A COMPUTATIONAL MODEL OF THE CAPGRAS DELUSION

Ryan McKay[1], Fabio Paglieri[2]

[1]Tufts University (Medford, USA)
[2]Institute for Cognitive Sciences and Technologies-CNR (Roma, Italy)

Delusions are false beliefs that are held with strong conviction despite the efforts of others to dissuade the deluded individual. They are first-rank symptoms of schizophrenia and feature prominently in numerous other psychiatric and neurological conditions. Thematically speaking, delusions range from the bizarre and exotic (e.g. «I am the left foot of God») to the *relatively* humdrum (e.g. «My neighbours are persecuting me»). Few delusions, however, have captured both the professional and public imaginations like the Capgras delusion, which involves the belief that a spouse or close relative has been replaced by a physically identical impostor.

Theories of the Capgras delusion tend to take the form of either *motivational* or *deficit* accounts (McKay et al., 2005a, b). Motivational formulations explain the delusion in terms of some putative psychological benefits it provides. For example, Enoch and Trethowan (1991) claim that the delusion resolves ambivalent feelings of love and hatred towards a loved one. In coming to believe that the loved one is in fact an impostor, the deluded individual can feel or express their unacceptable feelings without guilt or tension.

Deficit accounts, on the other hand, view the Capgras delusion as the consequence of a breakdown in the normal functioning of belief mechanisms, underpinned by neuroanatomical or neurophysiological abnormalities. For example, Langdon and Coltheart (2000) explain cases of Capgras delusion as resulting from the conjunction of two cognitive deficits. The first deficit involves damage to neural pathways underpinning the emotional component of face recognition (see Ellis et al., 1997). The ensuing discordance between experiences of the way someone «looks» and the way they «feel» is thought to be what provides the content of the impostor delusion. The second deficit is some kind of defect in the machinery of belief evaluation, and accounts for why the impostor hypothesis, once generated, is then *adopted* and *maintained* in the absence of appropriate evidence for that hypothesis.

A recent model of the Capgras delusion integrates both motivational and deficit perspectives (McKay et al., 2005b; see also Raschka, 1981). This model posits that the Capgras delusion may result when cognitive or perceptual deficits interact with salient motivational forces. For example, imagine a man who loves his wife deeply, and is thus *motivated* to believe that they have a happy and successful relationship. If such a man suffers a neurological disruption to pathways subserving the emotional component of face recognition, he may be confronted with a combination of doxastic inputs (1. This looks like my wife; 2. This doesn't *feel* like my wife) that, given his motivational state, is intensely disturbing. In the face of these inputs, the inference that there is something amiss in his relationship with his wife may be the logical one to make. However, if he is sufficiently motivated to avoid making this inference, he may be attracted to an alternative that is less logical yet in some sense still «fits the facts» («perhaps this is not my wife after all, but rather a physically identical impostor»).

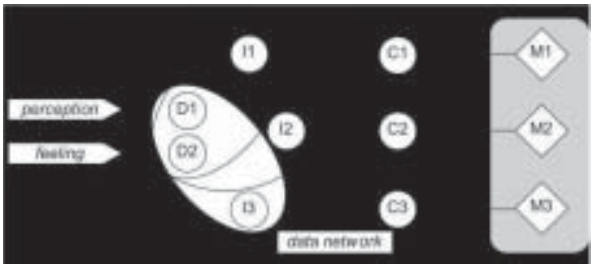
A possible means of testing this integrated explanation of Capgras delusion, and of contrasting it with competing accounts (i.e., purely motivational and deficit theories), involves building a computational model of the doxastic context in which Capgras delusions arise. Such a model should abstract the clinical features of typical Capgras cases, preserving only those factors considered relevant in delusion formation — namely, the available perceptual information, the patient's motives, and the cognitive machinery underpinning belief dynamics. By running computer simulations of this simplified model, it should be possible to assess the relative impact of each of these factors in the generation and maintenance of the Capgras delusion.

A major difficulty in such research concerns the choice of an appropriate formal architecture for representing belief dynamics. Most current approaches either ignore (Gardenfors, 1988) or marginalize (Boutilier, 1998) the influence of motives in belief formation and change, and hence are inadequate for critically assessing the role of motivational factors in the formation of the Capgras delusion (for a survey, see Paglieri, 2005). To overcome this limitation, we make use of an alternative approach to belief dynamics, *Data-oriented Belief Revision* (DBR). DBR was recently proposed by Paglieri and Castelfranchi, and has been successfully used to model motivational influences on belief formation and change (Paglieri, 2005; Paglieri, Castelfranchi, 2005). DBR is based on the ideas that (i) perceived information is believed or disbelieved on the basis of several specific properties, i.e. factual credibility, epistemic importance, motivational likeability and pragmatic relevance; moreover, (ii) belief change should be understood as an emergent effect of information dynamics, rather than being represented as an independent phenomenon as is customary in standard approaches.

In framing Capgras delusion within the DBR model, we will focus on two informational features: *credibility* (the amount of supporting evidence) and *likeability* (the motivational pressure exerted by the subject's goals). The relevant cognitive architecture will include 8 candidate data: 2 doxastic inputs (DI: This agent looks like my husband; D2: I feel uncomfortable with this agent), 3 potential conclusions (CI: There is something wrong with me; C2: There is something wrong with our relationship; C3: There is something wrong with the world, i.e. my husband has been replaced by a physically identical impostor), and 3 inferential rules (II: [DI & D2] →

CI; 12: [DI & D2] → C2; 13: [DI & D2] → C3). In addition, we consider three motivational factors potentially influencing the subject's belief dynamics: MI: I want to be sane (= not CI); M2: I want to have a good relationship with my husband (= not C2); M3:1 want the world to remain «as it should be» (= not C3).

This computational setting allows us to manipulate several different variables, in order to assess their relevance for the development of the Capgras delusion: (i) the amount of *perceptual evidence* regarding the physical appearance of the other person, i.e. the credibility of node DI; (ii) the strength of the *feeling of uneasiness,* i.e. the credibility of node D2; (iii) the credibility of the *different inferential rules* available to the agent, i.e. nodes II, 12, and 13; (iv) the strength of the *motivational influence* exerted over different conclusions, i.e. the likeability of nodes CI, C2, and C3; and (v) the *weight assigned to different factors* (i.e., credibility and likeability) in the process of belief selection.



In this paper, we present the results of a series of computer simulations of this model, to reproduce the behaviour of both Capgras patients and 'normal' subjects. More specifically, we wanted to test three competing hypotheses regarding the cognitive genesis of the Capgras delusion: whereas motivational theories assign to factor (iv) the main role in the formation of the delusion, deficit models emphasise the other variables. In partial contrast with both, the theoretical synthesis championed by McKay and colleagues (2005b) locates the genesis of this syndrome in the interaction of all of the variables.

Other authors (e.g. Sahdra, Thagard, 2003) have used computational modelling techniques to successfully simulate the effects of motivational factors on belief formation in garden variety cases of self-deception. This is the first reported effort, however, to model motivational influences in the genesis of delusions. In future work, we plan to extend this computational treatment to other belief pathologies, e.g. the Reverse Othello Syndrome (Butler, 2000), and cases of persecutory delusions in hearing-impaired subjects (McKay et al., 2005a).

References:

Boutilier, C. (1998). A unified model of qualitative belief change: A dynamical systems perspective. *Artificial Intelligence, 98,* 281–316;

Butler, P.V. (2000). Reverse Othello syndrome subsequent to traumatic brain injury. *Psychiatry: Interpersonal and Biological Processes, 63,* 85–92;

Ellis, H.D., Young, A.W., Quayle, A.H., & de Pauw, K.W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceeding of the Royal Society of London: Biological Sciences, B264,*1085–1092;

Enoch, M.D., & Trethowan, W. (1991). *Uncommon psychiatric syndromes* (3rd ed.). Oxford: Butterworth-Heinemann;

Gardenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states.* Cambridge: MIT Press;

Langdon, R., & Coltheart, M. (2000). The cognitive neuropsychology of delusions. *Mind & Language, 15(1),* 183–216;

McKay, R., Langdon, R. & Coltheart, M. (2005a). Models of misbelief: Integrating motivational and deficit theories of delusions. *Proceedings of the MSB 2005 symposium on «Agents that want and like: Motivational and emotional roots of cognition and action»,* 76–83;

McKay, R., Langdon, R. & Coltheart, M. (2005b). «Sleights of mind»: Delusions, defences and self-deception. *Cognitive Neuropsychiatry, 10(4),* 305–326;

Paglieri, F. (2005). See what you want, believe what you like: Relevance and likeability in belief formation. *Proceedings of the AISB 2005 symposium on «Agents that want and like: Motivational and emotional roots of cognition and action»,* 90–97;

Paglieri, F., Castelfranchi, C. (2005). Influence of social motivation over belief dynamics: A game-theoretical analysis. *Advances in Cognitive Economics.* Sofia: NBU Press, 202–213;

Raschka, L.B. (1981). The Capgras syndrome. *Canadian Journal of Psychiatry, 26(3),* 207–208;

Sahdra, B. & Thagard, P. (2003). Self-deception and emotional coherence. *Minds and Machines, 13,*213–231.

THE RELATION BETWEEN INDIVIDUALS' PRIMARY LANGUAGE AND THE NATURE OF THEIR SHORT-TERM MEMORY CODING OF WRITTEN WORDS

Paul Miller

University of Haifa (Israel)

A first objective of the present study was to determine the nature and efficiency of the strategy/ies used by prelingually deafened, native signers and a hearing control for the temporal retention of written words. A second objective was to reexamine the validity of the primary language coding hypothesis initially proposed by Shand (1979) with reference to the obtained findings. Data was gathered based upon a research paradigm asking the participants to remember series of six written, consecutively presented words and to recognize them