A Bayesian perspective on delusions: suggestions for modifying two reasoning tasks

Abstract

Background and Objectives. There are a range of mechanistic explanations on the formation and maintenance of delusions. Within the Bayesian brain hypothesis, particularly within the framework of predictive coding models, delusions are seen as an aberrant inference process characterized by either a failure in sensory attenuation or an aberrant weighting of prior experience. Testing of these Bayesian decision theories requires measuring of both the patients' confidence in their beliefs and the confidence they assign new, incoming information. In the Bayesian framework we apply here, the former is referred to as the prior while the latter is usually called the data or likelihood.

Methods and results. This narrative review will commence by giving an introduction to the basic concept underlying the Bayesian decision theory approach to delusion. A consequence of crucial importance of this sketch is that it provides a measure for the persistence of a belief. Experimental tasks measuring these parameters are presented. Further, a modification of two standard reasoning tasks, the beads task and the evidence integration task, is proposed that permits testing the parameters from Bayesian decision theory.

Limitations. Patients differ from controls by the distress the delusions causes to them. The Bayesian Decision theory framework has no explicit parameter for distress.

Conclusions. A more detailed reporting of differences between patients with delusions is warranted.

Introduction

Already in the 19th century, Hermann von Helmholtz described perception as an unconscious inference based on previous knowledge and incoming sensory data (1924). Seeing is believing and all seeing is influenced by what one expects to see. Indeed, one can “want to see” which means that the belief is weighted stronger than the actual information received from sensation: The perceived sensory input can be discounted to fulfil one’s prediction. The influence of what one wants to see is most obvious in the somewhat different case of viewing ambiguous figures, such as the Necker cube, or the duck rabbit. Even when people entirely fail to notice that there is more than one possible interpretation of sensory data, they can deliberately switch to the alternative interpretation once told what it is. In less ambiguous situations, where one interpretation of the evidence is more strongly favoured, it takes more to go against the evidence. More of what, though? A stronger expectation that one interpretation is true (Schwartenbeck et al., 2015) or a problem with the inference process (Hemsley and Garety, 1986)? In principle, both options could cause aberrant perception and beliefs. Given that beliefs at one level are evaluated at a higher level, it is not easy to disentangle the two possibilities. What looks like aberrant inferences might be caused by too strong or too weak higher order beliefs (Mathys et al., 2011). Indeed, with respect to explaining delusions an aberrant (over- or under-) weighing of belief has been postulated to be an underlying mechanism (Corlett, Frith, and Fletcher, 2009, Adams et al., 2013, Friston et al,. 2015; Teufel et al., 2015). These Bayesian decision theory accounts are hierarchical. Simply said, there is a Bayesian integration at the perceptual level, as well as there is a controlling or plausibility check at a higher cognitive level (Coltheart 2007). In the Bayesian decision theory terminology: there is an uncertainty about the precision of a belief. This uncertainty about the

precision of a belief is a measure of how certain the predictions from this belief are. For example, someone picking mushrooms needs to do more than decide whether the mushroom seen now is a better match to the memory of an edible or the memory of a poisonous mushroom. It is also necessary to know how variable the appearance of both species of mushroom is, and how precisely one remembers. And although there is an objective precision that might be measured by experiment, the mushroom picker must make a subjective estimate regarding the precision, and may be uncertain regarding that precision.

Further, at any time, there is not just one belief. As in hypothesis testing there are alternative options to believe in, each with its "strength". A person's model of the world contains many beliefs. The aim is to reduce uncertainties, find the appropriate model and hence make better predictions (Friston, 2005). Accordingly, there is ongoing learning. And how fast one learns depends on many factors. The difference in learning rate (attention, interests) leads to differences in the kind of beliefs (belief formation) one has as well as differences in the persistence of well-functioning beliefs (conviction stage). One may cling to some believes more than to others as some beliefs apply across various environments (are universal) whereas other beliefs are part of unstable environments (e.g. friendships). These two basic stages of belief formation and belief conviction also apply to delusion (see also Moritz et al., this issue).

In the next part, this article will illustrate belief formation and maintenance on a fictional example. This example shows that it is not the inference process itself that is aberrant. Rather, it appears to be a weak reflective, metacognitive assessment of the reliability of a belief that prevents the calibration of false beliefs and belief flexibility (Buck et al., 2012; Coltheart, 2007; Moritz and Woodward, 2006). That is, patients with delusions show epistemic irrationality, but intact instrumental rationality, i.e. they act according to their beliefs (Barch et al., 2013). Thereafter, I will describe modifications to two classical paradigms: the beads task and the evidence integration task. Knowing which parameter is impacted by delusions may provide individually tailored metacognitive therapy but also provide objective measures of treatment outcomes. It will allow measuring when all parameters are "normal".

Delusions as aberrant statistical inference

An advance from a descriptive towards a mechanistic understanding of delusions is crucial to advance understanding and treatment of this condition. The "Bayesian brain" framework provides such a mechanistic approach. In this view, all information processing in the brain is seen as an integration of previous knowledge with incoming new information. Continuously, knowledge/belief, is accumulated over various timescales: Prior experience and beliefs can be evolutionarily acquired (e.g. light comes from above), learned within the lifetime of an individual (e.g., chocolate is tasty) or fluctuate quickly on the order of seconds (e.g., the bird changed flight direction). Stereotypes or religions are examples of strongly learned beliefs: They can be held with great precision and be strongly robust against conflicting information. Mathematically, beliefs can conveniently be modelled as probability distributions over the space of possible events, allowing the study of the rather abstract concept of a "belief" in concrete terms. When representing beliefs as probability distributions the most likely value is its expectation[1]. Information about uncertainty of this parameter is contained in a dispersion parameter indexing the width of the distribution, its variance (the inverse of the variance is

[1] In the special case of a normal distribution the expectation is the mode, median and mean.

called precision). The stronger a belief, the more precise or narrow its corresponding distribution. A critical assumption of (Bayesian) reasoning is that beliefs are updated after perceiving new data. Further, probability distributions over the likelihood of different beliefs can be specified (beliefs about beliefs), expressing how reliable or appropriate one belief is compared to another. It follows that each belief also has a precision, and also a reliability in the precision of the belief. That is, the reliability is here how certain the agent is regarding the distribution of the belief, its shape, mean and variance. This reliability is a measure of how resistant a belief is to change. Reliabilities are thought to be set by sufficient experience, i.e. optimal agents become correctly calibrated (Huys et al., 2015, Pfuhl et al., 2011).

Data, in the form of novel observations, impact the internal representation of the world (the "model") by changing the associated probability distribution of the parameters by way of their "likelihood" (the probability of the data given the model). Depending on the (perceived) precision of this data as well as the current estimate of reliability, the internal update of a belief will be strong (in case of highly trustworthy or precise data) or weak. Any deviation between a predicted outcome based on one’s belief, i.e. how children will react to a cyclist, and data, i.e. how did the children react, is the prediction error. Any prediction error leads to a re-evaluation of the reliability and precision of the belief.

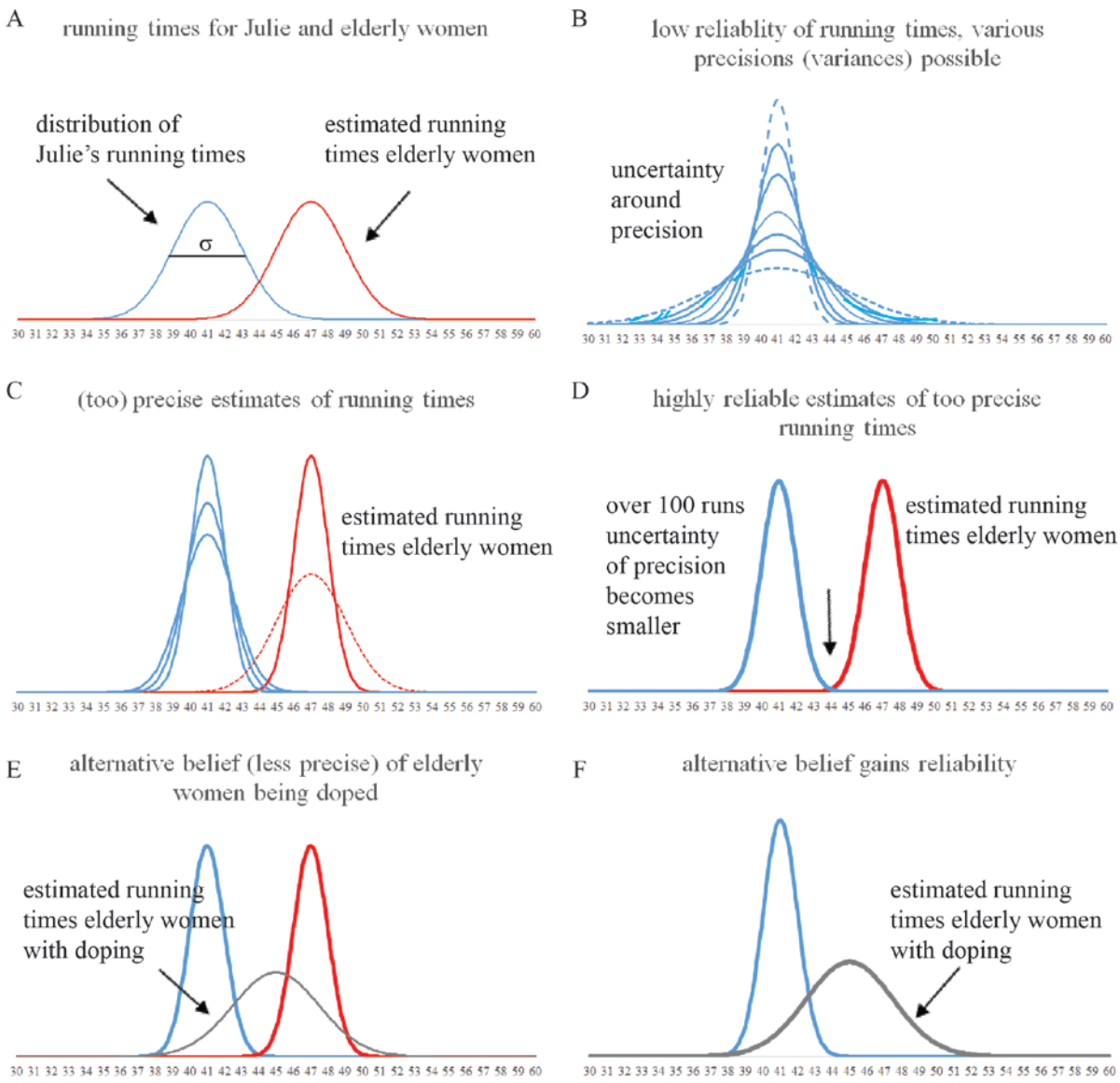

Figure 1 Schematic illustrating the precision of a belief and the reliability of the precision of a belief based on the running example (see text for details). The graphs show normal probability distributions that represent prior beliefs. A) Shown is the distribution of Julie's running times (blue) and her belief in the distribution of the running times of elderly women (red). These distributions have an expectation (41 min and 47 min, respectively), and a precision (inverse of variance) associated with it. The more precise the smaller the width of the distribution. B) Here, the reliability of the belief is the spread in the possible precisions of a belief (there can be also uncertainty around the mean, not illustrated). The larger the uncertainty about the range of possible running times the lower the reliability. Experience increases the reliability (better estimation) but not the precision. C) However, in this example training may increase precision or the stability of one's performance that might be extrapolated to the belief about running times of elderly women. D) Because there is real variation in the world, increased experience cannot just shrink the variance of estimated running times to 0. What experience can do is allow a more precise estimate of the variance of running times. Here illustrated as a very precise and very reliable estimate. This leads to nearly no overlap between Julie's expected running times and those of elderly women (arrow). E) A too precise belief results in large prediction errors, or discrepancies between predicted outcome and perceived outcome. Alternative beliefs result in lower discrepancies. F) Subsequent updating may strengthen the alternative belief. Horizontal axis: running times for 10 km in min.

How are delusions explained in terms of Bayesian inference?

The discussion of Bayesian inference so far has centered on inference processes as assumed in healthy individuals. In the following scenario, I will illustrate how such Bayesian reasoning

can go wrong and result in delusional beliefs. Consider the following belief Julie may hold about herself: Julie believes that all elderly women are slower runners than she is. From some local races her running time for 10 km is in the range of 39 to 43 min (Fig 1A). Now she takes part at a race with international starters. Obviously, there is some uncertainty associated with predicting her own speed, and perhaps some former Olympic athletes' performance declines less quickly with age than she thinks. Since it is an international race, she might be faster than usual (finding a pacemaker) or slower than usual (hindrance from other runners). This would be accommodated in the reliabilities (Fig 1B).

Even if Julie expects her mean running times in international races will be the same as in local races, and if she believes that the variance of running times will remain unchanged, in her first international race she may be less certain that these beliefs accurately reflect reality. She would consider her beliefs less reliable. The same applies to her beliefs regarding the running times of elderly women. For example, the variance in the running times of elderly women in international races might be greater if the age range is greater and if the performance of former Olympic athletes has declined less than Julie believes. Or the variance might be less if the organisers demand a minimum level of performance.

Assume that Julie finishes her run within 41 minutes, but she is beaten by an elderly woman, which contradicts at least one of her beliefs. Her own running time is as expected, so she has no reason to modify that belief. How she reconciles the new experience with her belief regarding the running times of elderly women depends both on the precision and on the reliability of this belief. If the new experience is only a moderate outlier, i.e. Julie believes that elderly women's running times in international races vary widely (low precision), Julie may recalibrate the variance of elderly women's expected running times (and also the mean, for simplicity not drawn here). If the new experience is an extreme outlier, meaning that Julie believes the expected variance in elderly women's running times is low, that she knows those times with high precision, she may still decide to recalibrate the variance and mean of those running time if a second condition holds: that Julie thinks her belief regarding elderly women's running times in international races is unreliable, that a wide range of variances and means are possible, due to her acknowledged ignorance of the factors mentioned above.

However, if the experience of being beaten by an elderly woman is an extreme outlier, *and* Julie believes that her belief regarding the possibly variance and mean of elderly women's running times is reliable, that she *knows* that variance and mean, then recalibrating that variance and mean is less reasonable than adopting an alternative belief that represents a different causal relationship. She may interpret some remark about doping to mean that doping occurred in this race, and an obvious candidate would be the elderly woman who did so unexpectedly well by beating Julie. In future races Julie may perceive some of the women that have beaten her as elderly and doped. Since her initial belief of running times for elderly women was very precise and reliable, this (perception of her) new experience further strengthens the belief in doping. For some period, Julie may hold both beliefs (Moritz et al., 2016b; Risen, 2016), being able to beat all elderly women who don't cheat and elderly women who do dope beating her. However, repeating the experience of being beaten by elderly women may convince Julie that doping is common. That is, the doping belief explains the data better than the elderly women running not faster than 44 min belief (Fig 1D). That is, an elderly women running 41 min is better predicted by the doping belief than the previous belief. Subsequently, the reliability (not necessarily the precision!) of the doping belief

increases. This belief of elderly women running between 38 to 52 min fast, may cause fewer prediction errors and hence is reinforced with every new experience. Any contrary evidence (friends telling about negative doping tests) might be explained away by "raising the stakes", for example that the doping was organized by the state. Such an exaggeration may lead to a conspiracist belief that goes beyond a race. The belief in the state-supported doping may go up the hierarchy and affect related beliefs, such as how trustworthy other claims of state institutions are (Lewandowsky, Oberauer & Gignac, 2013).

This example illustrates that the same information is treated differently, depending on the prior belief. It also highlights the importance of knowing the reliability and precision of a belief. Only if those are known can we infer how current situational information is being processed. In a selection stage, one may go through a stage of indifference or "hopping back and forth" between two alternative explanations. A process familiar to scientists when deciding which model explains a phenomenon. Subsequently, the model or belief resulting in the lower prediction error will be promoted and its reliability can increase. Indeed, there are two components here, one is the uncertainty about a belief (hierarchical Bayes), the other is the selection of a belief (model selection). Said differently, we can distinguish between which model and how well a model predicts data. "Which model" might be related to belief formation, whereas "how well" to belief conviction. If one does not have an alternative model, the existing model may describe the world (e.g. ether theory of radio waves).

Predictive coding theories of delusion do not see the problem in the prediction of the sensory input per se. The fault appears to be an imbalance between ascribed prior beliefs and incoming evidence (Adams et al., 2015, Friston, 2005). As a consequence, given that beliefs are learned and influence perception, one has to measure not only the belief per se but also the reliability (conviction/inflexibility to tolerate alternatives) of a belief and its precision (how narrowly defined is the belief, what counts as outlier for this belief). In other words, the precision of a belief causes prediction errors, the reliability determines how one acts on the prediction error. Here we can also see the interaction of precision and reliability. If the precision is low, prediction errors will be small and nearly every observation can be accommodated by the imprecise belief. There is no urge to update the precision or reliability of this belief. If the precision is high, rare events cause large prediction errors. If the belief has a low reliability its precision[2] may be recalibrated. Alternatively, if the belief has a high reliability, the rare event may be "explained away" by assigning it to an alternative belief. The original belief is protected against modification by interpreting the outlier as a consequence of a different cause. In this framework, delusions may be characterized as occurring as a consequence of an overly flexible or liberal acceptance of alternative beliefs which acquire a high reliability and become resistant to further changes. Note, that the reliability is a judgment about the belief. This process leads to epistemic irrationality in that the reliability of the belief is not verified against other beliefs.

Assessing reliability and precision of prior beliefs

The model above illustrates that one needs to assess both the precision of a belief as well as its reliability. Direct experimental tests that measure both the precision and the reliability are rare. Recently, experiments have been designed to measure some of these parameters. Schmack et al. (2013) exposed subjects to an environment of ambiguous objects, where one

[2] and the mean

particular type of object was more common than another type. A learning phase established an expectation or prior belief about the relative availability of the two types of objects (rotating either left or right). They found that for delusion-prone subjects, this induced prior was more stable when the environment has changed to an equal distribution of the relative frequency of the two objects. To some extent, this paradigm therefore measured the "belief in a belief".

Teufel et al. (2015) presented complex images to their subjects. In a later stage, fragments of these images were shown and had to be recognized by the subjects. Delusion-prone subjects were better in recognizing previously seen fragments compared to less delusion-prone subjects. These data were interpreted as supporting the hypothesis that a more precise, and therefore more persistent, prior was associated with delusions. In both studies, the response was binary, either left/right (Schmack et al., 2013) or yes/no (Teufel et al., 2015). Therefore, the precision could not directly be measured. To improve on this situation, Pfuhl et al. (2015) developed a task that measured the precision of a subjects' memory representations and their confidence about this precision. Subjects saw a squiggly shape and after a brief delay had to indicate which shape they saw. Next, they could set a confidence wedge that should include the shape they just saw. Immediate feedback was provided. The study found that patients with schizophrenia had a less precise visual memory (accuracy in degrees to identify the correct shape) and they too often set a too small confidence interval relative to their precision. That is, the perceived precision was smaller than the actual precision in patients. It follows, that this aberrant precision may be a failure or miscalibration of metacognition or the reflective mind (Fleming, Dolan and Frith, 2012).

As informative as these experiments are for the active inference account, they do not assess cognitive biases in reasoning. Accordingly, extrapolating from these perceptual tasks to delusional reasoning is risky. Delusions are complex (Brett-Jones, Garety and Hemsley, 1987) and there are multiple ways in which the parameters of Bayesian inference might be pathologically affected (Adams et al., 2015).

Delusional reasoning

A characteristic feature of delusion is resistance to updating one's belief in light of contradicting or disconfirming evidence (DSM-V, 2013). It is in this vein that two reasoning tasks have been developed to assess a belief's resistance to be updated: The evidence integration task (Woodward et al., 2006, 2007) and the beads task, a probabilistic inference task introduced by Huq et al. (1988). These tasks assess probabilities, which can be interpreted as posterior beliefs because they are measures of beliefs after new information or evidence was provided.

Evidence integration task

The evidence integration task has been developed by Woodward, Moritz and colleagues (2006, 2007). The logic is as follows: First, an ambiguous scenario is presented followed by four possible interpretations, or causes. These four causes are, in the statistical sense, possible models of the world. One model is deemed as absurd by common standards, three explanations are likely whereof two are "lures" and one is the – by common standards – most appropriate or true explanation. The participant's task is to rate the plausibility for each of the explanations after having read the scenario. It is important to stress that the first plausibility rating is already a conditional probability. That is, the rating is a posterior belief. After the first plausibility rating, the ambiguous scenario is made progressively less ambiguous by

presenting another piece of information S2, and a third final piece of information S3 that resolves all ambiguity. In the example from Woodward et al. (2007) the first evidence is: "Jenny can't fall asleep"; the second is "Jenny can't wait until it is finally morning". And the third is "Jenny wonders how many presents she will find under the tree". The four possible causes are: ''Jenny is nervous about her exam the next day'' (neutral lure); ''Jenny is worried about her ill mother'' (emotional lure); ''Jenny is excited about Christmas morning'' (true); and ''Jenny loves her bed'' (absurd) would be updated based on all evidence. The ambiguity initially favors the lure interpretations. The absurd interpretation should receive no high plausibility rating throughout. The plausibility of the true interpretation should increase with evidence. Three biases have been classified (Table 1, McLean et al., 2016). Subjects who do not downgrade the lure interpretation when the ambiguity is resolved commit a "bias against disconfirmatory evidence" (BADE). Subjects who do not increase their subjective probability of the true interpretation with more incoming information commit a "bias against confirmatory evidence" (BACE). Finally, subjects that rate the absurd interpretation as too plausible are thought to show a "liberal acceptance" (LA) of false beliefs. A crucial extension of this approach would be the assessment of the subjects' probability in the possible causes before any data / information is presented, even if asking for this rating without a scenario seems odd to subjects. This would measure a subjects' prior belief about those causes more directly (Fig 2). If the prior belief in the absurd interpretation is very high (e.g. "Jenny loves her bed" rated as very plausible / above 25%), then it is possible (and rational) to remain at a high probability for the absurd belief after having seen the evidence (Fig 2B). Mathematically, Bayes theorem states that the probability of the absurd interpretation after having read the scenario is proportional to the product of the probability of the scenario under the absurd interpretation and the probability of the absurd interpretation, $p(A|S1) \propto p(S1|A) * p(A)$. Here, S1 is the first piece of evidence from the scenario, and A is the absurd interpretation (state of the world if the absurd interpretation is true). When measuring p(A) beforehand, and p(A|S1) – which is the plausibility rating – we can infer p(S1|A), the likelihood/data or a subject's belief in this scenario occurring under this cause. This would allow measuring directly how much weight the data receives. It might be that patients with delusions do appear to weight the data not much due to a too strong prior belief – a finding contrary to some proposed accounts of delusion (Speechley et al., 2010, Adams et al., 2013).

A second, highly informative extension of this task would be the introduction of a meta-plausibility rating. This would assess how confident a subject was in the plausibility rating they just gave. One would not only assess the plausibility but also the belief in the plausibility. A related but not similar measurement is the variance of the plausibility ratings per cause in the 10 scenarios. If patients vary a lot with how plausible they think the absurd cause is, then they are in a stage of low reliability (liberal acceptance) for alternative beliefs.

Table 1: Evidence integration task and classification of possible outcomes, the two lure interpretations are collapsed

plausibility rating	information 1; p(S1|I)	information 2; p(S1,S2|I)	information 3; p(S1,S2,S3|I)	BADE/BACE/LA
Lure interpretation; p(I=L)	medium-high	medium	low	no BADE
	medium	medium	medium	yes, BADE, no down-regulation of lure

True interpretation; p(I=T)	low-medium	medium-high	high	no BACE
	medium	medium	medium	yes, BACE, no up-regulation of true
absurd interpretation; p(I=A)	low	low	low	no LA
	medium	medium	medium	yes, LA, plausibility of absurd too high

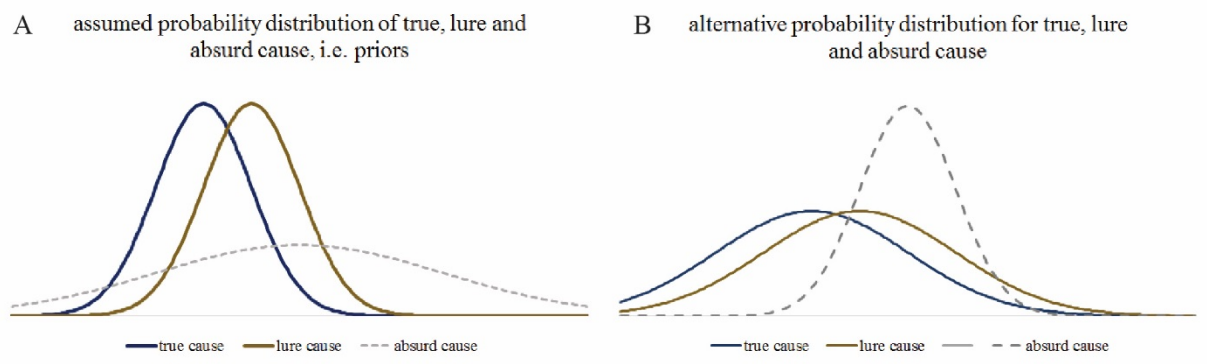

Figure 2. Evidence integration task: this schematic illustrates the importance to first assess the prior beliefs for the provided interpretations in the evidence integration task. Any piece of evidence will update the beliefs according to Bayes theorem. Even if the data is not very likely under the absurd cause, a high initial probability of of the absurd interpretation being true yields a non-negligible plausibility rating.

In summary, measuring a subject's prior belief for each of the four possible interpretations could confirm that the integration process is intact but that the initial strength of the beliefs is aberrant. By assessing how sure they are about their plausibility rating, the reliability can be measured.

The beads task: a probabilistic inference task

A second task that is usually used for assessing cognitive biases in delusions is the beads task. The standard version consists of two jars with different proportions of colored beads. One of the jars contains more white beads, whereas the other jar contains more black beads. The ratios vary between and within studies but are usually symmetrical. For example, jar 1 may contain a ratio of 85:15 of white:black, whereas jar 2 has exactly the opposite ratio, i.e. 15:85 of white:black. Beads are drawn one at a time and replaced such that there are always 100 beads in the jar. Therefore, the subject knows the probability with which a bead will be drawn from jar 1 or 2, i.e., p(black/white bead | jar i) and they have to estimate which of the two jars is more likely to be the source of the drawn beads. Studies vary in whether a maximum of 10 or 20 beads can be drawn and beads might be disguised as fish or sheep to provide a more intuitive appeal (Moritz et al., 2016a, Speechley et al., 2010). There are two versions of the task: In the draws-to-decision (DtD) version of the task, subjects are required to indicate when they are sure they know from which jar the bead has been drawn (Huq et al., 1988). Alternatively, subjects are instructed to indicate the probability of the bead being drawn from jar 1 or jar 2 – the decision threshold (DT) or graded estimate version (Moritz and Woodward 2004). In the DtD version, deciding after fewer than two beads is classified as "jumping to conclusion" and a data gathering bias is diagnosed (Huq et al., 1988, Garety and Freeman, 1999). Indeed, there is consistent evidence that persons with delusions decide on a source jar after having seen fewer beads than do non-deluded persons (McLean et al. 2016, Ross et al. 2015, Dudley et al., 2016). In the graded estimate version, Moritz et al. (2016a) found that

patients assessed the probability similarly to controls but a more lenient threshold was applied to decide from which jar the bead was drawn, e.g. 82% compared to 93%. Speechley et al. (2010) also asked subjects for the probability of the evidence under two possible hypotheses: a lake with predominantly black fish (80:20 ratio) and a lake with equal black and white fish (50:50). They did not ask for draws to decision. Patients with delusions gave the most extreme probability estimates for the hypothesis favoring the data but there was no difference in estimating conflicting data.

An advantage of the beads task relative to the evidence integration task is that the prior beliefs have a normative value. Before any bead has been drawn the a priori estimate of a white bead being drawn should rationally be 50%. Fear and Healy (1997) found no difference in correctly stating this probability among patients and controls. In easy versions of this task (85:15 ratio) persons with delusions do not differ from the Bayesian norm (Dudley et al., 1997). In the 60:40 versions of the task persons with delusions do request too little information. Although, they need more draws to decision the closer the ratio is to 50:50 (e.g. Garety et al., 2015).

Additionally, to elucidate whether subjects base their decision solely on one piece of information, i.e., they have decreased motivation to gather more evidence, one can use two jars with three bead colors (Fig 3). The ratio could, for example, be set to 70:20:10 and 10:20:70 with the colors white:black:red. There are two interesting conditions: First, if one presents a white bead the posterior probability is 87.5% instead of 70% in a 70:30 symmetrical jar setup. More interesting is how subjects respond, in the case in which one would draw first a black bead. Since black beads occur with 20% in both jars, the posterior probability for the bead originating from jar 1 is 50%. This should prevent deciding based on "the correct jar is the jar that has most beads of the drawn bead color". Further, if the ratio in the second jar is 20:10:70 then the posterior probability for jar 1 after having seen one black bead is 67%. Various ratios can be employed to measure the decision threshold. This three-bead color task (and two jars) also minimized miscomprehension of jars (Balzan et al., 2012) changing when the color changes. Appendix A provides the calculations.

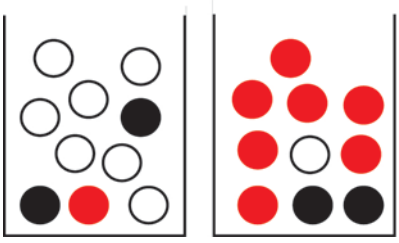

Figure 3. Modified beads task with three colors in the ratio 70:20:10 and 10:20:70 for white, black and red beads

Finally, the beads task is also an ideal tool to measure BADE. Using the graded estimate version of the task, a sequence favouring first jar 1 can be presented. At a point in the sequence, unbeknownst to the subject, the beads will be drawn from jar 2. The participant will be told that the jar can change and their task is to indicate when a change in jar is suspected. It is not about preventing miscomprehension (Balzan et al., 2012) but would measure whether subjects integrate previous data or treat all beads sequentially (Moutoussis et al., 2011). Two studies have employed similar designs. Langdon, Ward and Coltheart (2010) did change the jar and measured the probability but did not ask when the subject was sure that a change in jar occurred nor did they inform subjects about this possibility. Pfuhl et al. (2015) did inform their subjects that a change can occur and measured the probabilities but did not ask for when

a change occurs. In effect, they measured the response to a change but not whether a subject was aware of a change. In a self-monitoring task, Knoblich et al. (2004) found a reduced ability to detect a mismatch but not to act on it in patients with delusions. By asking for probability judgements as well as when they think a change in jar has occurred, one might also be able to dissociate acting from perceiving. The change detection version can also be varied by providing different probabilities of changes and asking for how certain they are that they have detected the change. This comes closer to measuring simultaneously a belief and the belief in a belief.

These modifications may help to find out what the beads task actually measures. Three explanations have been put forward: The data gathering bias might be due to ignoring possible future outcomes (Moutoussis et al., 2011), hypersalience of evidence (Speechley et al., 2010), or a liberal acceptance criterion (Moritz and Woodward, 2004).

Conclusion

The Bayesian brain account is a general framework to explain perception and cognition. Variants of it have been applied to psychopathology, especially psychosis and autism (Adams et al., 2013, 2015, Pellicano and Burr, 2012, van de Cruys et al., 2014). The non-technical account presented here is based on a crucial distinction of how narrowly defined a belief is (its precision) and the reliability or confidence in this being a belief about the true state of the world. This dissociation is important to identify whether patients with delusions have a too precise belief, a too reliable belief or both, and what happens at which stage. This distinction is also important to develop tasks measuring the parameters of the account more directly. Nevertheless, the mechanistic account cannot explain why some abnormal beliefs are causing distress for some but not for other people. Indeed, delusion-prone subjects do not show a similarly strong bias to draw premature conclusions (data gathering bias) than do patients with delusions (McLean et al., 2016, Ross et al., 2015). And believers in the “New religious movement” show a similarly strong bias (Lim et al., 2012). Thus, the Bayesian decision theory is a promising tool but it needs to be extended to incorporate a distress factor. So far, a mechanistic model including an affective component is lacking. Finally, more emphasize should be given to identify differences within delusional patients. This may identify protective as well as risk factors of relapse.

Acknowledgements

I thank Matthias Mittner and Robert Biegler for helpful comments and feedback on the draft manuscript. I am also very grateful to two anonymous reviewers whose comments improved the manuscript greatly.

Declaration of interest

There is no interest to be declared.

Appendix

A) Jar 1 contains white:black in a ratio of 80:20, jar 2 contains white:black in ratio 60:40

Bead sequence: w w w b w b w w w w (favouring 80:20)
P(jar 1) = P(jar 2) = 50%; P(white| jar 1) = 80%, P(white| jar 2) = 60%
After first bead drawn (white bead):

P(jar 1| white) = P(white| jar 1)*P(jar 1) / [P(white| jar 1)*P(jar 1) + P(white| jar 2)*P(jar 2)
P(jar 1|white) = .8*.5 / [.8*.5 + .6*5] = .4/.7 = .57 or 57%; This is the new prior for P(jar 1), noted as P_1(jar1) and 1-P_1 is new prior for P(jar2), i.e. P_1(jar2)

P(jar 1| white, white) = P(white| jar 1)*P_1(jar 1) / [P(white| jar 1)*P_1(jar 1) + P(white| jar 2)*P_1(jar 2)
P(jar 1|white, white) = .8 * .57 / [.8*.57 + .6*(1-.57)] = .64 or 64%; and so on

Table 1 provides the 10 posterior probabilities

Bead	White	White	White	Black	White	Black	White	White	White	white
Jar 1	.57	.64	.70	.54	.61	.44	.51	.58	.65	.71
Jar 2	.43	.36	.30	.46	.49	.56	.49	.42	.35	.29

B) Jar 1 contains white:black:red in ratio of 70:20:10 and jar 2 contains white:black:red in ratio 10:20:70

Bead sequence 1: w b w r w w b w w w
P(jar 1) = P(jar 2) = 50%; P(white| jar 1) = 70%, P(white| jar 2) = 10%, P(black| jar 1) = P(black| jar 2) = 20%, P(red|jar 1) = 10%, P(red|jar2) = 70%
After first bead drawn (white bead)
P(jar 1| white) = P(white| jar 1)*P(jar 1) / [P(white| jar 1)*P(jar 1) + P(white| jar 2)*P(jar 2)
P(jar 1|white) = .7*.5 / [.7*.5 + .1*5] = .35/.4 = .875 or 87.5%. This is the new prior for P(jar 1), noted as P_1(jar1) and 1-P_1 is new prior for P(jar2), i.e. P_1(jar2)
P(jar 1| white, black) = P(black| jar 1)*P_1(jar 1) / [P(black| jar 1)*P_1(jar 1) + P(black| jar 2)*P_1(jar 2)
P(jar 1| white, black) = .2*.875 / [.2*.875 + .2*(1-.875)] = .875 (black is uninformative as it occurs equally in both jars)

Bead sequence 2: b w w w r w b w w w
After first bead drawn (black bead)
P(jar 1| black) = P(black| jar 1)*P(jar 1) / [P(black| jar 1)*P(jar 1) + P(black| jar 2)*P(jar 2)
P(jar 1|black) = .2*.5 / [.2*.5 + .2*5] = .1/.2 = .5 or 50%, again black is uninformative, hence the second draw is the first informative one:
P(jar 1| black, white) = P(white| jar 1)*P(jar 1) / [P(white| jar 1)*P(jar 1) + P(white| jar 2)*P(jar 2)
P(jar 1|black, white) = .7*.5 / [.7*.5 + .1*5] = .35/.4 = .875 or 87.5%

Table 2 proves the 10 posterior probabilities for sequence 1 and 2

Seq 1	White	Black	White	Red	white	white	Black	White	White	white
Jar 1	.875	.875	.98	.875	.98	.997	.997	.999	.999	.999
Jar 2	.125	.125	.02	.125	.02	.003	.003	0	0	0
Seq 2	Black	White	White	White	Red	White	Black	White	White	white
Jar 1	.5	.875	.98	.997	.98	.997	.997	.999	.999	.999
Jar 2	.5	.125	.02	.003	.02	.003	.003	0	0	0

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. Front Psychiatry, 4, 47. doi:10.3389/fpsyt.2013.00047

Adams, R. A., Brown, H. R., & Friston, K. J. (2015). Bayesian inference, predictive coding and delusions. Avant, 5(3), 51-88. doi:10.12849/50302014.0112.0004

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: American Psychiatric Association

Balzan, R. P., Delfabbro, P. H., Galletly, C. A., & Woodward, T. S. (2012). Over-adjustment or miscomprehension? A re-examination of the jumping to conclusions bias. Australian and New Zealand Journal of Psychiatry, 46(6), 532-540. doi:10.1177/0004867411435291

Barch, D. M., Bustillo, J., Gaebel, W., Gur, R., Heckers, S., Malaspina, D., . . . Carpenter, W. (2013). Logic and justification for dimensional assessment of symptoms and related clinical phenomena in psychosis: Relevance to DSM-5. Schizophrenia Research, 150(1), 15-20. doi:http://dx.doi.org/10.1016/j.schres.2013.04.027

Brett-Jones, J., Garety, P., & Hemsley, D. (1987). Measuring delusional experiences: a method and its application. British Journal of Clinical Psychology, 26, Pt 4/.

Buck, K. D., Warman, D. M., Huddy, V., & Lysaker, P. H. (2012). The Relationship of Metacognition with Jumping to Conclusions among Persons with Schizophrenia Spectrum Disorders. Psychopathology, 45(5), 271-275. doi:10.1159/000330892

Coltheart, M. (2007). The 33rd sir Frederick Bartlett lecture cognitive neuropsychiatry and delusional belief. Quarterly Journal of Experimental Psychology, 60(8), 1041-1062. doi:10.1080/17470210701338071

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: A Bayesian framework for understanding models of psychosis. Psychopharmacology, 206(4), 515-530. doi:10.1007/s00213-009-1561-0

Dudley, R. E. J., John, C. H., Young, A. W., & Over, D. E. (1997). Normal and abnormal reasoning in people with delusions. British Journal of Clinical Psychology, 36(2), 243-258.

Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2016). Psychosis, delusions and the "Jumping to Conclusions" reasoning bias: A systematic review and meta-analysis. Schizophrenia Bulletin, 42(3), 652-665. doi:10.1093/schbul/sbv150

Fear, C. F., & Healy, D. (1997). Probabilistic reasoning in obsessive-compulsive and delusional disorders. Psychological Medicine, 27(1), 199-208. doi:10.1017/S0033291796004175

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and function. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1594), 1280-1286. doi:10.1098/rstb.2012.0021

Friston, K. (2005). A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci, 360(1456), 815-836. doi:10.1098/rstb.2005.1622

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. Cognitive Neuroscience, 6(4), 187-214. doi:10.1080/17588928.2015.1020053

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: A critical review of theories and evidence. British Journal of Clinical Psychology, 38(2), 113-154. doi:10.1348/014466599162700

Garety, P., Waller, H., Emsley, R., Jolley, S., Kuipers, E., Bebbington, P., . . . Freeman, D. (2015). Cognitive Mechanisms of Change in Delusions: An Experimental Investigation Targeting Reasoning to Effect Change in Paranoia. Schizophrenia Bulletin, 41(2), 400-410. doi:10.1093/schbul/sbu103

Hemsley, D. R., & Garety, P. A. (1986). The formation of maintenance of delusions – a Bayesian analysis. British Journal of Psychiatry, 149, 51-56. doi:10.1192/bjp.149.1.51

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic Judgements in Deluded and Non-Deluded Subjects. The Quarterly Journal of Experimental Psychology Section A, 40(4), 801-812. doi:10.1080/14640748808402300

Huys, Q. J. M., Guitart-Masip, M., Dolan, R. J., & Dayan, P. (2015). Decision-Theoretic Psychiatry. Clinical Psychological Science, 3(3), 400-421. doi:10.1177/2167702614562040

Knoblich, G., Stottmeister, F., & Kircher, T. (2004). Self-monitoring in patients with schizophrenia. Psychological Medicine, 34(8), 1561-1569. doi:10.1017/S0033291704002454

Langdon, R., Ward, P. B., & Coltheart, M. (2010). Reasoning anomalies associated with delusions in schizophrenia. Schizophrenia Bulletin, 36(2), 321-330. doi:10.1093/schbul/sbn069

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA Faked the Moon Landing-Therefore, (Climate) Science Is a Hoax: An Anatomy of the Motivated Rejection of Science. *Psychological Science, 24*(5), 622-633. doi:10.1177/0956797612457686

Lim, M. H., Gleeson, J. F., & Jackson, H. J. (2012). The jumping-to-conclusions bias in new religious movements. Journal of Nervous and Mental Disease, 200(10), 868-875. doi:10.1097/NMD.0b013e31826b6eb4

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. Frontiers in Human Neuroscience, 5, 20. doi:10.3389/fnhum.2011.00039

McLean, B. F., Mattiske, J. K., & Balzan, R. P. (2016). Association of the Jumping to Conclusions and Evidence Integration Biases With Delusions in Psychosis: A Detailed Meta-analysis. Schizophr Bull. doi:10.1093/schbul/sbw056

Moritz, S., & Woodward, T. S. (2004). Plausibility judgement in schizophrenic patients: Evidence for a liberal acceptance bias. German Journal of Psychiatry, 7(4), 66-74.

Moritz, S., & Woodward, T. S. (2006). Metacognitive control over false memories: A key determinant of delusional thinking. Current Psychiatry Reports, 8(3), 184-190. doi:10.1007/s11920-006-0022-2

Moritz, S., Woodward, T. S., & Lambert, M. (2007). Under what circumstances do patients with schizophrenia jump to conclusions? A liberal acceptance account. British Journal of Clinical Psychology, 46(2), 127-137. doi:10.1348/014466506X129862

Moritz, S., Scheu, F., Andreou, C., Pfueller, U., Weisbrod, M., & Roesch-Ely, D. (2016a). Reasoning in psychosis: Risky but not necessarily hasty. Cognitive Neuropsychiatry, 21(2), 91-106. doi:10.1080/13546805.2015.1136611

Moritz, S., Pfuhl, G., Lüdtke, T., Menon, M., & Andreou, C. A two-stage cognitive theory of the positive symptoms of psychosis. Highlighting the role of lowered decision thresholds. Journal of Behavior Therapy and Experimental Psychiatry. doi:http://dx.doi.org/10.1016/j.jbtep.2016.07.004

Moutoussis, M., Bentall, R. P., El-Deredy, W., & Dayan, P. (2011). Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. Cognitive Neuropsychiatry, 16(5), 422-447. doi:10.1080/13546805.2010.548678

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. Trends in Cognitive Sciences, 16(10), 504-510. doi:10.1016/j.tics.2012.08.009

Pfuhl, G., Tjelmeland, H., & Biegler, R. (2011). Precision and Reliability in Animal Navigation. Bulletin of Mathematical Biology, 73(5), 951-977. doi:10.1007/s11538-010-9547-y

Pfuhl, G., Sandvik, K., Biegler, R., & Tjelmeland, H. (2015). Identifying the Computational Parameters Gone Awry in Psychosis. In Y. Guo, K. Friston, A. Faisal, S. Hill, & H. Peng (Eds.), Brain Informatics and Health (Vol. 9250, pp. 23-32).

Risen, J. L. (2016). Believing what we do not believe: Acquiescence to superstitious beliefs and other powerful intuitions. Psychological Review, 123(2), 182-207. doi:10.1037/rev0000017

Ross, R. M., McKay, R., Coltheart, M., & Langdon, R. (2015). Jumping to Conclusions About the Beads Task? A Meta-analysis of Delusional Ideation and Data-Gathering. Schizophrenia Bulletin. doi:10.1093/schbul/sbu187

Schmack, K., de Castro, A. G. C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J. D., . . . Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. Journal of Neuroscience, 33(34), 13701-13712. doi:10.1523/JNEUROSCI.1778-13.2013

Schwartenbeck, P., FitzGerald, T. H. B., & Dolan, R. (2016). Neural signals encoding shifts in beliefs. Neuroimage, 125, 578-586. doi:http://dx.doi.org/10.1016/j.neuroimage.2015.10.067

Speechley, W. J. M. A., Whitman, J. C. M. A., & Woodward, T. S. P. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. Journal of Psychiatry & Neuroscience : JPN, 35(1), 7-17.

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., . . . Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early

psychosis and psychosis-prone healthy individuals. Proceedings of the National Academy of Sciences of the United States of America, 112(43), 13401-13406. doi:10.1073/pnas.1503916112

van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. Psychological Review, 121(4), 649-675. doi:10.1037/a0037665

Woodward, T. S., Moritz, S., Cuttler, C., & Whitman, J. C. (2006). The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions in schizophrenia. Journal of Clinical and Experimental Neuropsychology, 28(4), 605-617. doi:10.1080/13803390590949511

Woodward, T. S., Buchy, L., Moritz, S., & Liotti, M. (2007). A bias against disconfirmatory evidence is associated with delusion proneness in a nonclinical sample. Schizophrenia Bulletin, 33(4), 1023-1028. doi:10.1093/schbul/sbm013

Von Helmholtz, H (1924). Treatise on physiological optics, 3rd edn, translation by J P C Southall, Optical Society of America