. (2001). "Cortical responsiveness during talking and listening in schizophrenia: An event-related brain potential study." Biological Psychiatry **50**(7): 540-549.
- Friston, K. (2002). "Functional integration and inference in the brain." Progress in Neurobiology **68**: 113-143.
- Frith, C., S.-J. Blakemore, et al. (2000). "Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action." Brain Research Reviews **31**: 357-363.
- Frith, C. D. (1992). The Cognitive Neuropsychology of Schizophrenia. Hillsdale, NJ, Lawrence Erlbaum Ass.
- Frith, C. D. and R. J. Dolan (1997). "Brain mechanisms associated with top-down processes in perception." Philosophical Transactions of the Royal Society London B **352**: 1221-1230.
- Frith, C. D. and D. J. Done (1989). "Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action." Psychological Medicine **19**: 359-363.
- Gerrans, P. (2001). "Delusions as Performance Failures." Cognitive Neuropsychiatry **3**: 161-173.
- Hohwy, J. and R. Rosenberg (submitted). "Unusual experiences, reality testing, and delusions of control."
- Lipton, P. (1991). Inference to the Best Explanation. London, Routledge.
- Maher, B. A. (1974). "Delusional thinking and perceptual disorder." Journal of Individual Psychology **30**: 98-113.
- Murray, S. O., D. Kersten, et al. (2002). "Shape perception reduces activity in human primary visual cortex." Proceedings of the National Academy of Science **99**(23): 15164-15169.

- Spence, S. A., D. J. Brooks, et al. (1997). "A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control)." Brain **120**: 1997-2011.
- Spence, S. A., S. R. Hirsch, et al. (1998). "Prefrontal cortex activity in people with schizophrenia and control subjects. Evidence from positron emission tomography for remission of 'hypofrontality' with recovery from acute schizophrenia." British Journal of Psychiatry **172**: 316-323.
- Wolpert, D. M., K. Doya, et al. (2003). "A unifying computational framework for motor control and social interaction." Philosophical Transactions of the Royal Society London B **358**: 593-602.