

The Erotetic Theory of Delusional Thinking

Matt Parrott¹ and Philipp Koralus²

1/7/15

Abstract: In this paper, we argue for a novel account of one cognitive factor implicated in delusional cognition. Specifically, we propose that a less inquisitive version of a question/answer process that has been independently proposed to account for ordinary reasoning underlies delusional thinking. The erotetic theory of delusion holds that the central cognitive factor in delusion is impaired endogenous question raising. We also argue that the erotetic theory has considerable advantages over alternative models of the cognitive factor in delusion, specifically Bayesian models. More specifically, we argue that the erotetic theory offers a superior explanation of the onset of the Capgras delusion and of experimentally observed patterns of reasoning exhibited by delusional and schizophrenic subjects.

(Abstract word count: 114)

(Main text word count: 6449)

1. 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

¹ King's College London

² University of Oxford, Laboratory for the Philosophy and Psychology of Rationality and Decision (LPPRD)

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).

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 such 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). Despite the widespread popularity of this Bayesian approach in the cognitive sciences, we will argue that it faces serious challenges.

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.

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.

2. 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 that 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 can allow 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. What separates reliably valid reasoning from fallacy prone reasoning is the extent to which we raise enough questions as we reason.

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 this means is that delusional and schizophrenic subjects raise the same sorts of default questions as typical individuals, 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. Delusional and schizophrenic subjects are simply less inquisitive. The result would be that delusional thinking is just a more extreme manifestation of a general human tendency to answer our questions quickly, 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. 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 the Bayesian would, that if neural damage causes a delusion, it has to bring about a new, intrinsically mistaken epistemic attitude, such as, 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 raising questions is an inhibitory mechanism that ordinarily moderates a general human tendency to jump to conclusions 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 fact manifest *improved* reasoning performance, relative to non-psychiatric populations, in special cases. 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.

3. Capgras delusion

The Capgras delusion is a condition in which someone believes that an imposter has replaced one of her close friends or relatives.³ 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.⁴

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 implicated in the etiology of the Capgras delusion.

3.1. Capgras delusion in the erotetic theory

³ 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).

⁴ 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)

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 delusory misidentification, 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 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 can be represented as follows, as a set of alternative possible answers (following Koralus and Mascarenhas 2013):

- (1) {C-feature&wife&short&..., doctor&brown hair&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. 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 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 patients 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 in patients, 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 Capgras delusion patients 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 patients through this reasoning does not seem to have lasting effects (Colheart, et. al., 2007; Colheart, 2007). This pattern of delusional belief maintenance is not a surprise if, as above, the impostor belief keeps re-generating from a question that keeps arising by default whenever a person walks into the room. Lasting avoidance of the delusional conclusion would have to involve the patient him or herself raising the relevant follow-up questions whenever the patients is momentarily mislead by jumping to conclusions on a default question. However, on the erotetic theory of delusion, it is precisely endogenous question-raising that is hypothesized to be deficient.

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.

3.2. Contrasting Bayesian explanations

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 given 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 a “spouse” hypothesis. So either the subject's prior probability in *spouse* is comparatively low or the degree of confidence she has in the likelihood of *E* given *imposter* is comparatively high.

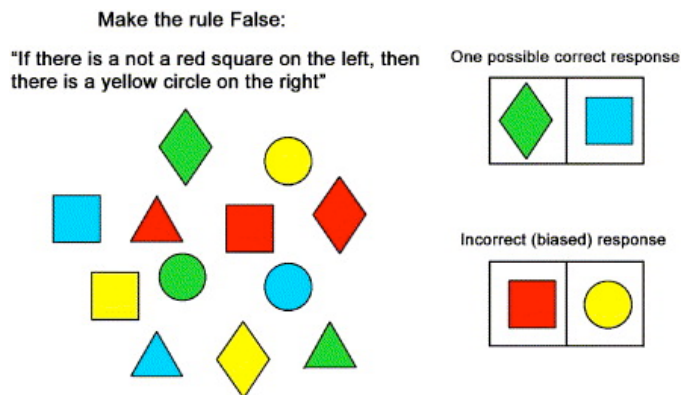
The central question for the Bayesian approach is therefore what values should be assigned to the prior probabilities within our model (cf. Parrott, forthcoming). Some models set assume that delusional subject assign a comparatively higher prior probability to *Imposter* (cf. Coltheart, et. al. 2010) whereas others claims the subject’s prior is very low and discounted in her subsequent reasoning (McKay, 2012).

Now, adopting priors on a particular proposition that are absurdly low in light of background knowledge or discounting one’s prior probabilities on that proposition entirely are both irrational. If we have to build those irrationalities concerning a particular proposition into a Bayesian model for the sole purpose of making sense of the irrationality of delusional thinking, it looks like the Bayesian model is merely kicking the explanatory problem upstairs rather than solving it. In addition, delusional patients often explicitly acknowledge that the impostor hypothesis seems rather improbable (Alexander, et. al., 1979; Stone and Young, 1997; cf. Startup, 1997). By contrast, 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 an endogenous tendency to raise further questions.

4. Reasoning with conditionals

In certain cases, a delusion-related diagnosis of schizophrenia can in fact improve performance on certain reasoning tasks. 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.

4.1. The erotetic theory and reasoning with 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 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 enough 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 supposition, the erotetic theory of delusion suggests that schizophrenic subjects will not spontaneously adopt a suppositional reasoning strategy. If you have deficits 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 superficially 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. 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 one 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 proposal for how normal 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 reasoners do better in this case. In addition, it explains the particular mistake made by control participants.

The erotetic theory is 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.

4.2. Contrasting a Bayesian approach

As we have seen, an adequate model 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 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.

5. 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).

5.1. 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, R, R, B, B, B, B, B, etc.?” The minimal “sequence” would be just a single draw, so the minimal question of this sort 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 general tendency to be less inquisitive, we might expect them to approach the task with a question that explicitly represents relatively fewer sequences than the relevant questions raised by control participants. 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. As it turns out, having a less fleshed out version of the in-principle irrational strategy of looking for patterns representative of the different jars happens to yield performance that is closer to an ideal rational response. In other words, the suggestion is that the delusional subjects get a closer-to-right answer for the wrong reasons.

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.

5.2. Contrasting a Bayesian model

The first step for a Bayesian is to give an idealized model for how a perfectly rational subject would respond in the 'draws to decision' task. In 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. 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, reflecting the proportions of beads. 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.

Consecutive draws of a red bead constitute very good evidence that the experimenter has chosen Jar A. Indeed, delusional participants in the Beads Task perform roughly as a Bayesian algorithm would. This suggests that it is the *non-psychiatric controls*, rather than delusional or schizophrenic subjects, whose behavior is irrational (cf. Maher and Spitzer, 1993). From a Bayesian perspective, they seem to be too conservative in their assessment of the evidence. Nevertheless, despite being more rational from a Bayesian point of view, the jumping to conclusions behavior exhibited by delusional participants is statistically irregular.

A problematic aspect of the Bayesian account seems to be that the difference between control and delusion participants is sometimes explained by a *failure of Bayesian rationality*, as in the Capgras case, it is, in this case explained by *particularly strong Bayesian rationality*. This makes the relevant Bayesian "explanations" seem more like formal illustrations than explanations of a particular cognitive factor in delusion. We believe that the erotetic theory is more promising in this regard.

6. Conclusion

In this paper, we have proposed the erotetic theory as a novel account of the cognitive factor in a multi-factor model of delusion. We argued that our 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 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 more plausible framework for explaining delusional cognition than the Bayesian alternative.

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 conclusions? 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.

Korolus, 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 al.(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 journal 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.

$\{\text{red_square_left} \ \& \ \sim\text{yellow_circle_right}, \sim\text{red_square_left}\}[\{\text{red_square_left}\}]\text{SUP}$

$=\{\text{red_square_left} \ \& \ \sim\text{yellow_circle_right}\}[\{\text{yellow_circle_right}\}]\text{SF}$

=contradiction.

Alternatively, you suppose $\{\text{red_square_left} \ \& \ \text{yellow_circle_right}\}$.

$\{\text{red_square_left} \ \& \ \sim\text{yellow_circle_right},$

$\sim\text{red_square_left}\}[\{\text{red_square_left} \ \& \ \text{yellow_circle_right}\}]\text{SUP}$

=contradiction

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

Target case (negation in antecedent) for normal participants

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

$\{\sim\text{red_square_left} \ \& \ \text{yellow_circle_right}, \sim\sim\text{red_square_left}\}$

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.

$$\{\sim\text{red_square_left} \& \text{yellow_circle_right}, \sim\sim\text{red_square_left}\}[\{\text{yellow_circle_right}\}]\text{SUP}$$

$$=\{\sim\text{red_square_left}\&\text{yellow_circle_right}\}$$

$$[\{\text{red_square_left}\}]\text{SF} = \text{contradiction.}$$

$\{\sim\text{red_square_left} \& \text{yellow_circle_right},$
 $\sim\sim\text{red_square_left}\}[\{\text{yellow_circle_right}\&\text{red_square_left}\}]\text{SUP}$
 $=\text{contradiction}$ (but only because $\sim\sim\text{red}$ 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

$$\begin{aligned} \text{Neg}\{\text{red_square_left} \& \sim\text{yellow_circle_right}, \sim\text{red_square_left}\} &= \\ =\{\sim\text{red_square_left}, \sim\sim\text{yellow_circle_right}\} \times \{\sim\sim\text{red_square_left}\} &= \\ =\{\sim\text{red_square_left} \& \sim\sim\text{red_square_left}, \sim\sim\text{yellow_circle_right} \& \sim\sim\text{red_square_left}\} & \\ []\text{F} = \{\text{yellow_circle_right} \& \text{red_square_left}\} & \end{aligned}$$

The only set of objects 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

$$\begin{aligned} &\text{Neg}\{\sim\text{red_square_left} \& \text{yellow_circle_right}, \sim\sim\text{red_square_left}\} = \\ &= \{\sim\sim\text{red_square_left}, \sim\text{yellow_circle_right}\} \times \{\sim\sim\sim\text{red_square_left}\} = \\ &= \{\sim\sim\text{red_square_left} \& \sim\sim\sim\text{red_square_left}, \sim\text{yellow_circle_right} \& \sim\sim\sim\text{red_square_left}\} \\ &[]F = \{\sim\text{yellow_circle_right} \& \sim\text{red_square_left}\} \end{aligned}$$

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.