The Erotetic Theory of Delusional Thinking

(DRAFT)

Forthcoming in Cognitive Neuropsychiatry

Dr Matthew Parrott (corresponding author)
Department of Philosophy
King's College London
The Strand
London, WC2R 2LS
United Kingdom
matthew.parrott@kcl.ac.uk
mobile: +44 (0) 7729 613 690

Philipp Koralus
Fulford Clarendon Associate Professor
Laboratory for the Philosophy and Psychology of Rationality and Decision
St. Catherine's College
University of Oxford
Manor Road
Oxford, OX1 3UK
United Kingdom
philipp.koralus@stcatz.ox.ac.uk

Keywords: Delusion; Erotetic Theory; Reasoning; Schizophrenia; Jumping to Conclusions

Acknowledgments**:** Parts of this work were presented at a graduate seminar at the University of Oxford and at the Language and Cognition Seminar at King's College London. We would like to thank the participants on both occasions for their helpful feedback. We also would like to thank Bill Fulford, Matthew Broome, and Salvador Mascarenhas for many thoughtful comments on an earlier version.

(Abstract word count: 196)

(Main text word count: 6891)

Abstract

Introduction. In this paper, we argue for a novel account of one cognitive factor implicated in delusional cognition. According to the erotetic theory of delusion we present, the central cognitive factor in delusion is impaired endogenous question raising.

Method. After presenting the erotetic theory, we draw on it to model three distinct patterns of reasoning exhibited by delusional and schizophrenic patients, and contrast our explanations with Bayesian alternatives.

Results. We argue that the erotetic theory has considerable advantages over Bayesian models. Specifically, we show that it offers a superior explanation of three phenomena: the onset and persistence of the Capgras delusion; recent data indicating that schizophrenic subjects manifest superior reasoning with conditionals in certain contexts; and evidence that schizophrenic and delusional subjects have a tendency to "jump to conclusions". Moreover, since the cognitive mechanisms we appeal to are independently motivated, we avoid having to posit distinct epistemic states that are intrinsically irrational in order to fit our model to the variety of data.

Conclusion. In contrast to Bayesian models, the erotetic theory offers a simple, unified explanation of a range of empirical data. We therefore conclude that it offers a more plausible framework for explaining delusional cognition.

Keywords: Delusion; Erotetic Theory; Reasoning; Schizophrenia; Jumping to Conclusions

Introduction

In cognitive neuropsychology and neuropsychiatry, it is typical for both schizophrenia and delusions to be characterized in terms of an abnormality in a subject's ability to reason (e.g., Coltheart et. al., 2011; Davies and Egan 2013; Garety and Freeman, 1999). There is strong empirical support for this characterization. First, widely replicated studies have shown that both delusional subjects and subjects with schizophrenia exhibit a tendency to 'jump to conclusions' on probabilistic reasoning tasks (Huq et. al., 1988; Garety, et. al. 2005; Garety and Freeman, 1999, Fine et. al., 2007; cf. So, et. al. 2012, Langdon, et. al. 2010). Second, in spite of research highlighting cognitive or neurobiological disturbances that plausibly contribute to the generation of certain monothematic delusions (Stone and Young, 1997, Blakemore et. al., 2002, Kapur, 2003), evidence indicates that abnormal reasoning is also implicated in the onset of these delusions (for review see Coltheart et. al. 2011 or Bell et. al. 2006). Finally, recent experiments suggest that schizophrenic subjects exhibit irregular performance on reasoning tasks involving conditionals, including better performance than controls on certain problems (Mellet et. al. 2006; cf. Kemp, et. al. 1997, Owen, et. al. 2007).

To improve our understanding of these irregular patterns of reasoning behavior, we need to understand precisely how the reasoning capacities of psychiatric subjects differ from those of normally functioning individuals. Yet, existing theoretical models have failed to offer a clear picture of this. Investigations of reasoning in delusional and schizophrenic subjects have largely been conducted from within a Bayesian framework (Davies and Egan, 2013; Fine et. al. 2007; Parrott, forthcoming). In this framework, human reasoning consists in cognitive processes governed by the probability calculus because the Bayesian approach conceives of the aim of reasoning as solving informational problems in conditions of uncertainty (Oaksford and Chater, 2007). Deficits in reasoning are conceptualized as departures from a Bayesian ideal and detailed formal models are developed to capture the specific ways in which delusional or schizophrenic subjects represent probabilities or exhibit probabilistic biases. Similarly, to the extent that more typical patterns of human reasoning and decision making seem less than fully rational, these are also explained by formal models which illustrate precisely how they deviate from some Bayesian ideal (Oaksford and Chater, 2007). Despite the widespread popularity of this Bayesian approach in the cognitive sciences, we will argue that it faces serious challenges.[1]

Our primary objective of this paper is to draw on the recently developed erotetic theory of reasoning (Koralus and Mascarenhas, 2013) in order to model distinct patterns of anomalous reasoning exhibited by psychiatric patients. The erotetic theory conceives of the aim of reasoning as asking questions and answering them as quickly as possible. Thus, on the erotetic theory, reasoning deficits of the sort we find exhibited by psychiatric patients are conceptualized in terms of the way they ask questions or in terms of how they go about answering those questions. In brief, we propose that we can make sense of the pattern of reasoning in delusional patients as stemming from reluctance to endogenously raise questions during the reasoning process. The idea is that in ordinary people, endogenous, or "self-generated" questions mitigate the frequency of various reasoning fallacies (Koralus and Mascarenhas 2013), and that this barrier is impaired in delusional patients.

We will begin by sketching the erotetic theory of delusional thinking. We will then explain how this theory can be used to model key experimental data points. In each case, we will argue that the explanation offered by the erotetic theory is superior to that available to the Bayesian.

The Erotetic Theory of Delusion

Any adequate theory of our capacity to reason has to solve both the problem of success and the problem of failure. By the problem of success, we mean the problem of explaining how our capacity for reasoning is robust enough to make science and modern societies possible. By the problem of failure, we mean the problem of explaining the fact that humans systematically commit fallacies of reasoning.

The intuitive idea at the foundation of the erotetic theory of reasoning (Koralus and Mascarenhas, 2013) is that human reasoning fundamentally proceeds by raising questions and trying to answer them as quickly as possible. This idea is made mathematically rigorous using tools from set theory and formal semantics, making it possible to calculate concrete predictions. Koralus and Mascarenhas have argued that the erotetic theory is both clearer on its predictions and more empirically accurate than the best competing theories of propositional reasoning, though this debate is not at issue for our purposes in this paper. The erotetic theory captures well-documented systematic fallacies of reasoning that are surprisingly compelling. For example, given the premises "John and Bill are in the garden, or else Mary is" and "John is in the garden," up to 90% of participants conclude "Bill is in the garden" (Walsh & Johnson-Laird, 2004). The erotetic theory of reasoning (Koralus and Mascarenhas, 2013) holds that naïve reasoners treat successive premises as questions and maximally strong answers to them, even if they do not look like questions. A reasoner will therefore take the disjunctive premise "John and Bill are in the garden, or else Mary is" to pose the question of which of the disjuncts is the case. In effect, the reasoner is asking, "am I in a John and Bill situation or in a Mary situation?" If she then accepts as a second premise "John is in the garden," she will interpret it to be as strong an answer as possible to the question in context. As luck would have it, "John is in the garden" is part of the first answer to the question at hand, and not the second, so she will conclude that the question in context has been answered: 'John *and Bill* are in the garden.' However, this is a fallacy, as it neglects the possibility, compatible with the premises, that Mary and John are in the garden but Bill is not. The foregoing example is the tip of an iceberg of systematic fallacies captured by the erotetic theory (Koralus and Mascarenhas, 2013).

The erotetic theory does not just predict fallacies. It also explains how our natural reasoning capacities allow for the possibility of valid reasoning by classical standards. There is an idealized reasoning strategy using our natural cognitive resources that provably yields classical soundness and completeness (Koralus and Mascarenhas, 2013). What allows naïve reasoning to respect classical validity is the systematic posing of further questions in the reasoning process. In a formally precise sense, questions make us rational. In particular, what separates reliably valid reasoning from fallacy prone reasoning is the extent to which we raise enough further questions *endogenously*, or on our own, as we reason with what is directly prompted by our premises. Different individuals may be better or worse at raising enough questions on their own in their reasoning process at different times, accounting for differences in performance. Some individuals may in fact be particularly impaired in their ability to endogenously generate their own questions to facilitate correct reasoning.

We shall draw on the erotetic theory of reasoning to propose a model of the cognitive factor in a general multi-factor model of delusions (Coltheart, et. al., 2011; Davies, et. al., 2001; Davies and Egan, 2013). Our hypothesis is that the cognitive impairment responsible for the anomalous reasoning exhibited by delusional and schizophrenic subjects is their impaired endogenous question-raising. What we mean by a deficit of "endogenous" questioning is simply a lack of "self-initiated" questioning. The idea is that while someone with this deficit would have no problem taking on board and answering questions that are put to her by someone else, or default questions that are strongly associated with external stimuli (for example, we hypothesize that if we are presented with a person walking into a room, this by default raises the question "who, if anyone, is this among people I know?"), she would have trouble generating further questions on her own that are not as directly prompted by external influence. Distinctions between self-initiated and externally stimulated versions of cognitive operations seem to already have been observed elsewhere in medical science. One might draw a parallel to certain movement disorders that leave patients unable to initiate movement but allow them to, in some cases, execute motor programs that are directly prompted by an external stimulus (like catching a ball thrown at them).

What this means is that delusional and schizophrenic subjects raise the same sorts of default questions in response to external stimuli as typical individuals do, but with fewer alternatives envisaged or with fewer follow-up questions. In particular, we suggest that these patients have a much lower tendency to raise questions that would depend on abandoning or modifying the initial question directly prompted by what is presented to them. In this technical sense, delusional and schizophrenic subjects are simply less inquisitive. The result would be that delusional thinking is an extreme manifestation of a general human tendency to answer our questions quickly, which is unmitigated by a countervailing tendency to raise further questions to prevent missteps.

Some crucial aspects of this hypothesis are worth emphasizing. First, nothing about the proposed cognitive *processes* is intrinsically irrational or intrinsically different from those we would find in ordinary individuals. What differs is merely the extent to which certain processes (e.g. self-generated question-raising) are available. Koralus and Mascarenhas (2013) and Koralus (*under review*) have independently proposed that what accounts for differences between naïve fallacious reasoning and decision-making and ideally rational cognition is whether enough questions are raised in the reasoning and decision-making process. The account we propose of delusional and schizophrenic reasoning suggests patients have an extreme version of a tendency that already exists in the general population. This obviates the need for having to claim, as a Bayesian would, that if neural damage causes a delusion, it has to bring about a new, intrinsically mistaken epistemic attitude, such as, for example, in the case of Capgras delusion, implausibly high priors for the hypothesis that the person who everyone says is the patient's wife is not in fact his wife. On the erotetic theory, we can say something that seems more attractive. The damage that might lead to delusional cognition yields irrational beliefs through creating a failure to inhibit certain aspects of normal reasoning processes that ordinarily (but not without fail, even in normal populations) prevent us from drawing fallacious inferences. This fits with the more general observation that certain inhibitory cognitive control operations are impaired in schizophrenic patients (Chan, et. al., 2006; Henik and Salo, 2004; Orem and Bedwell, 2010). The idea here is that endogenously raising questions serves as an inhibitory mechanism that ordinarily moderates a general tendency to overestimate the extent to which given information answers our questions and that this mechanism is impaired in relevant patients.

A second aspect of the erotetic theory worth emphasizing is that the way cognition is proposed to differ in delusional and schizophrenic patients can make sense of the fact that patients can, in special cases, manifest *improved* reasoning performance, relative to non-psychiatric populations. For example, in one of the sections to follow, we will discuss studies on conditional reasoning tasks in which delusional patients performed better than typical individuals. As we will show, on the erotetic model, these sorts of performance advantages are actually to be expected. The key to our explanation is that a *moderate* amount of question-raising can sometimes yield worse results than both raising no questions at all and raising questions exhaustively. This may remind one of the old adage that *a little bit* of philosophy is a dangerous thing. We will now consider key data points on delusional thinking in turn.

Capgras Delusion

The Capgras delusion is a condition in which someone believes that an imposter has replaced one of her close friends or relatives.[2] A prominent theory in cognitive neuropsychiatry maintains that this delusion is caused in some way by the occurrence of an abnormal experience. In non-delusional subjects, visual recognition of a familiar face is typically associated with a response in a person's autonomic nervous system. Ellis and Young (1990) proposed that in the Capgras delusion, a subject's autonomic nervous system is disconnected from her facial recognition system, such that familiar faces do not elicit this response. This hypothesis has been experimentally confirmed (Brighetti, et. al., 2007; Ellis, et. al., 1997; Ellis, et. al., 2000; Stone and Young, 1997). Therefore, it is plausible that an abnormal experience is at least partly responsible for the onset of the Capgras delusion.[3]

However, an irregular experience is not sufficient for explaining the delusion. Subjects with damage to ventromedial regions of the frontal cortex also manifest diminished autonomic responsiveness to faces but do not adopt the delusional belief that their friend or family member is an imposter (Tranel, et. al., 1995). Some additional cognitive deficit is plausibly implicated in the etiology of the Capgras delusion. We are proposing an account of what this further *cognitive* deficit consists in.

Capgras Delusion in the Erotetic Theory

What needs to be explained in the Capgras delusion is why delusional patients are convinced that someone close to them, such as their wife, is a stranger. We suggest that generally if somebody appears in front of us, regardless of whether we are delusional, this naturally raises the question of who this person is among people we know. If it is not someone we can identify as someone we know, this then raises the question of who this stranger is. We think that this construal is plausible because "making sense" of a person who just walked in appreciably gives rise to a hierarchy of tasks. The first task is to retrieve the "file" in one's knowledge base that corresponds to that person. The second task, only arising if the first task fails, is to create a new "file" for the apparent stranger.

In the ordinary course of events, the question of who someone is among people we know is rapidly settled by familiar appearance, voice, and similar obvious information. But as we have already seen, the onset of the Capgras delusion is correlated with a highly anomalous experience, a "feeling" or "sense" of unfamiliarity. Since this is not sufficient to bring about a delusionary misindentification, we must explain how an additional cognitive factor would cause a patient to conclude that a person in front of them is a stranger due to a feeling of unfamiliarity about that person.

According to the erotetic theory, the question of who someone is among people we know, determines a fixed set of alternatives consisting of those people that we know. We propose that subjects represent those alternatives as bundles of features. For example, we might represent our doctor as having brown hair, being tall, wearing a white coat, speaking with an Australian accent, etc. and we might represent our friend, call him "Jack," as having blond hair, being tall, and speaking with a South African accent, etc. Note that every explanation of the Capgras delusion needs to start with the observation that the delusion seems to be limited to misidentifications of people with whom the patient has a special, close relationship. We suggest that it is plausible that we represent people with whom we have a special relationship of this sort in a way that includes a feature we might call "closeness" or "emotional connection" in the representation of the person in question. Let's call this feature the "C-feature." Phenomenologically, someone might represent his wife as having the C-feature in the same way in which he might represent her as having a certain eye color. That our putative C-feature has a quasi-perceptual nature seems to be supported by phenomenological reports. For example, Young and colleagues studied a subject who claimed that 'there's been someone like my son's double which isn't my son. I can tell my son because my son is different. ...but you have got to be quick to notice it (Young, et. al., 1993, pg. 696; cf. Coltheart, 2005; Stone and Young, 1997).

To make the example concrete, suppose the patient knows three people, his wife, his doctor, and his friend Jack. Then the question of 'who this is among people I know' can be represented along the following lines, as a set of alternative possible answers (following Koralus and Mascarenhas, 2013):

(1) {C-feature&wife&short, doctor&brownhaired&white_coat&tall, Jack&blond&SA_accent&tall}

Patients and non-patients alike would proceed to try to answer this question with whatever information about features is available to them. Suppose the patient notices that the person in front of him is wearing a white coat. If he treats that information as a maximally strong answer to his question, he will be disposed to conclude that the person is his doctor, since none of the other alternatives match this feature. Now, suppose we present the patient with his wife, who is short. The feature of being short eliminates both Jack and the doctor from consideration. Now, suppose that because of a neurophysiological impairment the patient represents the person in front of him as *not* *intimate*, as having a feature that amounts to *not-C*. This feature would eliminate his wife from the set of alternatives as well. As a result, the answer computed for "who is this among people I know?" is "nobody I know." Now, we have the conclusion that the person is a stranger. *If we hold this conclusion fixed*, it does not seem like a *further* mistake on the part of the patient to speculate on various impostor scenarios, the nature of which will be influenced by general anxiety levels and other beliefs.

We suspect that ordinary life is full of momentary instances of misidentification that we immediately correct because we tend to endogenously raise further questions. A colleague walks in after a makeover, we briefly mistake her for a stranger because of the radical change in appearance, but we quickly adjust after a moment's reflection. Our proposal is that this kind of reflection involves endogenously raising a question and that delusional patients fail to systematically raise these sorts of questions.

Many subjects seem to fully recognize that the impostor scenarios are extremely unlikely (Alexander, et. al. 1979; Stone and Young, 1997). On the proposed view, the problem is that this still does not suffice to make them raise further questions that would have *I am misperceiving due to illness* as a possible answer ("who is this among people I know?" does *not* admit of this answer!). Reports on patient conversations suggest that is in fact possible to momentarily lead individuals with the Capgras through a chain of reasoning (Breen, et. al. 2007; Coltheart, et. al., 2007; Coltheart, et. al., 2011; Landa, et. al. 2006) toward the conclusion that they are misperceiving. At the same time, however, leading subjects through this reasoning does not seem to have lasting effects (Colheart, et. al., 2007; Colheart, 2007). In sum, besides the onset of Capgras delusion, we need to explain its persistence.

We propose to explain the persistence of the Capgras delusion as follows: the impostor belief keeps re-generating from a question that keeps arising by default whenever a person walks into the room (e.g. “who is this among people I know?”). Lasting avoidance of the delusional conclusion would have to involve the patient him or herself raising the relevant follow-up questions whenever the patient is momentarily mislead by quickly answering this default question. However, on the erotetic theory of delusion, it is precisely endogenous question-raising that is hypothesized to be deficient. This explanation makes sense of the fact that patients can be momentarily argued out of the their delusional conclusion, only to fall back into it later. On our reading of the literature, the recalcitrance exhibited by patients tends to be like that of a rubber band that snaps back to its original position once external prompting is removed. Although sufficient questioning by clinicians can bring patients to doubt or momentarily give up their delusional conclusion that, e.g. their wife is an impostor,as soon as the external questioning is ended and the wife is encountered again afresh, the impostor conclusion returns.

We think that this explanation has the following virtues. First, it only appeals to mechanisms that have been proposed to independently make sense of ordinary reasoning. We have just used a simple question/answer process to arrive at an apparently delusional conclusion. Fundamentally the same process has been proposed to underlie ordinary reasoning (Koralus and Mascarenhas, 2013). Secondly, it does not appeal to intrinsically irrational processes or to intrinsically irrational epistemic attitudes, which reduces the burden of explaining how not-so-subtle brain damage could give rise to delusions.

Moreover, it allows us to acknowledge that delusion patients themselves often acknowledge that their delusional beliefs are antecedently very implausible.

Contrasting Bayesian Explanations of Capgras

The central methodological assumption of the Bayesian framework is that when presented with some piece of new evidence *E*, a rational subject's beliefs should be updated by a process of conditionalization such that the new probability the subject assigns to a hypothesis when faced with *E* should be equal to the prior conditional probability of that hypothesis on the evidence (Coltheart, et. al*.,* 2010; McKay, 2012; Davies and Egan, 2013). Since we know that a Capgras subject actually believes an "imposter" hypothesis, we can conclude that the subject's ratio of posterior probabilities favours it over alternative hypotheses. So either the subject's prior probability in those alternatives is comparatively quite low or the prior degree of confidence she has in the likelihood of *E* given the imposter hypothesis is comparatively high.

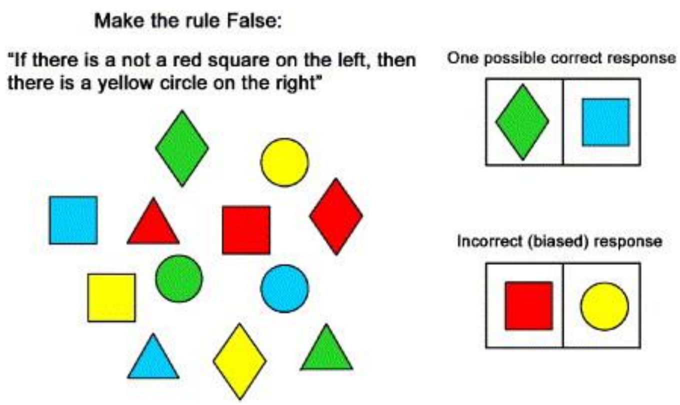
The central question for the Bayesian approach is therefore what values should be assigned to the prior probabilities within a formal model (cf. Parrott, forthcoming). Some models assume that delusional subjects assign a high prior probability to the imposter hypothesis (cf. Coltheart, et. al. 2010) whereas others claims the subject's prior in that hypothesis is low and therefore discounted in her subsequent reasoning (McKay, 2012).

Adopting priors on a proposition that are absurdly high in light of background knowledge or discounting one's prior probabilities entirely are both irrational. If we have to build those irrationalities concerning a particular proposition into a Bayesian model in order to make sense of the irrationality of delusional thinking, it looks like it is not really Bayesianism that is doing the explanatory work but the irrational epistemic state (absurdly high priors) that is postulated for the specific delusional belief to be explained. In other words, a Bayesian account of the Capgras delusion presupposes either some intrinsically irrational prior probability distribution or some kind of probabilistic bias. But, there is no clear way to give a Bayesian explanation of either of these. Additionally, either presupposition is difficult to reconcile with the fact that delusional patients often explicitly acknowledge that the impostor hypothesis seems to be rather improbable (Alexander, et. al., 1979; Stone and Young, 1997; cf. Startup, 1997).

By contrast, as we have seen, the erotetic theory explains delusional thinking as the consequence of a tendency to quickly answer questions that is also present in ordinary thinking. The difference between delusional and normal thinking is located in the *lack of a safeguard* against drawing misguided conclusions, namely a lack of a the tendency to endogenously raise further questions.

Reasoning with Conditionals

In certain cases, a delusion-related diagnosis of schizophrenia can in fact improve performance on reasoning tasks.[4] Mellet and colleagues designed an experiment that required participants to falsify conditional statements by manipulating colored shapes (Mellet, et. al., 2006). In their study, both schizophrenic subjects and non-psychiatric controls were presented with an array of colored shapes and a conditional rule pertaining to those shapes that they were then asked to falsify. There were two conditions: a control condition, in which the consequent of the conditional rule is negated, e.g., 'if there is a red square on the left, then there is not a yellow circle on the right', and an experimental condition, in which the antecedent is negated, e.g., 'if there is not a red square on the left, then there is a yellow circle on the right' (see Figure 1). Participants were asked to arrange the shapes in a manner that falsifies the conditional rule.



(Taken from Mellet, et. al. 2006)

Mellet and colleagues report that the performance scores of both schizophrenics and non-psychiatric controls in the control condition was nearly perfect; both groups are able to easily falsify a conditional rule with a negated consequent. However, they found a wide divergence in scores when 'not' is inserted into the antecedent of the rule. In this condition, schizophrenic participants performed significantly better at falsifying the conditional.

The Erotetic Theory and Conditionals

What needs to be explained is that schizophrenic patients were noticeably better than non-psychiatric controls at falsifying a conditional rule, but only with a negated antecedent, whereas their performance was equivalent to non-psychiatric controls when the rule lacked a negated antecedent. In addition to explaining the superior performance of schizophrenic patients, we want to understand why non-psychiatric controls fail to select the correct arrangement of shapes to falsify the conditional rule they are presented with only in the experimental condition (with the negated antecedent). We also want to explain why controls tend to select a particular arrangement, namely a red square on the left and a yellow circle on the right.

To account for these facts, we need some auxiliary hypothesis about the reasoning strategies employed by schizophrenic patients and controls on this particular task. An explanation of the three patterns mentioned above will be explanatorily interesting if we do not build the asymmetry between negated and non-negated cases into our auxiliary hypothesis.

We therefore propose that non-psychiatric participants adopt the following reasoning strategy:

1. Take the model of the statement about the left and right objects.
2. Make a supposition about the left object followed by a supposition about the right object, choosing, if you can, objects that some alternative in the model is committed to.
3. If those suppositions yield a contradiction, choose the corresponding objects.

4. Else, try again with a different set of suppositions.

We have already suggested that the difference between schizophrenic/deluded and non-deluded subjects is that the former have trouble endogenously raising questions. Suppositional reasoning, within the erotetic framework, is an inquisitive reasoning strategy (formalized in Koralus and Mascarenhas, 2013). It amounts to endogenously raising an additional question, roughly, "am I in a situation in which my supposition and its consequences hold, or in a situation in which my supposition is false?" Since, nothing in the experiment designed by Mellet and colleagues explicitly prompts participants to adopt a suppositional strategy, the erotetic theory of delusion suggests that schizophrenic subjects will not spontaneously adopt it. So, if you have a deficit in being endogenously inquisitive, you would try to solve the problem directly, using only what is suggested by the prompt. As it turns out, the direct reasoning approach is computationally more cumbersome (which is why we suspect more inquisitive "normal" participants adopt the suppositional strategy: it seems superficially easier as it requires fewer alternative possibilities to be represented. For details, see Koralus and Mascarenhas, 2013) but it yields the right results in this case, while the suppositional approach will lead one astray only in the version of the task involving negation in the antecedent (we relegate the formal derivations using the Koralus and Mascarenhas, 2013 system to an appendix). Moreover, the suppositional approach leads to the particular fallacious choice most control participants actually made, namely red square on the left and a yellow circle on the right. In sum, the erotetic theory of reasoning, together with a modest proposal for how participants strategize about this problem using suppositional reasoning, and the core idea that schizophrenic subjects do not readily have endogenous use of inquisitive reasoning, yields the observed pattern of data. It explains the relative difficulty non-psychiatric subjects have with falsifying rules with negated antecedents and provides an account of why schizophrenic subjects do better in this case. In addition, it explains the specific mistake made by control participants.

The erotetic theory is therefore able explain the complex pattern of data presented by Mellet and colleagues with the auxiliary hypothesis that by default, non-psychiatric subjects adopt a suppositional reasoning strategy. Since we have good reasons to think that individuals often use suppositional reasoning to address problems that otherwise require us to represent many alternative possibilities (Johnson-Laird 2008), we think this assumption is well-motivated. It is then a consequence of our hypothesis about a lack of endogenous question-raising, that schizophrenic subjects would not adopt suppositional reasoning and get a correct answer in the case at hand.

Contrasting a Bayesian Approach to Conditionals

As we have seen, an adequate model of the falsification data should explain three things. First, it should explain why *both* schizophrenics and non-psychiatric controls successfully falsify conditionals with negated consequents (if there is a red square on the left, then there is not a yellow circle on the right). Second, it should also account for the behavioral asymmetry for non-psychiatric controls when the negation is shifted to the antecedent of the conditional (if there is not a red square on the left, then there is a yellow circle on the right). Finally, the model should predict the statistically prevalent response given by non-psychiatric control subjects in the negated antecedent condition; it should help us understand why they tend to place a red square on the left and a yellow circle on the right. While we think Bayesian formal illustrations of the data could be given, we simply do not see how a Bayesian approach would fulfill our desiderata for an explanation.

The key question for the Bayesian would be how to capture the way participants interpret the instruction 'falsify this conditional'. From a Bayesian perspective, it is plausible to think of this as requesting subjects to find some relevant arrangement of shapes that has a conditional probability greater than the conditional probability of the statement they are being asked to falsify. Thus, we could formally model the observed behavior of schizophrenic subjects in the experimental condition as follows:

Falsification: P(yellow circle | ~red square) < P(~ yellow circle | ~ red square).

We can see that if it were true that the first conditional probability (yellow circle on the right given no a red square on the left) were less than the second conditional probability (something that is not a yellow circle on the right given no red square on the left), then the conditional rule would be falsified. The presence of something that is not a red square on the left would make it *more likely* that there is *not* a yellow circle on the right.

However, non-psychiatric controls exhibit a different pattern of behavior:

Erroneous falsification: P(yellow circle | ~red square) < P(yellow circle | red square).[5]

Since these two conditional probabilities are independent of each other, the value of the right-side cannot falsify the one on the left. Yet, even if though Erroneous falsification does formally illustrate the observed behavior of non-schizophrenic subjects, it does not help us understand why these subjects think P (yellow circle | red square) is *relevant* to the falsification task they are given. Non-psychiatric controls do not exhibit a general problem understanding the falsification task, for they perform very well in the control condition. So why would do they struggle to find a relevant conditional probability only in the experimental condition? The Bayesian framework does not explain the crucial issue, which is why this pattern of behavior is exhibited.

Jumping to Conclusions

Delusional subjects exhibit a strong tendency to 'jump-to-conclusions' when given a cognitive task that concerns probability (Fine et. al 2007; Garety and Freeman 1999; Garety et. al. 2005 Langdon et. al. 2010; So et. al. 2012). There is also evidence that this tendency is exhibited by subjects who are at risk of developing psychosis (Broome et. al. 2007) and by schizophrenic subjects who are not delusional (Fine et. al. 2007; Moritz and Woodward 2005; Menon et. al. 2006).

In the standard experimental setup to test for this jumping-to-conclusions bias, participants are given some version of the so-called 'Beads Task'. They are shown two jars, each of which contains a different ratio of colored beads. In a standard case, Jar A contains a ratio of 85% red beads to 15% blue beads and Jar B 85% blue beads to 15% red beads. In a 'draws to decision' condition, the experimenter selects beads from a predetermined jar, which she tells the participant has been chosen at random. After each selected bead has been shown to the participant it is returned to the jar so as to preserve the original ratio. In this condition, participants are asked to guess from which of the two jars the experimenter is choosing beads and to stop the experiment from continuing once they are certain of which jar has been chosen. Subjects who have a tendency to jump to conclusions require fewer draws before deciding from which jar beads are being selected. There is a widespread consensus that in the ‘draws to decision’ condition delusional subjects exhibit a clear tendency to jump to conclusions, requesting significantly fewer draws than controls (Garety and Freeman 1999). In the classical version of the Beads Task, they request between 2 to 3 draws compared with 4 to 6 draws for non-psychiatric controls.

The Erotetic Theory and Jumping to Conclusions

The observation to be explained is that delusion patients need much fewer consecutive draws to settle from which jar the beads are being selected. We can relate this task to asking and answering questions. There is independent motivation for the idea that we assess overrepresentation of one possible outcome in a probabilistic setting (e.g. “is the coin more likely to come up heads than tails?”) by looking at whether we can find patterns indicative of a bias (Bar-Hillel and Wagenaar, 1991; Falk and Konold, 1991; Lopes and Oden, 1997). This would make it plausible that the question of whether we are drawing from Jar A or Jar B ends up being represented as a set of alternatives in which we have Jar A and various draw sequences diagnostic of the bias inherent in Jar A, and alternatives in which we have Jar B and various draw sequences diagnostic of the bias inherent in Jar B. A possible gloss on this question might be, “do I have Jar A that supports sequences like R, R, R; B,B, R, R, R, etc. or Jar B that supports sequences like B B, B, R, R, B, B, B, B, B, etc.?” The minimal “sequence” would be just a single draw, so the minimal question would be tantamount to the question, "do I have Jar A, which has red on its first draw, or do I have Jar B, which has blue on its first draw?" If delusional and schizophrenic patients have a reduced tendency to endogenously raise questions, we might expect them to approach the task with a question that explicitly represents relatively fewer sequences than the questions raised by control participants. The prediction is that they would ask about fewer possible sequences beyond what is minimally required to ask which urn it is the source. Now, all other things being equal, the narrower the question the greater the chance that a shorter sequence of beads would yield an answer. The fewer questions (or, equivalently in the erotetic theory, questions with fewer alternatives) you ask, the fewer beads you need to see to find an answer. Thus, we would expect un-inquisitive delusional subjects to require fewer draws before completing the task.

A notable virtue of this explanation is that it does not posit any cognitive states or operations that are radically different from that of normal subjects, with the exception of the notion that delusional subjects are less inquisitive.

Contrasting a Bayesian Model of Jumping to Conclusions

The first step for a Bayesian is to give a model for how a rational subject would respond in the 'draws to decision' task. In the classical version of the Beads Task, participants are told the experimenter has selected a particular jar at random, so one's priors for Jar A and B should be equal (P (Jar A) = P (Jar B) = .50). Rational participants will also have prior subjective probabilities concerning the likelihood of a red bead being chosen given that the experimenter picked Jar A or Jar B, which reflect the proportions of beads (*P* (Red|Jar A) = .85; *P* (Red|Jar B) = .15). With these assumptions in place, Bayes' rule allows us to derive the posterior probabilities that a rational subject in the Beads Task ought to have in the experimenter having selected Jar A or Jar B:

After one red bead: P' (Jar A) = .85; *P*' (Jar B) = .15

After two consecutive red beads: P' (Jar A) ~ .97*; P'* (Jar B) ~ .03

After three consecutive red beads: P' (Jar A) ~ .9945*; P'* (Jar B) ~ .0054

As we can see, consecutive draws of a red bead would constitute good evidence that the experimenter has chosen Jar A. Indeed, given what the formal model illustrates, it seems that in the classical version of the Beads Task, delusional participants perform roughly as a Bayesian algorithm would, in the sense that having a .97 degree of confidence that the experimenter has chosen Jar A is plausibly above the threshold for having sufficient evidence to rationally decide that the experimenter has chosen Jar A (cf. Dudley, et. al. 1997; Garety, et. al., 1999; Huq, et. al., 1988). Similarly, it is difficult to see why having a .9945 degree of confidence would be insufficient for knowing that Jar A was chosen by the experimenter. One theoretical motivation for adopting a Bayesian framework is the general prevalence of uncertainty involved in forming beliefs about the external world, which means we should never expect a subject's degree of confidence to ever reach 1. So it therefore looks quite rational to make a decision about which jar has been chosen on the basis of an extremely high degree of confidence like .9945.

In the classical version of the Beads Task, there is no formally defined optimal decision point; no precise probability at which it is optimal for a subject to decide which jar has been selected. So our judgments as to how many draws are required in order to make a rational decision are largely based on our intuitions about the experiment. On one view, it is the *non-psychiatric controls*, rather than delusional or schizophrenic subjects, whose behavior is irrational (cf. Maher and Spitzer, 1993). In recent years, however, theorists have developed a more rigorous version of the Beads Task by adding an incentive structure. The standard experimental design is modified by presenting subjects with both a reward for guessing the correct jar and costs associated with requesting additional draws (cf. Furl and Averbeck, 2011; van der Leer, et. al., 2015; van der Leer and McKay, 2014). Incentivizing the experimental task in this way allows one to use a Bayesian utility maximization algorithm to compute a precise number of draws at which a decision becomes optimal from a decision-theoretic perspective (a number that varies depending on the value of the costs and rewards). Testing delusional subjects with the more rigorous version Beads Task seems to present them with a slightly different computational problem than the classical version (i.e., how to maximize overall utility rather than how to determine which jar the experimenter has chosen), but it seems to have interesting results. ,For example, a recent study that tested both high and low delusion-prone university students showed that they both requested fewer draws on an incentivized Beads Task than would be rationally ideal from a decision-theoretic point of view (van der Leer, et. al., 2015)[6]. Thus, once we add incentives to the Beads Task, it is less clear that delusional subjects manifest more rational behavior. Nevertheless, it remains the case that delusional subjects 'jump to conclusions' in comparison to non-psychiatric controls and thereby exhibit statistically irregular behavior.[7] That is to say, delusion-prone subjects still require fewer draws than non-delusional subjects even on the incentivized version of the Beads Task (van der Leer, et. al., 2015).

We think that a theoretically informed model should strive to explain a range of data in a fairly unified manner. A potentially problematic aspect of the Bayesian account seems to be that the difference between control and delusion participants is, in certain cases like the Capgras delusion, explained by a *failure* of Bayesian rationality, but, at least in the classical version of the Beads Task, it is explained by *particularly strong* Bayesian rationality. Moreover, in the incentivized version of the Beads Task in which both low and high delusion-prone subjects depart from a Bayesian ideal, it is not clear how to give a Bayesian explanation of this difference. Since both groups 'jump to conclusions' more than would prescribed by a normative Bayesian model, in what sense is their reasoning Bayesian? In each case, the relevant Bayesian "explanations" seem more like formal illustrations of the phenomena in question than explanations of a particular cognitive factor that may be implicated in delusional cognition.

Conclusion

In this paper, we have proposed the erotetic theory as a novel account of a cognitive factor in a multi-factor model of delusion. We argued that our central hypothesis, that the cognitive processes implicated in reasoning endogenously generate fewer questions in delusional and schizophrenic subjects, offers simple, unified explanation of a range of empirical data. Moreover, since the cognitive mechanisms we appeal to are independently motivated to make sense of ordinary patterns of reasoning, we have a natural way of explaining the behavior of non-psychiatric subjects. We also avoid having to posit a variety of distinct epistemic states or dispositions that are intrinsically irrational in order to fit our model to the variety of empirical data. We therefore conclude that the erotetic theory presented here offers a plausible framework for explaining delusional cognition.

References

Adams, R. A., et al. (2013). "The computational anatomy of psychosis." Frontiers in Psychiatry **4**.

Alexander, M., et al. (1979). "Capgras syndrome A reduplicative phenomenon." Neurology **29**(3): 334-334.

Ayton, P. and I. Fischer (2004). "The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?" Memory & cognition **32**(8): 1369-1378.

Bell, V., et al. (2006). "Explaining delusions: a cognitive perspective." Trends Cogn Sci **10**(5): 219-226.

Blakemore, S.-J., et al. (2002). "Abnormalities in the awareness of action." Trends Cogn Sci **6**(6): 237-242.

Breen, N., et al. (2000). "Towards an understanding of delusions of misidentification: Four case studies." Mind & language **15**(1): 74-110.

Brighetti, G., et al. (2007). "“Far from the heart far from the eye”: evidence from the Capgras delusion." Cogn Neuropsychiatry **12**(3): 189-197.

Chan, R. C., et al. (2006). "Specific executive dysfunction in patients with first-episode medication-naïve schizophrenia." Schizophrenia Research **82**(1): 51-64.

Chen, E., et al. (1994). "Semantic memory is both impaired and anomalous in schizophrenia." Psychological Medicine **24**(01): 193-202.

Clark, A. (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science." Behavioral and Brain Sciences **36**(3): 181-204.

Coltheart, M. (2005). "Delusional belief." Australian Journal of Psychology **57**(2): 72-76.

Coltheart, M., et al. (2011). "Delusional belief." Annual review of psychology **62**: 271-298.

Coltheart, M., et al. (2010). "Abductive inference and delusional belief." Cogn Neuropsychiatry **15**(1-3): 261-287.

Davies, M., et al. (2001). "Monothematic delusions: Towards a two-factor account." Philosophy, Psychiatry, & Psychology **8**(2): 133-158.

Davies, M. and A. Egan (2013). "Delusion: cognitive approaches–Bayesian inference and compartmentalization." The Oxford Handbook of Philosophy and Psychiatry: 688-727.

Dudley, R., et al. (1997). "Normal and abnormal reasoning in people with delusions." British journal of clinical psychology **36**(2): 243-258.

Edgington, D. (1995). "On conditionals." Mind **104**(414): 235-329.

Ellis, H. D., et al. (2000). "Automatic without autonomic responses to familiar faces: Differential components of covert face recognition in a case of Capgras delusion." Cogn Neuropsychiatry **5**(4): 255-269.

Ellis, H. D. and A. W. Young (1990). "Accounting for delusional misidentifications." British Journal of Psychiatry **157**(2): 239-248.

Ellis, H. D., et al. (1997). "Reduced autonomic responses to faces in Capgras delusion." Proceedings of the Royal Society of London. Series B: Biological Sciences **264**(1384): 1085-1092.

Falk, R. A. K., Clifford (1991). "Making Sense of Randomness." Psychological Review **104**(2): 301-318.

Fine, C., Craigie, J., and Gold, I. (2007). "Hopping, skipping or jumping ot conclucions? Clarifying the role of the JTC bias in delusions." Cogn Neuropsychiatry **12**(1): 46-77.

Garety, P. A. and D. Freeman (1999). "Cognitive approaches to delusions: a critical review of theories and evidence." British journal of clinical psychology **38**(2): 113-154.

Garety, P. A., et al. (2005). "Reasoning, emotions, and delusional conviction in psychosis." Journal of abnormal psychology **114**(3): 373.

Gilovich, T., et al. (1985). "The hot hand in basketball: On the misperception of random sequences." Cognitive psychology **17**(3): 295-314.

Hahn, U. and P. A. Warren (2009). "Perceptions of randomness: Why three heads are better than four." Psychological Review **116**(2): 454.

Henik, A. and R. Salo (2004). "Schizophrenia and the stroop effect." Behavioral and cognitive neuroscience reviews **3**(1): 42-59.

Hohwy, J. (2013). The Predictive Mind, Oxford University Press.

Huq, S., et al. (1988). "Probabilistic judgements in deluded and non-deluded subjects." The Quarterly Journal of Experimental Psychology **40**(4): 801-812.

Johnson-Laird, P. N. (2008). Mental models and deductive reasoning. Reasoning: Studies in human inference and its foundations. L. a. A. Rips, J. Cambridge, Cambridge University Press.

Koralus, P. and S. Mascarenhas (2013). "The erotetic theory of reasoning: Bridges between formal semantics and the psychology of deductive inference." Philosophical Perspectives **27**(1): 312-365.

Landa, R. and E. Garrett-Mayer (2006). "Development in infants with autism spectrum disorders: a prospective study." Journal of Child Psychology and Psychiatry **47**(6): 629-638.

Langdon, R., et al. (2010). "Reasoning anomalies associated with delusions in schizophrenia." Schizophrenia bulletin **36**(2): 321-330.

Lopes, L. L. and G. C. Oden (1987). "Distinguishing between random and nonrandom events." Journal of Experimental Psychology: Learning, Memory, and Cognition **13**(3): 392.

Maher, B. A. (1974). "Delusional thinking and perceptual disorder." Journal of individual psychology.

Maher, B. A. and M. Spitzer (1993). "Thought disorders and language behavior in schizophrenia." G. Blanken, J. Dittman et alt.(Eds.), Linguistic Disorders and Pathologies. An International Handbook: 522-533.

McKay, R. (2012). "Delusional inference." Mind & language **27**(3): 330-355.

Mellet, E., et al. (2006). "When a schizophrenic deficit becomes a reasoning advantage." Schizophrenia Research **84**(2): 359-364.

Menon, M., et al. (2006). "Probabilistic reasoning in schizophrenia: a comparison of the performance of deluded and nondeluded schizophrenic patients and exploration of possible cognitive underpinnings." Cogn Neuropsychiatry **11**(6): 521-536.

Mintz, A. R., et al. (2003). "Insight in schizophrenia: a meta-analysis." Schizophrenia Research **61**(1): 75-88.

Moritz, S. and T. S. Woodward (2005). "Jumping to conclusions in delusional and non-delusional schizophrenic patients." British journal of clinical psychology **44**(2): 193-207.

Oaksford, M. and N. Chater (2007). Bayesian rationality the probabilistic approach to human reasoning, Oxford University Press.

Orem, D. M. and J. S. Bedwell (2010). "A preliminary investigation on the relationship between color-word Stroop task performance and delusion-proneness in nonpsychiatric adults." Psychiatry research **175**(1): 27-32.

Parrott, M. (Forthcoming). "Bayesian Models, Delusional Beliefs, and Epistemic Possibilities." British journa for the philosophy of science.

Startup, M. (1997). "Awareness of own and others' schizophrenic illness." Schizophrenia Research **26**(2-3): 203-211.

Stone, T. and A. W. Young (1997). "Delusions and brain injury: The philosophy and psychology of belief." Mind & language **12**(3-4): 327-364.

Tranel, D., et al. (1995). "Double dissociation between overt and covert face recognition." Journal of Cognitive Neuroscience **7**(4): 425-432.

Walsh, C. R. and P. N. Johnson-Laird (2004). "Co-reference and reasoning." Memory & cognition **32**(1): 96-106.

Young, A. W., et al. (1993). "Face perception after brain injury: Selective impairments affecting identity and expression." Brain **116**(4): 941-959.

Young, A. W., et al. (1992). "Cotard delusion after brain injury." Psychological Medicine **22**(03): 799-804.

Appendix I: Derivations for conditional reasoning with and without false antecedents.

Control case (negation in consequent) for normal participants

If there is a red square on the left then there is no yellow circle on the right.

{red_square_left & ~yellow_circle_right, ~red_square_left}

Since we are, as far as we can, using objects that some alternative in the model is committed to, we should first try combinations involving red_square_left. There is no other particular object that any alternative is committed to. But if we use red_square_left as our first supposition, the only way to generate a contradiction is to next suppose yellow_circle_right.

{red_square_left & ~yellow_circle_right, ~red_square_left}[{red_square_left}]SUP

={red_square_left&~yellow_circle_right}[{yellow_circle_right}]SF

=contradiction.

Alternatively, you suppose {red_square_left & yellow_circle_right}. {red_square_left & ~yellow_circle_right, ~red_square_left}[{red_square_left&yellow_circle_right}]SUP =contradiction

So this predicts that the vast majority of participants should choose red_square_left and yellow_circle_right. As was observed.

<u>Target case (negation in antecedent) for normal participants</u>

If there is not a red square on the left then there is a yellow circle on the right.

{~red_square_left & yellow_circle_right, ~~red_square_left}

Again, we are making suppositions about the left and right objects, maximizing the use of objects that some alternative is committed to. This gives us red_square_left and yellow_circle_right. If we suppose yellow_circle_right and suppose further red_square_left, we get a contradiction.

{~red_square_left & yellow_circle_right, ~~red_square_left}[{yellow_circle_right}]SUP

={~red_square_left&yellow_circle_right}

[{red_square_left}]SF = contradiction.

{~red_square_left & yellow_circle_right,

~~red_square_left}[{yellow_circle_right&red_square_left}]SUP

=contradiction (but only because ~~red is not the same molecule as red, even though you could reason from one to the other).

So the prediction is that most normal participants should select red_square_left and yellow_circle_right, which is what was observed.

We suggest that schizophrenic participants have a deficit in endogenously calling up reasoning operations that are inquisitive, i.e. that tend to increase the number of alternatives without an external prompt. This would cover the use of supposition and of inquiry. This means that we would predict that schizophrenic patients have to rely on a different reasoning strategy in this case. We suggest that the most straightforward strategy would be to simply take the reasoning operations that are most directly related to the explicit instructions, namely to negate the conditional and then find a pair of objects that is consistent with the resulting model.

Control case (negation in consequent) for schizophrenic participants

Neg{red_square_left & ~yellow_circle_right, ~red_square_left} =

={~red_square_left, ~~yellow_circle_right} x {~~red_square_left}

={~red_square_left & ~~red_square_left, ~~yellow_circle_right & ~~red_square_left}

[]F = {yellow_circle_right & red_square_left}

The only set of objets that is compatible with this is obviously yellow_circle_right & red_square_left, so the prediction is that schizophrenic patients should respond just like the normal participants in this case, as was observed.

Target case (negation in antecedent) for schizophrenic participants

Neg{~red_square_left & yellow_circle_right, ~~red_square_left}=

={~~red_square_left, ~yellow_circle_right} x {~~~red_square_left}=

={~~red_square_left & ~~~red_square_left, ~yellow_circle_right&~~~red_square_left}

[]F={~yellow_circle_right&~red_square_left}

This result then leaves participants with the task of finding some set of objects that involves neither a yellow_circle_right nor a red_square_left. We suspect that this is not an insurmountable problem and thus predict that many participants should be able to do so, which would yield a correct answer.

---

[1] One potential challenge is to question the extent to which an explanation of ordinary human reasoning is really 'Bayesian', if it always appeals to some kind of deviation from an ideal model. Although this is an important question, we shall set it aside for the remainder of this essay (see Parrott (forthcoming) for further discussion).

[2] Our discussion here will focus on the Capgras delusion. Since Ellis and Young's influential suggestion, this delusion has been at the center of much work in cognitive neuropsychology and neuropsychiatry. However, most researchers tend to think that an adequate model of the Capgras delusion would be able to be extended to other monothematic delusions (cf. Coltheart 2007; Davies and Egan 2013).

[3] It is worth noting that this lack of responsiveness does not itself constitute the Capgras subject's anomalous experience. People are not consciously aware of their autonomic nervous system (Coltheart, 2005). Nevertheless, a disturbance in the autonomic nervous system could generate an irregular experience, perhaps an experience of something being different or wrong in some way. 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. One plausible hypothesis is that disturbance in the autonomic nervous system causes aberrant prediction error signaling, which in turn alters the character of a subject's conscious experience (cf. Adams et. al. 2013; Clark, 2013; Howhy, 2013)

[4] With respect to conditionals, there is some evidence that schizophrenic subjects are less susceptible to fallacious syllogistic reasoning due to belivability bias (cf. Owen, et. al. 2007);

although other experimental work indicates there may be no significant difference between schizophrenic subjects and non-psychiatric controls (Kemp, et. al., 1997). Although a discussion of these results is outside the scope of the current essay, we would like to note that the erotetic theory provides a model of believability bias (Koralus and Mascarhenas, 2013).

[5] This is not the only way to formally model their behavior. The following would also work: P (yellow circle | ~red square) < P (yellow circle ^ red square).

[6] In line with several experimental paradigms, the illustration of the classical Beads Task that we give in the previous paragraph assumes that subjects are presented with sequences of at least two if not three consecutive beads of the same colour cf. Fine, et. al., 2007; Garety and Freeman, 1999; Huq, et. al. 1988). It is worth noting that the sequences experimentally tested by van der Leer and colleagues do not begin with two consecutive colours.

[7] This may have something to do with delay discounting (cf. Heerey, et. al., 2007).