Bayesian Models, Delusional Beliefs, and Epistemic Possibilities

Matthew Parrott
King's College London

(DRAFT: Forthcoming in *British Journal for the Philosophy of Science*)

**Abstract:** The Capgras delusion is a psychiatric condition in which a person believes that an imposter has replaced some close friend or relative. Recent theorists have appealed to Bayesianism to help explain both why a subject with the Capgras delusion adopts this delusional belief and why it persists despite counterevidence. The Bayesian approach is useful for addressing these questions; however, the main proposal of this essay is that Capgras subjects also have a delusional conception of epistemic possibility, more specifically they think more things are possible, given what is known, than non-delusional subjects do. I argue that this is a central way in which their thinking departs from ordinary cognition and that it cannot be characterized in Bayesian terms. Thus, in order to fully understand the cognitive processing involved in the Capgras delusion, we must move beyond Bayesianism.

*1 The Simple Bayesian Model*

*2 Anomalous Evidence and the Capgras Delusion*

*3 Impaired Reasoning*

*4 Setting Priors*

*5 Epistemic Modality*

*6 Delusions of Possibility*

*7 Delusions of Possibility in Different Contexts*

Individuals with delusions report believing incredibly strange things that are nothing like what the rest of us believe. Because this sort of behaviour strikes most of us as obviously irrational, it raises at least two explanatory questions, both of which are at the centre of most contemporary discussions of the nature of delusions. First, there is the question of why someone adopts a delusional belief in the first place. Why do they believe something so obviously false, especially when it is obvious that there is absolutely no evidence in its favour? Secondly, there is the question of why, once it has been adopted, the delusional belief persists over time despite the presence of abundant counterevidence (Coltheart [2007]; Coltheart, *et. al.* [2011]; Langdon [2013]). In recent years, several leading theorists have turned to Bayesian modelling in order to develop answers to these questions (cf. Coltheart, *et. al.* [2010]; McKay [2012]; Davies and Egan [2013]). More specifically, they have adopted a Bayesian approach for explaining the Capgras delusion, the belief that one's close friend or family member, often one's spouse, is a qualitatively identical imposter. Although theorists have developed these models with specific reference to the Capgras delusion, it is commonly thought that a successful model can be extended to other monothematic delusions, which, like Capgras, involve subjects who have delusional beliefs concerning only a single theme (Coltheart, *et. al.* [2011]; Davies *et. al.* [2001]).[i]

Part of the reason Bayesian models looks especially attractive for understanding the Capgras delusion is that there is compelling evidence indicating that the delusion is generated in response to an anomalous experience (cf. Stone and Young [1997]). The

entire Bayesian framework is especially effective at modelling how a person ought to adjust his or her beliefs in response to experiential evidence, so by applying it to a case of irrational belief, like Capgras, we may be able to better understand both why a subject undergoing some kind of an anomalous experience adopts the belief that his or her spouse is an imposter and why this belief persists despite what looks like obvious counterevidence. The Bayesian approach therefore aims to illustrate more clearly in what respects the cognitive processes underlying the Capgras delusion are impaired.

As we shall see in the first four sections of this essay, using Bayesianism to model delusional cognition can indeed be instructive for understanding why someone believes that her acquaintance or spouse is an imposter. But, even though this may help us answer certain questions about the adoption and persistence of the Capgras delusion, there may be important ways in which this delusion is less susceptible to Bayesian modelling. Notably, a Bayesian answer to either the adoption or persistence question only makes sense if we assume that a Capgras subject considers certain candidate hypotheses to be potential explanations for her highly anomalous experiences, including, crucially, the hypothesis that her spouse or close friend is really an imposter. However, no matter how odd or unusual one's experience might be, it is very unclear why someone would ever think this is a candidate for explaining it. Why would the thought that one's spouse is an imposter even be considered as a potential explanation for some unusual experience? The Bayesian framework cannot help us answer this question.

The main proposal of this essay will be that subjects with the Capgras delusion have an abnormal conception of epistemic possibility. The basic notion of epistemic possibility will be clarified in Section 5 but, for now, we can think of it as being

equivalent to what is possible, given what is known - things incompatible with what is known are not epistemic possibilities. For example, I currently think it is epistemically possible that my wife is at home, which is to say that, as far as I know, she might be at home. However, since I also know my wife was at work one hour ago, I do not think it is epistemically possible that she is in China. Indeed, I think it is false that she might be in China.[ii]

In Section 6, I shall claim that Capgras subjects have an unusually broad conception of epistemic possibility in that they think certain things are epistemically possible that non-delusional subjects do not. In particular, I shall argue that non-delusional subjects do not think it is epistemically possible for an imposter to have replaced their spouse or close friend precisely because this is clearly incompatible with many things non-delusional subjects know to be true, including things like 'this is the person I married' and 'this person and I went on holiday last year'.[iii] For this reason, a non-delusional subject would not even seriously entertain the possibility that an imposter has replaced her spouse or friend, even in cases in which they must explain or make sense of some highly anomalous empirical data. If this is right, then simply considering a plausible explanation to be that one's spouse or friend is an imposter manifests a significant departure from ordinary cognition. This suggests that, in addition to modelling how a delusional subject comes to believe a specific hypothesis in the face of an unusual experience, if we hope to fully understand the cognitive processes implicated in the Capgras delusion, we must explain why a subject considers certain things to be potential explanations for her experience and this will require us to move beyond Bayesianism. In Section 7, I argue that this conception of delusional cognition has the

further advantage of allowing us to make sense of why we do not typically regard individuals from different cultures as delusional.

If the proposal of this essay is correct, we will ultimately want to understand what leads a delusional subject to develop such an irregular conception of epistemic possibility. This is a difficult question, a full discussion of which is beyond the scope of this essay. However, if there is a distinct cognitive impairment responsible for a subject's abnormal conception of epistemic possibility, this may appear to present a clear challenge to the well-known two-factor framework for explaining monothematic delusions (cf. Davies, *et. al.* [2001]; Coltheart, *et. al.* [2011]). I shall discuss this potential challenge in Section 8 and also briefly comment on two promising approaches for explaining how a person develops an irregular conception of epistemic possibility that seem worth further exploration.

## 1 The Simple Bayesian Model

The idea behind Bayesian modelling of belief processing is that a person's existing beliefs can be thought of in terms of subjective probabilities or levels of credence that she assigns to various hypotheses.[iv] How a person ought to rationally respond to new evidence can then be captured by a function relating her new evidence to the probabilities she assigns. For example, suppose we have the following two hypotheses:

H1: The mug on my desk contains water

H2: The mug on my desk contains gin

Let's also assume the probability that the mug has water rather than gin is fairly high: $P(H1) = 0.9$ and $P(H2) = 0.1$. This distribution models what Bayesian theorists call my *prior probabilities*, the levels of credence I have in competing hypotheses before considering any evidence. Now, suppose I acquire some new evidence by tasting the liquid in the mug and suppose further that it has a botanical flavour and slightly burns my palate. Call this evidence $E$. The general Bayesian idea is that, when confronted with $E$, a rational subject's beliefs should be updated by a process of conditionalization such that the new probability the subject assigns to each hypothesis is equal to the prior conditional probability of that hypothesis given $E$. According to Bayes's theorem, this can be formulated as follows:

$P'(H1) = P(H1).P(E|H1)/P(E)$

$P'(H2) = P(H2).P(E|H2)/P(E)$

As we can see, the posterior probability ($P'$) of any hypothesis in light of new evidence depends on two things; first how probable that hypothesis is before one acquires any evidence and, second, how well that hypothesis predicts the evidence, or how likely the evidence is given the hypothesis. Thus, in this example, the balance of epistemic reasons favours H2 over H1 if either the prior probability of H1 is comparatively low or the likelihood of $E$ given H2 is comparatively high. Let's suppose the latter is true, $P(E|H2) > P(E|H1)$. Even so, it will be rational to adopt H2 over H1 only if this likelihood ratio favours H2 enough to outweigh its comparatively low prior probability. For instance, if $P(E|H2) = .80$, then it will be rational to believe H2 only if $P(E|H1) < .08$.

Notice that this framework offers us a simple way of understanding how a subject should ideally adjust her levels of credence when faced with *any* kind of evidence. Even if one's evidence is not itself very probable, Bayesianism delivers a clear verdict on how one should respond to it. So even if someone's experience were highly abnormal or even hallucinatory, a Bayesian model would demonstrate the most rational way to adjust one's beliefs. If we apply a Bayesian model to a non-ideal or irrational case, we will therefore get a clear mismatch between how an ideally rational subject should adjust her beliefs and how an actual subject does adjust her beliefs. In certain cases, this may illustrate the ways in which ordinary human beings are less that ideally rational, for instance due to various biases (although see Oaksford and Chater [2007]). And, as we shall see in the following three sections, if the model is applied to a delusional subject, it can illustrate more clearly how someone's belief forming processes may be impaired.

## 2 Anomalous Evidence and the Capgras Delusion

A prominent theory in cognitive neuropsychiatry maintains that the Capgras delusion is caused by an abnormal experience. It has been well established that, in non-delusional subjects, visual recognition of a familiar face is associated with a response in a person's autonomic nervous system. Several years ago, Ellis and Young ([1990]) proposed that in the Capgras delusion the autonomic nervous system is disconnected from a subject's facial recognition system, such that visually familiar faces do not elicit this response. This hypothesis has been confirmed by several experiments (Ellis, *et. al.* [1997]; Ellis, *et.*

*al.* [2000]; Brighetti, *et. al.* [2007]) Thus it seems very likely that a lack of autonomic response to a familiar face is at least partly responsible for the Capgras delusion.

It is important to realize that this lack of autonomic response need not itself constitute the Capgras subject's anomalous experience. People are not consciously aware of their autonomic nervous system and so it would be difficult to see how they could be directly aware of a lack of responsiveness in this system (Coltheart [2005]). Nevertheless, it is not unreasonable to think the abnormality in the autonomic nervous system could generate some kind of irregular conscious experience, perhaps an experience of something being different or wrong in some way.[v] We need not be conscious of the internal operations of the autonomic nervous system in order for its outputs to factor in our conscious experiences.

Nonetheless, it is evident that an unusual experience of an otherwise familiar face would not be a sufficient explanation for the Capgras delusion. Patients who suffer damage to ventromedial regions of the frontal cortex also show diminished autonomic responsiveness to familiar faces (Tranel, *et. al.* [1995]) but they do not adopt the belief that their close friend or family member is a stranger or imposter. It therefore seems that the Capgras subject's belief processing must be impaired in some additional way that a ventromedially damaged subject's is not. For this reason, it is widely agreed that more is needed to explain why subjects adopt the Capgras delusion.

**3 Impaired Reasoning**

In recent years, theorists have developed different Bayesian models to explain how impaired empirical reasoning in response to an unusual experience could give rise to the Capgras delusion (Coltheart, *et. al.* [2010]; McKay [2012]; Davies and Egan [2013]; for a more general discussion see Adams, *et. al.* [2013] and Fletcher and Frith [2009]). To illustrate these models, consider the following two candidate hypotheses:

> *Spouse*: This person is my spouse.

> *Stranger*: This person is an imposter.

As we have seen in the simple the Bayesian picture, whether a subject adopts one these will depend in part on its prior probability and in part on how well it explains new evidence. We have also seen that subjects suffering from the Capgras delusion are presented with anomalous data caused by the fact that they have visual experiences of faces without the normal autonomic responses. For our purposes, let's assume that this generates a relatively unspecific experience:

> *E*: There is something odd about this person.

With this assumption in place, we can model the adoption of the Capgras delusion in a Bayesian framework. Since we know that the Capgras subject does adopt *Stranger*, we know that the ratio of posterior probabilities favours *Stranger* over *Spouse*. This means that either the prior probability of *Spouse* is comparatively low or the likelihood of *E* given *Stranger* is comparatively high. The former strikes most people as implausible, so let's assume the latter is true, $P(E|Stranger) > P(E|Spouse)$. Nevertheless, it would be

rational to adopt *Stranger* only if this ratio sufficiently outweighs its comparatively low prior probability.

When it comes to the Capgras delusion, it might seem obvious that subjects assign a low prior probability to *Stranger*. Indeed, this is the starting point for the Bayesian model developed by Coltheart, Menzies, and Sutton ([2010]). They claim, 'it would seem that a subject might give a very low prior probability to the stranger hypothesis Hs and a very high prior probability to the wife hypothesis Hw in view of the general implausibility of the general plausibility of the second.' ([2010], pp. 277-8) Nevertheless, despite the low prior probability of *Stranger*, Coltheart, Menzies, and Sutton believe that 'the delusional hypothesis provides a much more convincing explanation of the highly unusual data than the nondelusional hypothesis; and this fact swamps the general implausibility of the delusional hypothesis.' ([2010], p. 278) They demonstrate this by using the following probability distribution:

$P(Stranger) = 0.01$

$P(Spouse) = 0.99$

$P(E|Stranger) = 0.999$

$P(E|Spouse) = 0.001$

With these values, we can calculate the posterior probabilities as follows:

$P'(Stranger) \sim 0.91$

$P'(Spouse) \sim 0.09$

Thus, according to Coltheart and his colleagues, the adoption of the Capgras delusion in the face of 'highly unusual data' is not irrational. It is in fact roughly 10 times more

probable given *E*. Anyone faced with anomalous data like *E* ought to update her beliefs to include *Stranger*.

This does not mean we must think the delusion is a completely rational response to *E*. Rather, Coltheart, Menzies and Sutton go on to argue that their Bayesian model illustrates why it is irrational for a subject to *maintain* her belief in *Stranger*. According to their view, soon after adopting the belief, a Capgras subject is confronted with a lot of data that 'should undermine his belief in the stranger hypothesis' ([2010], p. 279). We might suppose this data includes things like friends and clinicians repeatedly telling the subject that he sees his wife or the fact that the alleged stranger knows things only his wife could know. It is independently plausible to think this set of counterevidence is better explained by *Spouse* than *Stranger*, *P*(counterevidence|*Spouse* > *P*(counterevidence|*Stranger*). If so, by standard Bayesian reasoning, a rational subject should discard *Stranger* and update her belief system to include *Spouse* once she becomes aware of the counterevidence. But, the Capgras subject does not do this. We can therefore surmise that her belief processing system is impaired at some stage of belief re-evaluation. How might this happen?

Coltheart and his colleagues suggest that Capgras subjects do not 'accept the evidence of their senses and the testimony of others.' ([2010], p. 281) So, although they respond to *E* in broadly rational way, they respond irrationally to counterevidence. As they describe it, 'it seems as if the new information does not even enter the deluded subject's belief system as data that need to be explained.' ([2010], p. 280) If this is right, then it suggests some kind of cognitive deficit prohibits subjects from appropriately incorporating sensory or testimonial information. Coltheart and his colleagues speculate

that the deficit is caused by damage to the right frontal lobe, specifically to the lateral region of the right frontal cortex. ([2010]; Coltheart [2007])

## 4 Setting Priors

Because the model presented by Coltheart and his colleagues rationalizes the adoption of *Stranger*, Ryan McKay ([2012]) complains that it rests on an implausible conception of prior probabilities. McKay thinks the hypothesis that one's spouse is really a stranger 'represents an exceedingly unlikely - almost miraculous - state of affairs.' ([2012], p. 340) For this reason, he believes it is far more realistic to assign it a prior probability of 0.00027. Correspondingly, he thinks a more realistic value for $P(Spouse)$ is 0.99973. But if we adopt McKay's values, then when the Capgras subject updates her beliefs in response to $E$ (assuming the same values for likelihood used in the previous model), the posterior probability of *Stranger* will be approximately .21, which would be much *lower* than the posterior probability of *Spouse*. Thus, according to McKay's model, adopting *Stranger* is irrational.

If this is right, however, why does the Capgras subject adopt it? One possibility, favoured by McKay, is that the subject's belief forming system is heavily biased toward explanatory adequacy. The general idea would be that Capgras subjects strongly favour explaining novel experiences at the expense of pre-existing beliefs, rather than balancing the demands of explanation with overall belief conservation (cf. Stone and Young [1997]). Aimola Davies and Davies describe this bias as a 'tendency towards acceptance of a hypothesis that explains a salient piece of evidence. ' ([2009], p. 293) In Bayesian terms, an individual who is biased in this way would update her beliefs in a manner

isometric to the likelihood ratio. Thus, in the face of *E* a biased subject will behave as if $P'(Stranger) = P(E|Stranger)/P(E)$.[vi] She will effectively discount her prior probabilities.

Interestingly, once the delusional belief is irrationally adopted in the way McKay proposes, one might think its persistence looks fairly normal. The reason has to do with what McKay seems to think is involved with incorporating a belief into one's belief system. For any hypothesis, if a subject fully incorporates it into her belief system, it would immediately affect her overall distribution of credence. To incorporate a belief in this way just is to adjust other beliefs so as to preserve overall coherence and consistency. Since this would alter the probabilities one assigns to a wide range of things, evidence that may have been very improbable at one time may no longer be; a point nicely stated by Davies and Egan:

It is improbable that a trusted friend should assert, concerning a stranger, that she is the patient's wife.  But it is not so improbable that a trusted friend should assert, concerning a stranger who looks just like the patient's wife and says that she is his wife (an imposter, and a good imposter at that), that she is the patient's wife. ([2013], p. 702)

Thus, if McKay's model is accurate, we may not need to appeal to any further cognitive *impairment* to explain why the Capgras delusion persists.

However, adopting a Bayesian framework does not preclude us from thinking that the persistence of one's delusional belief is irrational. Davies and Egan ([2013]) argue that McKay's model relies on an implausibly idealized picture of the belief system. They claim instead that rather than forming a single unified network, our beliefs are typically

fragmented or compartmentalized. This is what allows us to critically reflect on beliefs without having to acquire new evidence, which is especially useful in cases where beliefs are adopted automatically as pre-potent responses to perceptual stimuli (cf. Egan [2008]). If my entire web of beliefs were adjusted to cohere with every automatic perceptual belief, it would be difficult for me to ever re-evaluate, and subsequently discard, beliefs that are caused by visual illusions or hallucinations.[vii] However, by compartmentalizing, a subject is able to retain her prior levels of credence so that those may be used to reflectively assess automatic responses, which is why we are not stuck with beliefs in visual illusions.

Along these lines, Davies and Egan think adopting *Stranger* is a kind of automatic pre-potent response to *E*. They then argue that, like any belief formed in this way, it is immediately compartmentalized (cf. Gilbert [1991]). As a result, the delusional subject retains her prior levels of credence, which she could use to re-evaluate her belief in *Stranger*. However, whereas an epistemically rational subject would thereby reject *Stranger*, in the case of the Capgras delusion, the belief persists. Davies and Egan speculate that this is because the subject's belief evaluation system is impaired in some way. Upon reflection, the subject is unable to access 'an alternative to the imposter hypothesis that provides a better explanation of the patient's anomalous experience.'([2013], p. 719)

What sort of cognitive impairment might prohibit someone from accessing an alternative to *Stranger*? Davies and Egan offer two suggestions. First, they propose that the patient might suffer from impaired working memory or executive function (cf. Aimola-Davies and Davies [2009]; Feinberg and Roane [2005]), which may be

compounded by the fact that delusional subjects do not adequately understand their situation. For this reason, plausible hypotheses, such as that a stroke has disconnected the face processing system from the autonomic nervous system leading to an unusual experience, are not available to them. But it is not clear why someone would need such a sophisticated explanation for *E*. If the subject retains her prior level of credence in *Stranger*, which the model assumes to be extremely low, wouldn't any alternative be a better explanation?

Their second suggestion is that a Capgras subject's delusional beliefs have failed to be compartmentalized. In that case, the subject's belief in *Stranger* would irregularly become fully integrated with her other beliefs in a way that changes her prior levels of credence. However, notice that if there were this sort of compartmentalization failure, one would expect the belief in *Stranger* to be less circumscribed and to have more of a widespread effect on the belief system than it appears to in most cases. (cf. Tumulty [2011]) In stereotypical cases of the Capgras delusion, we regularly find subjects who fail to act on their delusional beliefs and who frequently report that the delusion is implausible. (cf. Bortolotti [2010])

We have now seen three different Bayesian models of the Capgras delusion, each of which answers the adoption and persistence questions in a slightly different way. These are not exhaustive. We could develop a model of the delusion in some other way, perhaps a way in which both adoption and persistence come out looking rational (cf. Maher [1988]). Yet, however we develop a model, the general Bayesian approach looks like it will be useful for understanding central aspects of the cognitive processes implicated in the Capgras delusion. We need not debate the details of these different

models any further because they all share a common assumption, which I think is worth questioning.

## 5 Epistemic Modality

In a Bayesian framework, if a subject's prior probabilities are fixed, a model will accurately predict how the subject should update her beliefs when confronted with new evidence. But, if we wish to know what specific value of *prior* probability to assign to a hypothesis, the Bayesian framework offers us no assistance.  An important limitation of the Bayesian approach is that it gives us a picture only of how a subject should respond to new information. It is completely silent on how the subject should assign prior probabilities to competing hypotheses before acquiring information.[viii]  However, if we wish to model either rational or irrational cognitive processes in Bayesian terms, we need some reasonable way to determine prior probabilities. When faced with the theoretical question of what probability to assign *Stranger* in our model, it is intuitive to think the most rational assignment would be very low. Indeed, as we have just seen, debates between Bayesian theorists mostly centre on whether the prior probability is set low enough. But what if all of the previously considered models set *P(Stranger)* too high? Indeed, we might reasonably ask why we should assign *Stranger* any positive degree of credence at all.

Given some new evidence *E,* one might think a fully rational subject would consider any metaphysically possible or perhaps any logically possible hypothesis that could explain *E.*  In some cases, this would make the set of candidate hypotheses infinite,

but it is not obviously impossible for a person to, in some sense, consider a countable infinite set of hypotheses.[ix] Nevertheless, even if this were the best way to think of some cases of total ignorance, it is typically more rational for background knowledge to constrain the range of viable hypotheses a subject considers. If I know that *p* is true, it would be straightforwardly irrational for me to consider any anything incompatible with *p* as a possible explanation of *E* by assigning it some positive probability incompatible with the credence I have in *p* (this would violate the standard probability axioms). This is true even though the process of considering a range of hypotheses often takes place unconsciously. Thus, on the standard picture, a rational person considers only those hypotheses that are not ruled out by background knowledge, which is to say only those that are epistemic possibilities.[x]

Epistemic possibilities are those things that are possible given what is known or, equivalently, those things that are compatible with what is known. With respect to probability space, it is quite natural to think of knowledge as having a probability equal to 1. On the resulting view of epistemic modality, which I favour, a proposition is epistemically impossible if and only if it should be assigned a probability equal to zero given what is known. However, certain epistemologists strongly resist this way of thinking about knowledge. They prefer to lower the threshold of probability that a mental state must meet in order to count as a state of knowledge. Although I am not sympathetic with this approach to epistemology, it may be tempting to someone with sceptical tendencies, someone who thinks we are certain of almost nothing but nevertheless know quite a bit. Such a person might think that since we count as knowing things despite having a level of credence less than 1, almost every hypothesis is an epistemic possibility

because almost every hypothesis has *some positive* probability. If one assigns a probability of 0.9 to *p,* then it is reasonable to assign some positive probability to ~*p*, anything less than or equal to 0.1, which might make ~*p* very improbable, even exceptionally so, but not really *impossible*. Therefore, someone might think that even an extremely improbable delusional hypothesis is nonetheless an epistemic possibility as long as it has some positive credence. This is a mistake.

On any view according to which knowledge falls within a range of subjective probabilities less than or equal to 1, we should not think of an epistemic impossibility as equivalent to a probability of zero. A given hypothesis is epistemically impossible/possible only relative to a given body of knowledge. So whether a hypothesis ~*p* (or any *q* that entails ~*p*) is epistemically impossible will depend on the degree of subjective probability one assigns to *p*. If *p* has a subjective probability of 0.9, then ~*p* is epistemically impossible if and only if one assigns it subjective probability over 0.1. Therefore, having *some* positive probability value does not automatically make ~*p* an epistemic possibility; rather it depends on whether the specific value is compatible with the probability one assigns to *p*. Strictly speaking, it is the comparative value of the subjective probability that one assigns to ~*p* that is epistemically possible or impossible. In Bayesian terms, considering an epistemically impossible hypothesis just means assigning it a credence that is incompatible with what is known. Whereas on the traditional picture, this would be any value over zero, on the more relaxed view we are considering here the value depends on the threshold one sets for knowledge. Regardless, on either view, a person considers a hypothesis that is epistemically impossible in virtue

of having a level of credence in it greater than what is permitted by a given body of knowledge.

But *whose* knowledge is relevant for determining whether or not a hypothesis is epistemically possible? It is very natural to think it is only the knowledge of the individual considering the hypothesis. So, we might think that *H* is an epistemic possibility for an agent *a* if and only if *H* is compatible with everything *a* knows. It is, however, widely agreed among philosophers that epistemic possibility depends on more than what any single individual knows. One reason for this is that a person can come to learn that she was wrong about *H* being epistemically possible. This might happen if *a* were to acquire some new information that rules out *H*. For example, if I claim that Peter might be in Paris for the weekend, but then learn from you that he stayed in the UK, it is natural for me to retract my initial assertion. But if what was epistemically possible for me before acquiring this information depended only on what *I* knew, then, at the earlier time, my belief and assertion that Peter might be in Paris would have been correct. It would therefore be wrong for me to retract the previous claim.[xi] Since retraction seems warranted in these cases, the epistemic possibility of *H* cannot depend only on what *a* knows.

A similar reason that the knowledge base relevant for determining epistemic possibilities must include more than what a single person knows is that different people disagree about what is epistemically possible. It seems, for instance, that *a* could believe *H* is epistemically possible and *b* could disagree or contradict *a* on the basis of information *b* possesses. However, if the truth of *a'*s belief depends only on what *a* knows, and the truth of *b'*s belief only on what *b* knows, this would not make sense. *H*

would be epistemically possible *for a* and impossible *for b;* so their disagreement would not be real-they would be talking past each other.[xii]

For these reasons, we should expand the relevant body of background knowledge in our analysis of epistemic possibility. The resulting view is typically that epistemic possibility is determined by the knowledge of some contextually salient group. Thus, Keith DeRose ([1991]) suggests that whether or not *H* is epistemically possible depends on whether any member of a 'relevant community' knows *H* is false. If they do, *H* is not a genuine epistemic possibility, regardless of what *a* individually knows. But, in addition to what the relevant community *actually* knows, there are cases in which it looks like the community could easily come to learn some information that bears on the question of whether *H* is epistemically possible. To accommodate this intuition, Andy Egan proposes that epistemic possibilities depend on what is within the relevant community's epistemic reach:

The idea, though, is pretty clear: *It might be the case that P* is true iff it's compatible with all of the facts that are within some group's *epistemic reach* that P, where what it takes to be within one's epistemic reach can vary across contexts.[xiii] ([2007], p. 8)

One question for Egan's proposal is what counts as being within a community's epistemic reach. I am currently sitting by a computer and can easily access the Internet. Does this extend my epistemic reach to all Internet-accessible facts? It might, but if it does, it is hard to see what difference my immediate spatial proximity to the computer makes. In many contexts, it would not be *that* difficult for me to use some technology to access information relevant to a particular question. Does all this information constrain what is

epistemically possible for me? Again, it might, but then the notion of 'epistemic reach' is not really doing much work.

There are also questions about who counts as a member of the contextually relevant group.[xiv] Does the group consist of only those people in the same room as *a* or just those people with whom *a* intends to be communicating? Or does the relevant group include anyone who could listen to *a*? And why should the group be restricted at all? Why not include the background knowledge of absolutely everyone?

Although these are interesting questions, however we resolve them, it will be true that *a'*s epistemic possibilities are fixed by a body of knowledge that includes more than what *a* knows; it will depend on what is known by others.[xv] We can therefore modify our analysis in accord with Egan's proposal:

*H* is an epistemic possibility for an agent *a* if and only if *H* is compatible with everything that is within the epistemic reach of some group G.

For now, we do not need to settle who to include in G or in what sense information must be in G's epistemic reach. G will certainly include anyone in close proximity to *a* but may include more people, some of whom *a* may not even be aware of. For our purposes, it only matters that the relevant background knowledge includes more than what any single person knows.

**6 Delusions of Possibility**

If the analysis in the previous section is on the right track, *Stranger* is not an epistemically possible hypothesis. The individuals that constitute a typical Capgras subject's epistemic community know many things that are incompatible with *Stranger*. For instance, they regularly have thoughts that depend on their knowingly re-identifying a delusional subject's spouse, including thoughts like 'it was nice to see the two of you [the subject and her spouse] last week,' 'yesterday I saw your [the subject's spouse] keys on the dresser,' or 'this person went to the shop with me last Tuesday' (cf. Evans [1982]). In order for someone to know things like this, they must be in a position to knowledgeably re-identify the Capgras subject's spouse. And, if the friends and colleagues of a Capgras subject do know things that imply that the person claiming to be the delusional subject's spouse is in fact the subject's spouse, *Stranger* is not an epistemically possible hypothesis. It is incompatible with what the delusional subject's epistemic community knows. Since someone with the Capgras delusion believes *Stranger,* it is clear that she takes it to be an epistemic possibility but she is wrong.

This suggests that the Capgras subject has an abnormal conception of epistemic modality. She envisions the space of epistemic possibility to include more that it actually does.[xvi] I think we might naturally think of this as a manifestation of delusional cognition, regardless of whether or not the subject actually comes to believe *Stranger*. Simply entertaining *Stranger* as a candidate explanation, assigning it too high of a prior probability, demonstrates a subject's thinking is irregular. Indeed a quite common reaction to someone with the Capgras delusion is incredulity precisely because it is hard to imagine how anyone within our community could seriously entertain the *possibility* that his or her spouse had been replaced by a duplicate, nonetheless actually believe it.

Compare what would happen to a non-delusional person experiencing *E*. Even if we suppose such a person would want to explain *E* in some way, the first step of such an process would be to consider a set of epistemically possible hypotheses, each of which has some *prima facie* plausibility as an explanation. Through some cognitive process, one would then zero in on the best explanation of *E*. What happens in the Capgras case, however, is that a delusional subject starts out by considering a different set of candidate hypotheses as explanations for *E*. So the way the delusional subject generates potential explanations is itself manifests a departure from normal cognition.

We can even imagine someone who does not actually believe *Stranger* but nevertheless believes it is epistemically possible. In conversations, this person might report things like, 'someone might replace my spouse with a duplicate very easily but luckily for me this hasn't yet happened.' Or, she might anxiously say, 'every morning, I am extremely worried that my spouse might be a duplicate. It has never happened, but it might.' This kind of behaviour would, I think, strike us as delusional. What difference could it make whether or not a person literally comes to believe the proposition? We naturally think something about this way of thinking is wrong simply in virtue of the fact that the person seriously considers hypotheses that we would have ruled out.

Someone may object that even if it is not a genuine epistemic possibility, we cannot be certain that ordinary subjects do not consider *Stranger*; perhaps they do so within a modular subsystem rather than consciously. The idea behind this objection is that a non-delusional subject might consider epistemic impossibilities like *Stranger* within something like a face-recognition module, and, because that module would be unable to access everything the person knows, it would not have access to the knowledge

that would rule out the epistemic possibility of *Stranger*. If this line of objection were

right, a non-delusional subject would, within a face-recognition module, assign a level of

a probability to *Stranger* that is incompatible with what is known outside of the modular

system. However, even if we assume this picture of belief formation as modular, as long

as the modular system operates in accord with Bayesian principles, any level of

probability it assigns to *Stranger* must be compatible with the overall distribution of

probabilities within that system and there is little reason to think *Stranger* would be

epistemically possible *relative* to information contained within a typical face-recognition

module. The sort of knowledge that rules out the epistemic possibility of *Stranger,* for

example, information sufficient for re-identifying a person's face ('this person went to the

shop with me last Tuesday') is plausibly accessible to a face-recognition module (cf.

Davies and Egan [2013], p. 713).  So even though such a module would not have access

to more elaborate hypotheses, such as those about brain damage, it would nevertheless

have sufficient information to rule out the epistemic possibility of *Stranger*. As we have

seen, on some pictures, this would not mean that the level of credence assigned to

*Stranger* within the module is equal to zero, only that it is sufficiently low to be

compatible with what is known.[xvii]

A different objection would claim that no one is ever in a position to know that

some particular individual is not an imposter, so it is not delusional to think she *might* be.

However, this line of objection sets the standard for knowledge at an extremely high

level. It is widely accepted, at least outside of sceptical contexts, that we all know a great

deal. Importantly, we seem to know a great deal about the individual objects we perceive

in our immediate environment, including, crucially, enough information to

knowledgeably re-identify them.[xviii] I know for instance that this fruit in front of me is the red apple I bought at the store on Monday and that I am drinking from the same mug I drank from yesterday. If we have enough evidence to know facts like these, then we typically have enough to know that a friend's spouse is not a qualitatively identical imposter. The sceptic may wish to resist the idea that other member's of a delusional subject's epistemic community know things that are incompatible with the subject's spouse being an imposter, but this will mean that they do not know a great deal.

One might wonder whether it is right to think that a Capgras subject is a member of the same epistemic community as those who know things incompatible with *Stranger*. Might we not think instead that a delusional subject has adopted some different set of epistemic standards, perhaps *because* she is having such highly irregular experiences? Indeed, reporting a delusion in the face of counterevidence could be seen as a symptom of a kind of withdrawal from one's epistemic community. If the membership of G relevant for determining whether *Stranger* is possible for the Capgras subject did not include individuals who know things incompatible with *Stranger*, then it would be premature to conclude that it is an epistemically impossible hypothesis.

However, it is very hard to see whom to include in G if not the individuals with whom a Capgras subject regularly interacts. Most plausibly, discussing the possibility of a hypothesis with someone seems sufficient for the interlocutor to become a member of a contextually salient G and most Capgras patients regularly engage in conversations with family, friends, co-workers, and clinicians, all of whom know things incompatible with *Stranger*. It is therefore highly likely that the background knowledge of these individuals determines what is epistemically possible for a Capgras subject.

## 7 Delusions of Possibility in Different Contexts

One advantage of the proposal that a individual's conception of epistemic possibility can be delusional is that it offers the conceptual resources to help us understand why certain beliefs can be delusional in some cultures but not in others. Dominic Murphy discusses a case of people in the Sudan who believe that trees convey information. As Murphy describes them, these people believe that 'trees record conversations, and are privy to the plans of witches. You can learn what they know by burning an ebony twig, dipping it in water and reading the pattern of ashes in the water.' ([2013], p. 119) Murphy rightly claims that we do not think that individuals belonging to this culture are delusional but he also thinks there is a serious risk this cultural exemption is *ad hoc*. That is, if someone in our culture were to believe trees conveyed information or recorded conversations in the face of counterevidence, we would take them to be delusional. But if our only criteria for classifying people as delusional are evidential, it is hard to see how this distinction could not be *ad hoc*. If someone counts as delusional in virtue of having a belief that is both not based on evidence and resistant to counterevidence, then people from different cultures should also be classified as delusional.

I think we can avoid the risk of making *ad hoc* exceptions by basing them on whether or not a subject's conception of epistemic possibility is delusional. In the Sudanese culture Murphy discusses, it is presumably compatible with what is known by the community for trees to record conversations, which is why the belief that they actually do is not obviously delusional--although notice that we could imagine a case in

which it would be. By contrast, in a very different culture, the notion that trees record conversations is incompatible with what the relevant epistemic community knows, so either believing it or seriously entertaining the idea appears sufficiently delusional. It is right to let cultural considerations affect our assessment of whether an individual's beliefs are delusional, but this is because those considerations determine what is a genuine epistemic possibility for the individual.[xix]

One worry with this line of thought is that it might seem to make it rather easy for entire communities to become delusional.[xx] Suppose that a particular member of the Sudanese community, Juliet, becomes exposed to some on-line lectures in biology. She comes to learn that trees do not really convey information, nor do they record conversations. Since Julie has learned this through testimony, we can assume that she knows trees don't convey information. Nevertheless, Juliet continues to regularly interact with the same people in her community. Sometimes she tells the fellow members of her community what she has learned about trees, yet fails to convince them. Because Juliet is most plausibly a member of the same G as the rest of her community, her actual knowledge that the trees don't convey information or record conversations is incompatible with the community's widespread belief that they do. But it seems implausible to think that Juliet can make the entire community delusional simply be learning about trees.

This objection illustrates how there is a crucial difference between saying someone's conception of epistemic possibility is delusional and saying it is false. Once we acknowledge that epistemic possibility depends on more than what any single individual knows, it is possible that many people, even an entire community, have a

mistaken conception of epistemic space. This is true of the Sudanese who continue to think trees convey information to them and record conversations even after Juliet has learned otherwise. So if we wish to say that these Sudanese individuals are mistaken about epistemic possibility, yet not delusional, but also say that a Capgras subject is delusional about epistemic possibility, we must mean something more than that the Capgras subject's conception of epistemic space is wrong.

The key difference between someone with a false conception of epistemic possibility and someone with a delusional one is that the former's is correctible. In non-pathological cases, learning information will alter one's conception of what is possible. So a non-delusional individual who mistakenly thinks $H$ is epistemically possible and learns some fact $q$ that is incompatible with $H$ or is made aware of some existing incompatibility between $H$ and a subset of what she knows will adjust her conception of epistemic modality by ceasing to think $H$ is possible. This is not something that tends to happen immediately. Juliet's community is not likely to change their beliefs overnight simply because Juliet tells them they are wrong. Indeed, given their conception of what is epistemically possible, they are likely to discount Juliet's comments about trees. However, if these people are not delusional, then persistent exposure to the sort of information or evidence that Juliet learned will lead them to change their way of thinking. If one is disposed to adjust a false conception of epistemic possibility in light of sufficient information, then it is not delusional—it is merely wrong. The Capgras subject is not disposed to behave in this way. Even in cases where she reluctantly acknowledges that her belief seems extremely odd, her conception of whether or not *Stranger* is possible

does not change. It is this irrational persistence of one's conception of epistemic modality that is indicative of delusional cognition.[xxi]

But what if someone with a mistaken conception of epistemic possibility is just extremely stubborn or opinionated? What if Juliet can *never* convince the members of her community that trees don't convey information, no matter how hard she tries? What if they see an abundance of biological evidence and just continue believing that trees convey information? In that case, I do think it is plausible to describe the members of Juliet's community as delusional. Their steadfast resistance no longer seems to be an understandable sociological fact, but instead seems like abnormal cognition. Of course whether or not this is the appropriate reaction to them will depend crucially on the fact the Sudanese are presented with clear evidence by a member of their epistemic community, but if they are, then continuing to believe something that is apparently epistemically impossible does seem to be delusional.

Someone might worry that this would make delusional cognition extremely widespread.[xxii] Suppose, for example, that an overly confident graduate student thinks that he is far superior to other students and suppose further that the entire faculty know this is wrong.[xxiii] It follows from this that the graduate student has a mistaken conception of what is epistemically possible. Is the student obviously delusional? It seems not. But now suppose that the graduate student isn't disposed to change his mind in the face of clear counterevidence. Despite what the faculty attempt to show him he continues to believe in his own superiority. Is the student's uncorrectable and mistaken conception of what is epistemically possible really sufficient for being delusional? I think it may be and I think that any reluctance we might have to categorizing the student as delusional

comes from the fact that the *content* of his belief is not especially bizarre, like the content of the Capgras delusion. We can easily imagine a case in which a graduate student would not be delusional (though he may be arrogant) in thinking he might be superior because it is not incompatible with what is known (perhaps the faculty do not *know* but merely believe the student is not intellectually superior). For this reason, when we are faced with an overconfident student, we may be naturally less likely to react with puzzlement and more hesitant to intervene than in the case of the Capgras subject. But our more measured response does not demonstrate that the student is not actually delusional. Whether they are or are not will depend on how they respond to clear and reasonable contradictory evidence to the belief that they are intellectual superior.

The notion of irrational persistence is a familiar theme in discussions of delusions and this last objection could be equally raised concerning whether or not a subject's strongly held empirical belief is delusional. The worry is that it is not clear at precisely what point a stubbornly held belief becomes delusional. For instance, how much evidence does someone have to ignore before she counts as delusional? I think there is probably no bright line to be drawn here and that our intuitions will vary between different cases. But, especially if the same cognitive processes are implicated in both delusional and ordinary cognition, it should not be surprising if there turn out to be borderline cases. Nevertheless, I think there will also be clear cases in which the way someone thinks about epistemic possibility manifests a delusional pattern of thought, just as there are clear cases in which someone's belief is obviously delusional.

**8 How Many Factors?**

In addition to understanding why a delusion is adopted and why it is not discarded in the face of counterevidence, it now seems we could ask a third, equally important, question about why it is even considered in the first place as a possible explanation. Raising this question might be thought to cause some problems for one of the leading approaches to understanding delusions in cognitive neuropsychiatry, the two-factor framework (cf. Coltheart, *et. al.* [2011]; Davies, *et. al* [2001]). That is we might think that we need a distinct cognitive 'factor' or deficit to answer each of these three questions. The central methodological assumption of the two-factor approach is not that there are only two explanatory questions to be asked about delusional cognition, but the commitment to only two cognitive deficits or impairments being needed to answer these questions. Therefore, whether or not we need to abandon the two-factor approach will depend on how many pathological departures from ordinary cognition are needed to fully explain the Capgras delusion (cf. Davies and Egan [2013]).

The principal claim of this essay has been that the cognitive processes implicated in the Capgras delusion involve a delusional sense of epistemic possibility and that this contributes to the aetiology of the delusion.[xxiv] If this is right, it seems that at least two factors are needed to adequately answer the adoption question: the occurrence of an anomalous experience and whatever causes the subject to assign an irrationally high prior probability to *Stranger*. If these two deficits sufficiently explain why a belief in *Stranger* is adopted then the adequacy of the two-factor framework would depend on whether or not the persistence of a belief in *Stranger* is normal. However, even assuming a delusional conception of epistemic possibility, a Capgras subject's prior level of credence

in *Stranger* may be significantly lower than her credence in *Spouse*. In that case, the cognitive processes responsible for adopting *Stranger* would have to exhibit some kind of bias. Following McKay, we might think of this bias as a third-factor, even before we consider the persistence question. Obviously if a distinct cognitive impairment were then needed to explain the delusion's persistence, it would push us further in the direction of a multi-factor account.

When it comes to explaining the specific factor that is responsible for a subject having a delusional conception of epistemic possibility, I think there are two avenues worth exploring. First, delusional subjects might reason according to some kind of non-standard inference rules. If a Capgras subject were unable to properly deduce the consequences of known truths because she used a different set of inference rules, this could help explain why she assigns an abnormally high positive prior probability to *Stranger*. However, though there is some evidence that schizophrenics operate with different inference rules in certain contexts, there is currently no evidence for thinking a Capgras subject exhibits unusual inference patterns (Selesnick and Owen [2012]; Owen, Cutting, and David [2007]).

A more plausible suggestion for why someone develops a delusional conception of epistemic modality is that the subject lacks the ability to apply relevant background knowledge. The central idea would be that a Capgras would be unable to use what she knows to appropriately restrict the range of hypotheses she considers as explanations. Because her thinking about which things are possible is cognitively isolated from pertinent information, even if we were to emphasize the implausibility of *Stranger*, this would have little to no effect on her conception of epistemic possibility.

Using a body of background information to restrict a range of hypotheses requires some amount of cognitive resources and we have already seen that Capgras patients manifest deficiencies in executive function and working memory (cf. Broome, *et. al.* [2009]; Feinberg and Roane [2005]). These deficits could prohibit someone from appropriately applying knowledge that is incompatible with *Stranger*. However, there is also neurobiological evidence that could help explain why delusional subjects have difficulties cognitively restricting epistemic possibilities. It is fairly well documented that delusional symptoms are correlated with striatal dopamine elevation. The standard account of this is that the aberrant dopamine firing causes inappropriately high saliency to be attributed to experiences. (Corlett, *et. al.* [2007]; Corlett, *et. al.* [2009]) But a high level of striatal dopamine would affect more than experiences. Specifically, it would also plausibly cause people to misattribute salience to passing thoughts as well, which could contribute to those thoughts seeming to be serious possibilities. According to this hypothesis, the surge of dopamine would more or less overwhelm whatever process normally inhibits certain thoughts from becoming candidate hypotheses for explanation.

Once we explain why *Stranger* is generated as a candidate hypothesis, we still need an account of why it is adopted and why it persists. But we have already seen that Bayesian models are helpful for addressing these questions. The limitation of the Bayesian approach is that it does not help us understand a subject's abnormal distribution of prior probabilities. But once we understand why a set of candidate hypotheses includes *Stranger*, a Bayesian framework can help us see why *Stranger* is adopted and maintained.

It is tempting to think that answering the adoption and persistence questions will provide a complete account of a delusion like Capgras. Indeed, from the perspective of

cognitive neuropsychology, it can be difficult to see what else would need to be explained once we have answers to these questions. One aim of this essay has been to show that we need to understand delusional patterns of thinking much more broadly and this requires expanding the range of our inquiry to address additional questions. It is possible that we will discover more than two cognitive factors are implicated in the aetiology of certain delusions but, if we do, that would surely be a step forward.

St Catherine's College
Manor Road
Oxford OX1 3UJ, UK

Faculty of Philosophy
University of Oxford
Radcliffe Humanities, Woodstock Road
Oxford OX2 6GG, UK

matthew.parrott@philosophy.ox.ac.uk

**References**

Adams, R. Stephan, K. E., Brown, H. R., Frith, C. D. and Friston, K. [2013]: 'The Computational Anatomy of Psychosis', *Frontiers in Psychiatry*, **4**, pp. 1-26.

Aimola-Davies, A. and Davies, M. [2009]: 'Explaining pathologies of belief', in M. Broome and L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, Oxford: Oxford University Press, pp. 285-323.

Bayne, T. and Pacherie, E. [2004]: 'Bottom-Up or Top-Down?: Campbell's Rationalist Account of Monothematic Delusions', *Philosophy, Psychiatry, and Psychology*, **11**, pp. 1-11.

Bortolloti, L. [2010]: *Delusions and Other Irrational Beliefs*, Oxford: Oxford University Press.

Broome, M., Matthiasson P., Fusar-Poli, P., Woolley, J., Johns, L., Tabraham, P., Bramon, E., Valmaggia, L., Williams, S., Brammer, M., Chitnis, X., and McGuire P. [2009]: 'Neural correlates of executive function and working memory in the "at-risk mental state"', *British Journal Of Psychiatry,* **194**, pp. 25-33.

Brighetti, G., Bonifacci, P., Borlimi, R., and Ottaviani, C. [2007]: '"Far from the heart far from the eye": Evidence from the Capgras delusion', *Cognitive Neuropsychiatry*, **12**, pp. 189-97.

Campbell, J. [2002]: *Reference and Consciousness*, Oxford: Oxford University Press.

Christensen, D. [2004]: *Putting Logic in its Place: Formal Constraints on Rational Belief*, Oxford: Oxford University Press.

Coltheart, M., Langdon, R. and McKay, R. [2011]: 'Delusional Belief', *Annual Review of Psychology,* **62**, pp. 271-98.

Coltheart, M. Menzies, P. and Sutton, J. [2010]: 'Abductive inference and delusional belief', *Cognitive Neuropsychiatry*, **15**, 261-87.

Coltheart, M. [2005]: 'Conscious Experience and Delusional Belief', *Philosophy, Psychiatry and Psychology*, **12**: 153-57.

Coltheart, M. [2007]: 'Cognitive Neuropsychiatry and Delusional Belief', *Quarterly Journal Experimental Psychology*, **60**, pp. 1041-62.

Corlett, P.R., Krystal, J, Taylor, J., and Fletcher, P. [2009]: 'Why do delusions persist?', *Frontiers of Human Neuroscience* **3**, p. 12.

Corlett, P., Honey, G., & Fletcher, P. [2007]: 'From Prediction Error to Psychosis: Ketamine as a Pharmacological Model of Delusions', *Journal of Psychopharmacology*, **21**, pp. 238-52.

Davies, M. and Egan, A. [2013]: 'Delusion: Cognitive approaches, Bayesian inference and compartmentalization', in K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Sadler, G. Stanghellini, and T. Thornton (eds.), *The Oxford Handbook of Philosophy of Psychiatry*, Oxford: Oxford University Press.

DeRose, K. [1991]: Epistemic Possibilities", *Philosophical Review* **100**: pp. 525-54.

Dowell, J. [2011]: 'A Flexible Contextualist Account of Epistemic Modals', *Philosophers' Imprint*, **11**, 1-25.

Egan, A. [2008]: 'Seeing and Believing: Perception, Belief formation, and the Divided Mind', *Philosophical Studies*, **140**, pp. 47-63.

Egan, A. [2007]: 'Epistemic Modals, Relativism and Assertion', *Philosophical Studies*, **133**, pp. 1-22.

Egan, A., Hawthorne, J. and Weatherson, B. [2005]: 'Epistemic Modals in Context', in G.

Preyer and G. Peter (eds.), *Contextualism in Philosophy,* Oxford: Oxford University Press, pp. 131-68.

Easwaran, K. [2011]: 'Bayesianism II: Applications and Criticisms', *Philosophy Compass* **6**, pp. 321-32.

Ellis, H., Lewis, M., Moselhy, H., and Young, A. [2000]: 'Automatic Without Autonomic Responses to Familiar Faces: Differential Components of Covert Face Recognition in a Case of Capgras Delusion', *Cognitive Neuropsychiatry*, **5**, pp. 255-69.

Ellis, H. and Young, A. [1990]: 'Accounting for Delusional Misidentifications', *British Journal of Psychiatry*, **157**, pp. 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*: *Biological Sciences*, **B264**, pp. 1085-92.

Evans. G. [1982]: *The Varieties of Reference*, Oxford: Oxford University Press.

Feinberg, T. and Roane, D. [2005]: 'Delusional Misidentification', *Psychiatric Clinics of North America*, **25**, pp. 665-83.

Fletcher, P. and Frith, C [2009]: 'Perceiving is Believing: a Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia', *Nature Reviews Neuroscience*, **10**, pp. 48-58.

Gilbert, D. [1991]: 'How Mental Systems Believe', *American Psychologist*, **46**, pp. 107-19.

Kratzer, A. [2012]: *Modals and Conditionals: New and Revised Perspectives*, Oxford: Oxford University Press.

Langdon, R. [2013]: 'Folie a Deux and its Lesson for Two-Factor Theorists', *Mind and Language*, **28**, pp.113-24.

Maher, B. [1988]: 'Anomalous Experience and Delusional Thinking: The Logic of Explanations', in T. Oltmanns and B. Maher (eds.), *Delusional Beliefs*, Chichester: John Wiley and Sons, pp. 15-33.

MacFarlane, J. [2011]: 'Epistemic Modals are Assessment-Sensitive', in A. Egan and B. Weatherson (eds.), *Epistemic Modality*, Oxford: Oxford University Press. pp. 144-178.

MacFarlane, J. [2007]: 'Relativism and Disagreement', *Philosophical Studies*, **132**, pp. 17-31.

McKay, R. [2012]: 'Delusional Inference', *Mind and Language*, **27**, pp. 330-55.

Murphy, D. [2013]: 'Delusions, Modernist Epistemology, and Irrational Belief', *Mind and Language*, **28**, pp. 113-24.

Oaksford, M. and Chater, N. [2007]: *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*, Oxford: Oxford University Press.

Owen, G., Cutting, J., & David, A. [2007]: 'Are People with Schizophrenia More Logical than Healthy Volunteers?', *British Journal of Psychiatry*, **191**, pp. 453-454.

Samuels, R. [2005]: 'The Complexity of Cognition: Tractability Arguments for Massive Modularity', in P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind: Structure and Contents*, Oxford: Oxford University Press, pp. 107-21.

Selesnick, S. and G. Owen [2012]: 'Quantum-Like Logics and Schizophrenia', *Journal of Applied Logic*, **10**, pp. 115-26.

Stone, T. and Young, A. [1997]: 'Delusions and Brain Injury: The Philosophy and Psychology of Belief', *Mind and Language*, **12**, pp. 327-364.

Sturgeon, S. [2008]: 'Reason and the Grain of Belief', *Nous*, **42**, pp.139-65.

Tumulty, M. [2011]: 'Delusions and Dispositionalism about Belief', *Mind and Language* **26**, pp. 596-628.

Tranel, D., Damasio, H. and Damasio, A. [1995]: 'Double Dissociation Between Overt and Covert Recognition', *Journal of Cognitive Neuroscience*, **7**, pp. 425-32.

Williamson, J. [1999]: 'Countable Additivity and Subjective Probability', *British Journal for the Philosophy of Science*, **50**, pp. 401-416.

---

[i] This essay will assume that delusions involve beliefs (for arguments, see Bortolotti [2010]; Bayne and Pacherie [2004]). As we will see, however, it would be a mistake to think that being delusional is equivalent to pathological believing. This essay will also be concerned with the Capgras delusion. However, the main proposal should apply *mutatis mutandis* to monothematic delusions that are structurally similar to Capgras in that their etiology also involves an anomalous experience.

[ii] Notice, however, that I may nevertheless think it is logically or metaphysically possible that she is in China. These are senses of 'possibility' that are not restricted by a subject's background knowledge. In ordinary discourse, there are many contexts in which people express their beliefs about epistemic possibilities by saying that certain things 'might' or 'might not' be the case. (cf. Kratzer [2012])

[iii] This is not to say that a non-delusional subject thinks it is *absolutely* impossible for an imposter to replace her spouse. However, the set of things that are epistemically possible is smaller than the set of absolute or logical possibilities.

[iv] All or nothing beliefs may be thought of as the upper and lower limits of probability space (0 and 1). The relation between all or nothing beliefs and subjective probabilities raises a number of issues that cannot be addressed in this essay (for discussion see Christensen [2004] and Sturgeon [2008]).

[v] Coltheart ([2005]) suggests that this experience might be caused by a prediction error signal. For discussion of prediction error signaling, especially as it relates to Bayesian modeling see Adams, *et al.* ([2013]).

[vi] As McKay notes, this is technically a much more sophisticated function that can capture all hypotheses under consideration. Since in this example we are assuming there are only two hypotheses, we can simplify. The general point is that someone with a bias toward explanatory adequacy will update beliefs in a way that mimics the likelihood ratios. For more detailed discussion, see (McKay [2012]).

[vii] See Egan ([2008]) for further discussion of vision.

[viii] Cf. Easwaran ([2011]). This essay focuses on the sort of Bayesian framework found in contemporary discussions of the Capgras delusion. However, similar questions will arise for more dynamic models (cf. Weatherson [2007]).

[ix] Suppose I tell you I am thinking of some specific natural number. It might seem most rational for you to distribute your levels of credence evenly among the set of natural numbers. (cf. Williamson [1999]). This does not entail that any computational operation on such a set will be tractable, which will likely depend on some further factors (cf. Samuels, [2005]).

[x] Again, this is an idealization in the model. Ordinary subjects may be irrational in certain ways by having epistemically incompatible priors, as we saw is possible in the previous

section's discussion of belief compartmentalization. If belief formation is implemented by non-conscious modular systems, the set of hypotheses a particular module considers may include some that are incompatible with those in other systems.

xi For extended discussion, see (MacFarlane [2011]; DeRose [1991]; Egan, *et. al.* [2005].

xii For further discussion of disagreement, see MacFarlane ([2007]).

xiii DeRose includes a similar clause in his own account of epistemic possibility but phrases it in terms of what the contextually salient community 'can come to know' ([1991], p. 594). Egan's intends for his concept of 'epistemic reach' to do the work of both aspects of DeRose's definition. This is because, according to Egan, both the information of the contextually relevant community and what that community can easily come to know are within *a*'s epistemic reach.

xiv This question is discussed at length in (Dowell [2011]). She argues that the group can fixed by *a*'s intentions. This proposal is difficult to reconcile with certain intuitions people have about cases where eavesdroppers are assessing *a*'s claims about epistemic possibility (cf. Egan, *et. al.* [2005]).

xv Might this make it too difficult for *a* to know what is epistemically possible? I don't see why it would. People learn from others, both about what is actually the case and what is epistemically possible. Naturally, *a* will deliberate from whatever she *thinks* is possible, but, in most ordinary cases, *a* is correct.

xvi Might she also envision it to include less than it actually does? Perhaps, but having a subjective conception of epistemic space that is a subset of what is actually possible does not seem to be delusional. There is a more interesting question of whether the Capgras subject takes the space of epistemic possibility to be broader than it actually is quite

generally or whether this irregularity is restricted to the theme of her delusion. I believe this is an open empirical question. Since the subject actually believes *Stranger*, we know that she has an abnormal conception of epistemic possibility at least with respect to the theme of her delusion, but she may have an abnormally broad conception of epistemic space more generally. In that case, we would predict that were she to have other kinds of unusual experiences these would also generate delusional beliefs. That is, if the Capgras subject has an irregular conception of epistemic possibility generally, the reason her delusional thinking is not more widespread is because she does not have a sufficiently wide range of anomalous experience. Until this is tested, we simply do not know.

xvii What if the knowledge accessible to the module fails to make *P(Stranger)* low enough? Suppose the information accessible to a face-recognition module implies that *P(Stranger)* should be less than or equal to .10 but the knowledge of the entire subject (some of which is inaccessible) implies that *P(Stranger)* should be less than or equal to .08 and that the module's actual credence in *Stranger* is .09? This value would be an epistemic impossibility, even though it would be permissible relative to the information accessible to the module. According to the resulting picture, a non-delusional subject considers an epistemically impossible hypothesis (*Stranger*) at the sub-personal level, even though it never turns up in conscious thought. However, there is little reason to think ordinary cognition works this way. It is very difficulty to envision a case in which some piece of knowledge *K* is both *inaccessible* to a cognitive module and also would lower the probability one ought to assign to a specific hypothesis like *Stranger* by only a very small amount. So although there is no proof that actual empirical reasoning does not

work this way at a subpersonal level (how could there be?), I do not think there is much to be said in its favor.

xviii Indeed there are reasons to think that if we couldn't knowledgably re-identify particulars over time that we would not be able to acquire perceptual knowledge of them, nor would we be able to act on them (cf. Campbell [2002])

xix This is why religious beliefs typically do not strike us as delusional. The epistemic possibilities determined by one's community allow for typical religious beliefs.

xx Thanks to an anonymous referee for both this objection and the following one.

xxi Naturally, if the community's empirical beliefs about trees conveying information to them were revised upon confronting counterevidence, it would not be *ad hoc* to claim they were not delusional. In the context of taking Murphy's worry seriously, however, we are supposing that the empirical belief persists despite evidence to the contrary.

xxii It is important to keep in mind here that delusions are symptoms and not psychiatric conditions. Thus delusions can be present not only in psychosis but also in conditions like obsessive-compulsive disorder and dementia (although the lines between diagnostic categories are often blurry).

xxiii It is important for the objection that the faculty *knows* this and does not merely believe it.

xxiv It is worth emphasizing again that this thesis predicts certain results from an experiment that tests whether subjects have a delusional conception of epistemic modality. It can therefore be empirically disconfirmed.