Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

British Journal of Psychology (2010), 101, 621–635
© 2010 The British Psychological Society

The British Psychological Society

www.bpsjournals.co.uk

Subclinical delusional ideation and appreciation of sample size and heterogeneity in statistical judgment

Niall D. Galbraith*, Ken I. Manktelow and Neil G. Morris
Department of Psychology, University of Wolverhampton, UK

Previous studies demonstrate that people high in delusional ideation exhibit a data-gathering bias on inductive reasoning tasks. The current study set out to investigate the factors that may underpin such a bias by examining healthy individuals, classified as either high or low scorers on the Peters *et al.* Delusions Inventory (PDI). More specifically, whether high PDI scorers have a relatively poor appreciation of sample size and heterogeneity when making statistical judgments. In Expt 1, high PDI scorers made higher probability estimates when generalizing from a sample of 1 with regard to the heterogeneous human property of obesity. In Expt 2, this effect was replicated and was also observed in relation to the heterogeneous property of aggression. The findings suggest that delusion-prone individuals are less appreciative of the importance of sample size when making statistical judgments about heterogeneous properties; this may underpin the data gathering bias observed in previous studies. There was some support for the hypothesis that threatening material would exacerbate high PDI scorers' indifference to sample size.

Proponents of the continuity model of mental illness (e.g., Claridge, 1988; Claridge & Beech, 1995; van Os, Linscott, Myin-Germeys, Delespaul, & Krabbendam, 2009) argue that psychotic symptoms represent extreme versions of everyday behaviours. By this rationale, there are individuals within the normal population who have a predisposition towards psychosis, but whose experiences are not extreme enough to be classed as mental illness. This subclinical category of experience is referred to as schizotypy (Claridge, 1988; Meehl, 1962, 1990). Some authors have suggested that it may be more fruitful to study people who fall within the subclinical range of psychosis than patients (Linney, Peters, & Ayton, 1998; see also Bentall, 2003). As Claridge (1988) notes, research on patients can be hampered by medication effects and motivation, plus the nature and severity of symptoms and experiences can vary considerably over time, and some individuals can also suffer general cognitive decline affecting memory and attention. The development of schizotypy questionnaires (e.g., Oxford–Liverpool

* *Correspondence should be addressed to Dr Niall D. Galbraith, Department of Psychology, School of Applied Sciences, University of Wolverhampton, Wolverhampton WV1 1SB, UK (e-mail: n.galbraith@wlv.ac.uk).*

DOI:10.1348/000712609X479384

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Inventory of Feelings and Experiences (O-LIFE): Mason, Claridge, & Jackson, 1995; the Peters *et al.* Delusions Inventory (PDI): Peters, Day, & Garety, 1996) has enabled researchers to investigate some of the traits associated with schizophrenia whilst avoiding these methodological difficulties. In line with the continuity model of mental illness, the current study investigates subclinical delusional ideation and its effect upon reasoning.

A growing body of literature demonstrates that patients who experience delusions often exhibit a number of reasoning biases. Using a probabilistic reasoning problem known as the 'beads' task, Huq, Garety, and Hemsley (1988) found that delusional patients tended to make decisions on the basis of less evidence when compared to controls. Participants have to decide which of two jars containing different ratios of red and black beads (85:15 and 15:85) has been chosen by the experimenter, by asking the experimenter to draw a sequence of beads, one at a time, from the jar. Huq *et al.* (1988) found that patients with delusions tended to request fewer beads than the controls before making a decision.

This jump-to-conclusions bias has since been replicated in numerous other studies using the beads task (e.g., Broome *et al.*, 2007; Moritz & Woodward, 2005; Peters, Thornton, Siksou, Linney, & MacCabe, 2008). Dudley, John, Young, and Over (1997) found that people with delusions requested fewer beads than controls whether the ratio of black to red was 60:40 or the relatively easier 85:15 ratio. They concluded therefore that the bias did not merely reflect impulsiveness, as the people with delusions took account of the base rate change and delayed their decision in the harder 60:40 ratio version; however, they were still deciding on the basis of less evidence than the other groups. It was noted by Dudley *et al.* (1997) that although the deluded patients were more hasty, in terms of Bayes' theorem, the amount of information they requested was closer to the optimum amount. However, more importantly it signifies that deluded patients reason differently from controls, and it has been well demonstrated that humans are not always intuitively strong Bayesian reasoners (Gigerenzer & Hoffrage, 1995). John and Dodgson (1994) developed an inductive task which they called the 'Twenty Questions Game', and they found that deluded participants requested less information before making judgments that were also of poorer quality when compared to controls.

From the studies reviewed so far, deluded patients appear consistently to base judgments on less information than controls. However, on the beads task, patients were just as likely as controls to guess the correct jar. Indeed when the number of beads viewed is determined by the experimenter and not the participant, there are often no differences between the deluded and control groups (Dudley *et al.*, 1997, Expt 2; Young & Bentall, 1997). Dudley *et al.* (1997) also failed to find group differences on a coin tossing task designed to test for appreciation of base rate shift. This has led to the suggestion by Dudley *et al.* (1997) that delusional patients do not have a general reasoning deficit, but rather they are biased to gather less data before making their decisions. Linney *et al.* (1998), found further evidence for a data gathering bias in high PDI scorers and also speculated on the processes underpinning it. Using an adaptation of the coin-tossing task (Blackmore & Troscianko, 1985) high PDI scorers were less likely to change their ratings of how biased the coin was in response to different sample sizes. Linney *et al.* (1998) therefore speculated that the data-gathering bias is perhaps linked to an insensitivity to the effects of random variation, whereby such individuals may be less likely to appreciate that statistical judgments are more reliable when based upon a larger sample. This failure to appreciate the importance of sample size may in

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

turn lead to early termination of data gathering, perhaps resulting in decisions or conclusions that are based on weaker evidence. Furthermore, a failure to appreciate the effects of random variation may prevent delusional individuals from attributing events merely to chance, perhaps leading to a search for alternative causes instead.

However, numerous studies of statistical reasoning in the normal population have revealed that it is common for people to be insensitive to varying sample sizes (Jepson, Krantz, & Nisbett, 1983; Kahneman & Tversky, 1972; Nisbett & Borgida, 1975; Nisbett, Krantz, Jepson, & Fong, 1982; Nisbett, Krantz, Jepson, & Kunda, 1983; Nisbett & Ross, 1980; Sedlmeier & Gigerenzer, 1997). Studies have demonstrated that one of the most important factors concerns people's beliefs about the homogeneity of the sample (e.g., Nisbett *et al.*, 1983). Objects that are homogeneous share characteristics that are universal and this justifies a generalization from a small sample. In contrast, objects that are heterogeneous have features that vary across the population, and in such cases generalization from small samples is more risky (Nisbett *et al.*, 1983). Nisbett and colleagues asked their participants to imagine that they encountered imaginary minerals, animals, and people on a previously undiscovered island. The objects had distinctive characteristics with varying degrees of homogeneity and were encountered in various sample sizes (1, 3, and 20). Nisbett *et al.* found that their participants were less likely to generalize from a sample of 1 if the characteristic was expected to be heterogeneous (e.g., the degree to which tribes-people were obese). However, larger samples of heterogeneous characteristics led to greater generalization. Conversely, participants were more likely to generalize from samples of 1 if the object/characteristic was homogeneous (e.g., the electric conductivity in a chemical element).

If high PDI scorers are less likely to appreciate the importance of sample size in statistical inference, one might expect that they would also be more likely to generalize from small heterogeneous samples than low PDI scorers are. The authors are not aware of any studies which have examined the importance of heterogeneity/homogeneity in delusional thinking. The current experiment aimed to address this by presenting the Nisbett *et al.* (1983) task to high and low PDI scorers. All participants were asked to imagine that they encountered certain minerals/animals/people, with particular characteristics, in samples of 1, 3, and 20. For each object and for each of the three sample sizes, participants were asked to estimate what percentage of the population shared the same characteristic. It was predicted that high PDI scorers would make higher (and therefore more confident) estimates from heterogeneous samples of 1 than low PDI scorers.

Method

Participants

There were 70 undergraduate volunteers (58 females), with an average age of 22 years ($SD = 6$ years 2 months).

Design

A mixed design was used with PDI group (high and low) as the between-groups factor, and with sample (1, 3, and 20), and property (high homogeneity, medium homogeneity, and low homogeneity) manipulated within groups.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Materials and procedure

The PDI was used as a measure of delusional ideation (Peters *et al.*, 1996). The PDI is designed specifically for the normal population rather than psychiatric patients. Participants respond 'yes' or 'no' to each question. To every question where they answer 'yes', participants are also required to indicate levels of distress, preoccupation, and conviction on a five-point scale. Scores range from 0 to 336. An advantage of the PDI is that it measures all of the common types of delusional ideation, whilst also capturing the multidimensionality of beliefs. The PDI has also demonstrated good reliability and validity (Peters, Joseph, Day, & Garety, 2004). The AH4 (Alice Heim 4) was employed as a measure of verbal intelligence and has been demonstrated as reliable and valid (Heim, Watts, & Simmonds, 1975). The AH4 presents the testee with 65 questions on verbal reasoning to be completed within 10 min, and then 65 on spatial reasoning. Here, only the verbal reasoning section was presented to participants. The participants' truthfulness when completing questionnaires was assessed using the lie scale from the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1976). The participants were tested in groups of up to 15 in a large practicals laboratory.

There were three versions of the reasoning task: one for each of the three sample sizes. The $N = 1$ version read as follows:

> Imagine that you are an explorer who has landed on a little known island in the Southeastern Pacific. You encounter several new animals, people and objects. You observe the properties of your 'samples' and you need to make guesses about how common these properties would be in other animals, people or objects of the same type.
>
> Suppose you encounter a new bird, the 'shreeble'. It is blue in colour. What percentage of all shreebles on the island do you expect to be blue (*write your answer on the line below*)?
>
> Suppose the shreeble you encounter is found to nest in a eucalyptus tree, a type of tree which is fairly common on the island. What percent of all shreebles on the island do you expect to nest in eucalyptus trees?
>
> Suppose you encounter a native to the island, who is a member of a tribe he calls the Barratos. He is obese. What percent of male Barratos do you expect to be obese?
>
> Suppose the Barratos man is dark skinned. What percent of male Barratos do you expect to be dark skinned (as opposed to very dark skinned, olive skinned or light skinned)?
>
> Suppose you encounter what the physicist on your expedition describes as an extremely rare element called floridium. Upon being heated to a very high temperature, it burns with a green flame. What percent of all samples of floridium found on the island do you expect to burn with a green flame?
>
> Suppose the sample of floridium, when drawn into a filament, is found to conduct electricity. What percent of all samples of floridium found on the island do you expect to conduct electricity?

The $N = 3$ and $N = 20$ conditions were phrased identically except for the number of examples encountered. In the original Nisbett *et al.* (1983) study, the participants found the floridium flame colour and its conductivity, as well as the skin colour of the Barratos tribesmen as highly homogenous characteristics. Nisbett *et al.*'s (1983) participants regarded the shreeble's colour and its tendency to nest in the Eucalyptus tree as less homogeneous but still not as heterogeneous as the obesity of the Barratos. Here, it was expected that the two PDI groups will differ in the probability estimates for the low homogeneity characteristic of obesity based on a sample of 1.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

The order of presentation was counterbalanced and each version of the reasoning task was interspersed with the PDI, lie scale and the AH4.

Results

Participants were divided into high and low scoring groups on the basis of their scores on the PDI. Those scoring up to and including the median of 55 were placed into the low PDI group, those scoring above the median comprised the high PDI group. An Independent *t* test revealed no significant difference between the groups in terms of age ($t(68) < 1$): the high PDI group mean was 21 years and 9 months ($SD = 5$ years, 7 months), the low PDI group mean was 22 years and 3 months ($SD = 6$ years, 9 months). Nor was there a significant difference between-groups on the lie scale ($t(68) < 1$): the high PDI group mean was 7.20 ($SD = 4.15$), the low PDI mean was 6.94 ($SD = 5.54$). The groups also scored similarly on the AH4 ($t(68) < 1$): the high PDI group's mean score was 41.06 ($SD = 9.18$), the low PDI group's mean was 40.71 ($SD = 8.31$). The high PDI group's mean PDI score was 85.00 ($SD = 24.73$), the low PDI group's mean PDI score was 39.97 ($SD = 11.75$). There were 4 males and 31 females in the low PDI group, with 8 males and 27 females in the high PDI group. The males' PDI scores did not differ significantly from the females' ($t(68) < 1$; male mean = 69.42, $SD = 30.51$; female mean = 60.26, $SD = 29.17$).

A three-way, mixed design ANOVA examined the effects of PDI group (high and low), sample (1, 3, and 20), and property (high homogeneity, medium homogeneity, and low homogeneity) upon probability estimates. Sample and property were manipulated within groups (see Figure 1). The property factor was formed by grouping the six properties (shreeble colour, nest, Barratos skin colour, obesity, floridium flame colour, floridium conductivity) in accordance with the homogeneity estimates provided by the participants in the original Nisbett *et al.* (1983) study: high homogeneity (flame, conductivity, skin), medium homogeneity (shreeble colour, nest), and low homogeneity (obesity). Partial eta squared (η_p^2) was included as a measure of effect size.[1]

There was no main effect of PDI Group ($F(1, 68) = 1.65$; $p = .204$; $\eta_p^2 = .024$; MSE = 2450.21). There was a main effect of Sample ($F(2, 136) = 68.25$; $p = .001$; $\eta_p^2 = .51$; MSE = 505.02). Simple effects showed that the samples of 1 produced lower estimates than samples of 3 ($F(1, 69) = 44.94$; $p = .001$; $\eta_p^2 = .39$; MSE = 126.58) and 20 ($F(1, 69) = 105.31$; $p = .001$; $\eta_p^2 = .60$; MSE = 218.63). There was also a main effect of property ($F(2, 136) = 72.77$; $p = .001$; $\eta_p^2 = .52$; MSE = 481.57) with simple effects showing that the low homogeneity category produced lower estimates than both the medium homogeneity category ($F(1, 69) = 54.88$; $p = .001$; $\eta_p^2 = .44$; MSE = 210.01) and the high homogeneity category ($F(1, 69) = 102.06$; $p = .001$; $\eta_p^2 = .60$; MSE = 214.31). There was no significant interaction between PDI Group and Sample ($F(2, 136) < 1$). However, there was a significant interaction between PDI group and property ($F(2, 136) = 4.56$; $\eta_p^2 = .06$; $p = .012$; MSE = 481.57). Simple effects showed that the PDI groups did not differ in their estimates of high homogeneity categories (flame, conductivity, skin; $F(1, 68) < 1$). However, they did differ in terms of their estimates of the low homogeneity category (obesity): the high PDI scorers' estimates

[1] *According to Cohen (1988), when interpreting partial eta squared, a value of 0.01 constitutes a small effect size, 0.06 is a medium effect size and 0.14 a large effect size.*

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

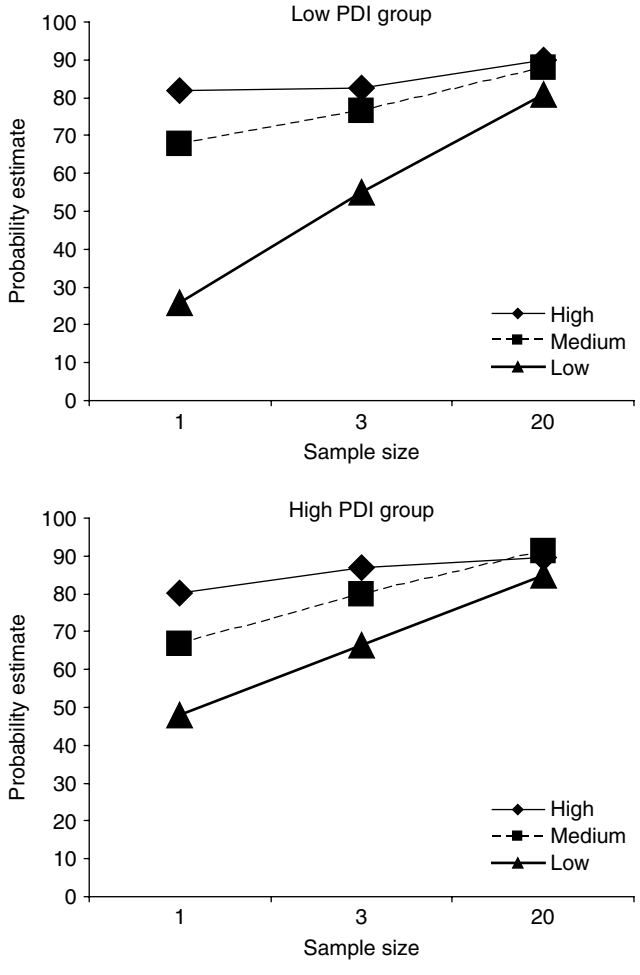


Figure 1. High and low PDI scorers' probability estimates from Expt 1.

were significantly higher ($F(1, 68) = 5.77$; $p = .019$; $\eta_p^2 = .08$; MSE = 474.22). There was also a significant interaction between property and sample ($F(4, 272) = 45.56$; $p = .001$; $\eta_p^2 = .40$; MSE = 135.38). Contrasts show that both the high homogeneity properties and the low homogeneity property obesity had inflated estimates from a sample of 20 compared with a sample of 1, however, this pattern was much more pronounced in the low homogeneity category ($F(1, 69) = 118.47$; $p = .001$; $\eta_p^2 = .63$; MSE = 811.28; see Figure 1). Finally, as predicted, there was also a three-way interaction ($F(4, 272) = 5.26$; $p = .001$; $\eta_p^2 = .07$; MSE = 135.38). Simple effects analyses revealed that the high PDI group made higher probability estimates than the low PDI group with regard to the low homogeneity property in the $N = 1$ condition ($F(1, 68) = 10.88$; $p = .002$; $\eta_p^2 = .14$; MSE = 786.70). As can be seen from Figure 1, the groups' probability estimates across the three sample sizes are similar for the high and medium homogeneity properties but not for the low homogeneity property. The groups did not differ in their $N = 1$ estimates for the high homogeneity properties ($F(1, 68) < 1$) or for the medium homogeneity properties ($F(1, 68) < 1$).

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Discussion

The findings show good support for the predictions in that high PDI scorers produced more confident probability estimates from a sample of 1 for the heterogeneous characteristic of obesity. The high PDI scorers were more confident that the entire population of Barratos tribesmen were obese when faced only with a sample of 1. This tentatively suggests that they were less likely to regard obesity as a highly variable characteristic. Linney *et al.* (1998) speculated that people high in delusional ideation might be less sensitive to random variation and less sensitive to the importance of sample size in statistical judgment. The current experiment supports this and goes beyond Linney *et al.*'s suggestions, by revealing the cognitive processes which underlie this insensitivity: a relatively poor appreciation of the variability of heterogeneous properties and that generalization about such properties is unsafe when based upon small samples. The findings may also shed new light on the cognitive processes which underlie the data-gathering bias in delusional people (Dudley *et al.*, 1997; Peters *et al.*, 2008). Individuals who are less appreciative of the importance of sample size and the variability of heterogeneous properties are more likely to terminate data gathering early thus basing their conclusions on a smaller (and less reliable) sample of data, perhaps making them more vulnerable to developing bizarre beliefs.

The overall results are also consistent with the findings of Nisbett *et al.* (1983): their data revealed that probability estimates from a sample of 1 for the obesity trait were the lowest of all the items and this is the pattern of responding observed in both groups here in the current study. Nisbett *et al.* (1983) found that probability estimates from $N = 1$ varied according to the item in question. They found that with regard to floridium conductivity, flame colour, and skin colour, participants' probability estimates from a sample of 1 were almost as high as estimates from a sample of 20. This again mirrors the current findings. With regard to shreeble colour and nest, although Nisbett *et al.* found that probability estimates were slightly lower, they were still considerably higher than estimates of obesity when based on a sample of 1. Hence the current participants' intermediate $N = 1$ probability estimates for shreeble colour and nest (medium homogeneity) reflect participants' judgment that these features of shreebles are neither particularly homogenous nor heterogeneous.

EXPERIMENT 2

Previous studies suggest that the reasoning of delusional patients is particularly disrupted when reasoning tasks contain emotional or threatening content (Kemp, Chua, McKenna, & David, 1997; Mujica-Parodi, Malaspina, & Sackeim, 2000; Young & Bentall, 1997). Both Kemp *et al.* (1997) and Mujica-Parodi *et al.* (2000) discovered that delusional patients endorsed more invalid conclusions on syllogistic reasoning tasks when the task content was emotional. Young and Bentall (1997) found that deluded participants gave more certain statistical judgments during early trials on the beads task when meaningful material was introduced (although depressed participants displayed this pattern also). In light of this work, the aim here was to examine whether adding threatening content to the task used in Expt 1 would lead to greater differences between the high and low PDI groups than the neutral content. The participants were presented with the same task used in the previous experiment but with an additional feature. The sample of Barratos tribesmen was described as aggressive, as well as obese and olive-skinned. Aggression was included firstly as it is a threatening characteristic

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

and secondly because it is likely to be regarded as a relatively heterogeneous trait. It was predicted firstly that compared to low PDI scorers, the high PDI group would make higher probability estimates from a sample of 1 with regard to obesity, as in Expt 1. In addition, if the high PDI group was likely to make higher probability estimates from $N = 1$ with regard to heterogeneous properties, one would expect this effect to be more pronounced with a threatening property such as aggression.

Method

Participants

There were 66 undergraduate volunteers (57 females), with an average age of 22 years 6 months ($SD = 5$ years 7 months) all recruited from a large urban UK university.

Design

The mixed design was identical to that of Expt 1 with PDI group (high and low) as the between-groups factor, and with sample (1, 3, and 20) and property (high homogeneity, medium homogeneity, and low homogeneity) manipulated within groups.

Materials and procedure

The PDI was once again used to measure delusional ideation. The AH4 was used as a measure of verbal intelligence. The participants' truthfulness when completing questionnaires was again tested by using the EPI lie scale (Eysenck & Eysenck, 1976).

The reasoning task was the same as that used in Expt 1, but for two changes. Firstly, the inclusion of aggression and secondly, the Barratos skin colour was this time 'olive' to avoid racial stereotyping. Once again the order of presentation was counterbalanced with each version of the reasoning task being interspersed with the PDI, lie scale and the AH4.

Results

Those scoring up to and including the median of 51 on the PDI comprised the low PDI group, those scoring above 51 comprised the high PDI group. An Independent *t* test revealed no difference between the groups in terms of age ($t(63) < 1$): the high PDI group's mean age was 22 years and 4 months ($SD = 4$ years 10 months), the low PDI mean age was 22 years and 6 months ($SD = 6$ years and 3 months). The groups did not differ significantly on the AH4 ($t(64) = 1.32$; $p = .192$): the high PDI group's mean AH4 score was 40.22 ($SD = 9.03$), the low group's mean AH4 score was 43.15 ($SD = 9.01$). Nor did the groups differ significantly on the lie scale ($t(64) = 1.59$; $p = .117$): the high PDI group mean Lie score was 5.28 ($SD = 3.64$), the low PDI mean was 6.85 ($SD = 4.35$). The high PDI group's mean PDI score was 89.69 ($SD = 29.83$), the low PDI group's mean was 31.68 ($SD = 11.78$). There were 3 males and 31 females in the low PDI group, in the high PDI group there were 5 males and 26 females. Again the males' PDI scores did not differ significantly from the females' ($t(63) < 1$; male mean = 69.00; $SD = 36.43$; female mean = 57.56; $SD = 36.46$).

A three-way, mixed design ANOVA examined the effects of PDI group (high and low), sample (1, 3, and 20), and property (high homogeneity, medium homogeneity, and low

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

homogeneity) upon probability estimates. Again, sample and property were manipulated within groups (see Figure 2). As in Expt 1, the property factor was formed by grouping the six properties (shreeble colour, nest, Barratos skin colour, obesity, floridium flame colour, floridium conductivity) in accordance with the homogeneity estimates provided by participants in the Nisbett *et al.* (1983) study. However, in this experiment the aggression property was grouped with obesity in the low-homogeneity category: high homogeneity (flame, conductivity, skin), medium homogeneity (shreeble colour, nest) and low homogeneity (obesity, aggression).

There was no main effect of PDI group ($F(1, 64) < 1$) but there was a main effect of sample ($F(2, 128) = 171.13$; $p = .001$; $\eta_p^2 = .73$; MSE = 194.90), with simple effects analyses showing that samples of 1 produced lower probability estimates than samples of 3 ($F(1, 65) = 89.57$; $p = .001$; $\eta_p^2 = .58$; MSE = 66.51) and 20 ($F(1, 65) = 241.01$; $p = .001$; $\eta_p^2 = .79$; MSE = 80.16). There was also a main effect of property ($F(2, 128) = 64.51$; $p = .001$; $\eta_p^2 = .50$; MSE = 585.11), with simple effects showing that the low homogeneity category produced lower estimates than both the medium

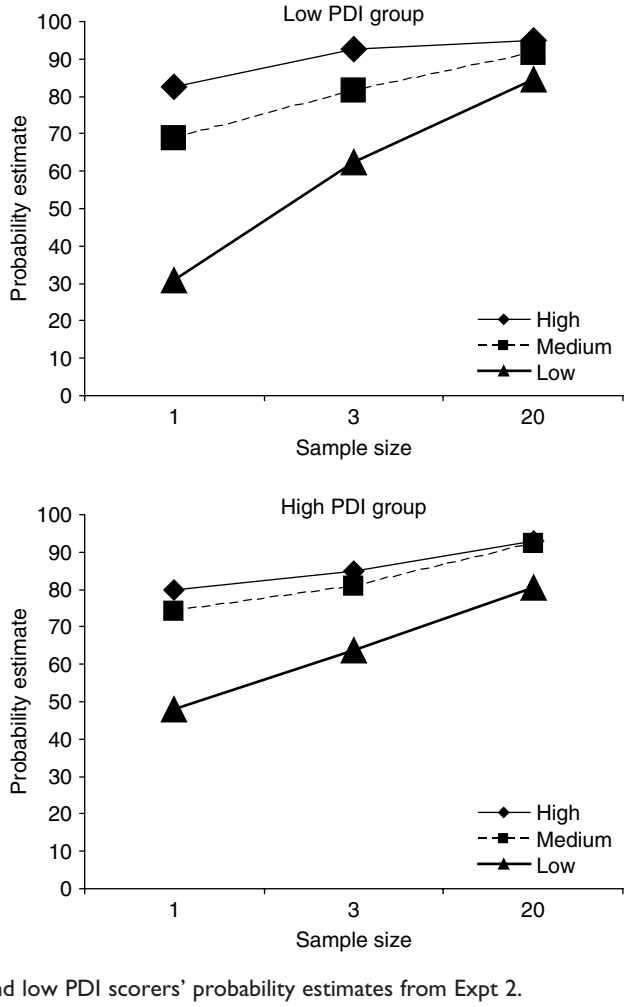


Figure 2. High and low PDI scorers' probability estimates from Expt 2.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

homogeneity category ($F(1, 65) = 59.66$; $p = .001$; $\eta_p^2 = .48$; MSE = 223.12) and the high homogeneity category ($F(1, 65) = 125.93$; $p = .001$; $\eta_p^2 = .66$; MSE = 194.24).

The two-way interactions also produced some effects. Firstly, the interaction between PDI group and property was not strong enough to reach significance ($F(2, 128) = 1.64$; $p = .20$; $\mu_p^2 = 0.03$; MSE = 585.11). However, there was a significant interaction between PDI group and sample ($F(2, 128) = 5.03$; $p = .008$; $\eta_p^2 = .07$; MSE = 194.90). A contrast showed that both groups gave lower estimates for $N = 1$ than $N = 20$ samples, however, this pattern was much more pronounced in the low PDI group ($F(1, 64) = 6.06$; $p = .017$; $\eta_p^2 = .09$; MSE = 127.61). There was also an interaction between sample and property ($F(4, 256) = 39.12$; $p = .001$; $\eta_p^2 = .37$; MSE = 106.33), contrasts showed the $N = 20$ estimates were greater than $N = 1$ estimates but this difference was more pronounced with the low homogeneity properties than with the high homogeneity properties ($F(1, 64) = 107.37$; $p = .001$; $\eta_p^2 = .63$; MSE = 560.79).

There was a three-way interaction ($F(4, 256) = 5.49$; $p = .001$; $\eta_p^2 = .08$; MSE = 106.33). A simple effects analysis revealed that the high PDI group made higher $N = 1$ probability estimates than the low PDI group with regard to the low homogeneity properties ($F(1, 64) = 8.63$; $p = .005$; $\eta_p^2 = .12$; MSE = 511.04). Unpacking the low homogeneity property, two further simple effects analysis examined the PDI groups on their $N = 1$ estimates for obesity and aggression. The high PDI group gave higher estimates in relation to both obesity ($F(1, 64) = 4.79$; $p = .032$; $\eta_p^2 = .07$; MSE = 588.45) and aggression ($F(1, 64) = 7.30$; $p = .009$; $\eta_p^2 = .10$; MSE = 871.13). Figure 2 shows that the groups are similar in their probability estimates across the three sample sizes for all properties apart from obesity and aggression.

A further two-way ANOVA was carried out, to examine whether aggression had a more pronounced effect upon the groups' estimates than obesity at $N = 1$. PDI group and property (within groups; obesity $N = 1$, aggression $N = 1$) were the factors. Although in the direction predicted (both groups gave higher $N = 1$ estimates for aggression but this pattern was more pronounced in the high PDI group), the interaction did not reach significance ($F(1, 64) < 1$; see Figure 3).

Discussion

The results again support the predictions in that the high PDI group produced higher probability estimates from a sample of 1 for the heterogeneous characteristics of obesity (thus replicating the effect from Expt 2) and aggression. The high PDI group was more confident (than the low PDI group) that the entire population of Barratos tribesmen were obese and aggressive when estimating merely from a sample of 1. This suggests that they were less apt than the low PDI scorers to regard aggression as a highly variable attribute. These results again extend the findings of Linney *et al.* (1998) by providing new evidence for the cognitive processes underlying high PDI scorers' insensitivity to sample size. The difference between the groups' $N = 1$ mean estimates was greater with regard to the threatening property of aggression than the neutral property of obesity, but a comparison of these effect sizes failed to reach significance. Hence, this study does not fully support previous studies which have found that reasoning in delusional patients is particularly affected when threatening or emotional content is used (Kemp *et al.*, 1997; Mujica-Parodi *et al.*, 2000; Young & Bentall, 1997).

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

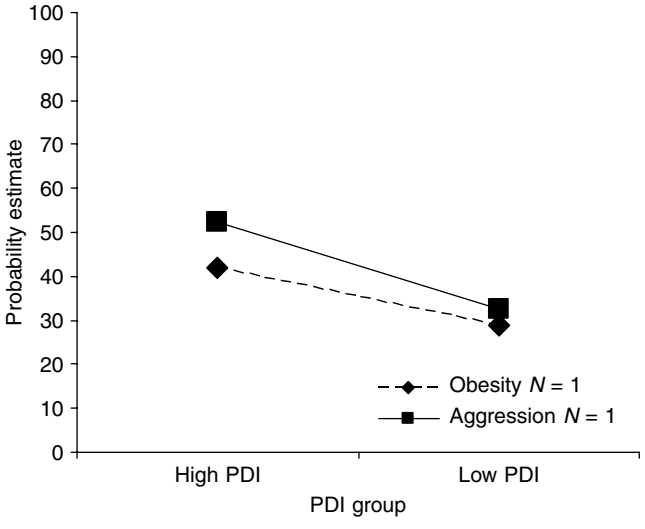


Figure 3. High and low PDI scorers' $N = 1$ probability estimates for obesity and aggression.

GENERAL DISCUSSION

Together the findings provide evidence for the cognitive processes which may underlie a poor appreciation of the importance of sample size in statistical judgment displayed by people high in subclinical delusional ideation. Firstly, in Expt 1, the high PDI scorers (compared to low PDI scorers) were more confident when generalizing about a heterogeneous human characteristic (obesity) from a sample of 1. This effect was replicated in Expt 2, where, compared to the low PDI scorers, the high PDI group was more confident when generalizing from a sample of 1 in relation to both obesity and aggression. These findings indicate that high PDI scorers are less cautious when generalizing about a highly variable property from just a single instance. Despite the two-way interaction between PDI group and sample in Expt 2, the presence of the three-way interaction in both experiments suggests that the difference between the groups does not merely represent a relative failure to appreciate the importance of sample size, but rather, the importance of sample size when making judgments about heterogeneous properties.

The findings are consistent with those of Linney *et al.* (1998), who found that high PDI scorers were less able to take account of varying sample sizes on a statistical reasoning problem. Linney *et al.* (1998) speculated that high PDI scorers are insensitive to the effects of random variation, and are less likely to appreciate that statistical judgments are more reliable when based upon a larger sample. The current findings go further than the suggestions of Linney *et al.*, by highlighting one of the specific cognitive processes which may underlie an insensitivity to sample size. That is, a poor appreciation of the variability of heterogeneous characteristics. Both groups reasoned statistically, in that their estimates for $N = 1$ samples were lower than those of 3 and 20, plus both groups' estimates were more cautious when generalizing about the low homogeneity properties from a single instance. However, this cautiousness was less pronounced in the high PDI group. This relative failure to appreciate the importance of sample size when generalizing about heterogeneous properties may lead to early termination of the data gathering process, resulting in hasty decisions or conclusions

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

that are based on a smaller, and therefore less reliable body of evidence. This process may indeed contribute to the formation or maintenance of delusional ideas as suggested by Garety (1991). Furthermore, a failure to appreciate the variability of heterogeneous characteristics may prevent delusional individuals from attributing events merely to chance, perhaps leading to a search for alternative (delusional) causes instead. The data gathering bias and how this might contribute to delusions has been well documented (Fine, Gardner, Craigie, & Gold, 2007), however, the current study is the first to demonstrate the importance of homogeneity/heterogeneity. The current results suggest that delusional ideas may arise from a relative failure to adjust confidence in judgments when those judgments are based around heterogeneous properties. The two groups were equally confident when generalizing about homogeneous properties. However, the results suggest that people low in delusional ideation are more cautious when making judgments about heterogeneous entities and therefore they may be less likely to form beliefs based upon unsound generalizations. Hence, the development of delusional beliefs may depend on limited data gathering, but more specifically, limited data gathering in relation to heterogeneous properties where larger sets of data are particularly important.

Evidence of insensitivity to random variation in those with subclinical delusional ideation is consistent with the continuity model of psychosis (e.g., Claridge, 1988; van Os *et al.*, 2009) and indicates that behaviours found in delusional patients may be evident in non-patients, albeit perhaps in a milder form. Furthermore, evidence of reasoning biases in the subclinical range also suggests that such biases have a role to play in delusion formation and that they may exist in people who may yet develop schizophrenia. The tasks used in the current study, also offer a methodological advantage over the beads task as a means of testing appreciation of sample size within a clinical population. The beads task requires participants to actually gather data, which may be an activity that is impaired in patients due to their general apathy or medication effects, as highlighted by Claridge (1988). By contrast, the tasks used here do not involve actual data gathering, but do allow one to test the participants' appreciation of sample size. Hence patients' performance on these tasks may be less impaired by the aforementioned confounding factors.

The effect sizes for the three way interactions in Expts 1 and 2 are of medium size according to Cohen (1988). However, given that both groups were drawn from the normal population, the level of delusional ideation even in the high PDI group is relatively moderate compared to that which would be expected from a clinical sample (Peters *et al.*, 2004). Therefore any group effects are likely to be subtle. However, this is not the first study to report cognitive differences between high and low PDI groups from the non-clinical population using sample sizes of less than 100 (e.g., Warman, Lysaker, Martin, Davis, & Haudenschield, 2007; Ziegler, Rief, Werner, Mehl, & Lincoln, 2008). The findings of the current study add strength to the idea that comparing the cognitive processes of high and low PDI scorers from the normal population may provide important insights into the processes that underpin clinical delusions.

There were no differences between the sexes; however, it is difficult to draw concrete conclusions from these data considering the relatively low number of males. Earlier work has demonstrated that females score more highly than males on magical thinking and other positive schizotypal features (Claridge & Hewitt, 1987), so future work should allow for more reliable comparisons between the sexes. There were no differences between the PDI groups on the lie scale scores. The scores recorded in all experiments were somewhat higher than the mean score (3.8) taken from normative

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

data (Goldberg & Williams, 1988); however, the truthfulness of the PDI groups relative to each other did not vary.

The inclusion of the threatening property 'aggression' in Expt 2 was associated with a greater effect size on estimations than that of the heterogeneous but emotionally neutral trait of obesity. However, this effect of aggression was not *statistically* greater than that of obesity, hence the current study cannot offer support for the notion that reasoning in high PDI scorers may be particularly disrupted by the inclusion of threatening material. This effect might be stronger on a task with greater emotional valence or perhaps when comparing delusional patients with healthy controls.

Although the current findings suggest that the observed differences in reasoning between the groups were due to levels of delusional ideation, one cannot rule out the possibility that other (unmeasured) personality variables may covary with delusional ideation and may better explain the effects. For example, the relationship between need for closure and delusional ideation is still unclear (Freeman *et al.*, 2006; McKay, Langdon, & Coltheart, 2006) and delusional ideation has also been associated with neuroticism (Larøi, Van der Linden, DeFruyt, van Os, & Aleman, 2006). Future studies of reasoning biases in delusional thinking may strengthen their conclusions by ruling out possible personality covariates.

To conclude, high PDI scorers from the normal population were found to make higher probability estimates from a sample of 1 with regard to two heterogeneous properties: obesity and aggression. Overall the findings provide support for the notion that people high in delusional ideation are less appreciative of the importance of sample size and heterogeneity in statistical judgment and that this may underpin the data gathering bias observed in previous studies.

Acknowledgements

We would like to thank Maxwell Roberts, Peter Mitchell, and two anonymous reviewers for their helpful comments on this paper.

References

Bentall, R. P. (2003). *Madness explained: Psychosis and human nature*. London: Penguin Books.

Blackmore, S., & Troscianko, T. (1985). Belief in the paranormal, probability judgements, illusionary control and the 'chance baseline shift'. *British Journal of Psychology*, *76*, 459–468.

Broome, M. R., Johns, L. C., Valli, I., Woolley, J. B., Tabraham, P., Brett, C., *et al.* (2007). Delusion formation and reasoning biases in those at clinical high risk for psychosis. *British Journal of Psychiatry*, *191*, 38–42.

Claridge, G. S. (1988). Schizotypy and schizophrenia. In P. Bebbington & P. McGuffin (Eds.), *Schizophrenia: The major issues* (pp. 187–200). Oxford: Heinmann Professional.

Claridge, G. S., & Beech, T. (1995). Fully and quasi-dimensional constructions of schizotypy. In A. Rain, T. Lencz, & S. A. Mednick (Eds.), *Schizotypal personality*. New York: Cambridge University Press.

Claridge, G. S., & Hewitt, J. K. (1987). A biometrical study of schizotypy in normal population. *Personality and Individual Differences*, *8*, 303–312.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

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

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Eysenck, H. J., & Eysenck, S. B. (1976). *Manual of the eysenck personality inventory.* San Diego, CA: Educational and Industrial Testing Service.

Fine, C., Gardner, M., Craigie, J., & Gold, I. (2007). Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry*, *12*(1), 46–77.

Freeman, D., Garety, P., Kuipers, E., Colbert, S., Jolley, S., Fowler, D., *et al.* (2006). Delusions and decision-making style: Use of the need for closure scale. *Behaviour Research and Therapy*, *44*(8), 1147–1158.

Garety, P. A. (1991). Reasoning and delusions. *British Journal of Psychiatry*, *159*(14), 14–18.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Goldberg, D. P., & Williams, P. (1988). *A user's guide to the general health questionnaire*. London: NFER Nelson.

Heim, A. W., Watts, K. P., & Simmonds, V. (1975). *Group test of general intelligence*. Windsor: NFER.

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. *Quarterly Journal of Experimental Psychology: Human Learning and Memory*, *40A*, 801–812.

Jepson, C., Krantz, D. H., & Nisbett, R. E. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences*, *6*, 494–501.

John, C. H., & Dodgson, G. (1994). Inductive reasoning in delusional thinking. *Journal of Mental Health*, *3*, 31–49.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kemp, R., Chua, S., McKenna, P., & David, A. (1997). Reasoning and delusions. *British Journal of Psychiatry*, *170*, 398–405.

Larøi, F., Van der Linden, M., DeFruyt, F., van Os, J., & Aleman, A. (2006). Associations between delusion proneness and personality structure in non-clinical participants: Comparison between young and elderly samples. *Psychopathology*, *39*, 218–226.

Linney, Y. M., Peters, E. R., & Ayton, P. (1998). Reasoning biases in delusion-prone individuals. *British Journal of Clinical Psychology*, *37*, 285–302.

Mason, O., Claridge, G. S., & Jackson, M. (1995). New scales for the assessment of schizotypy. *Personality and Individual Differences*, *18*(1), 7–13.

McKay, R., Langdon, R., & Coltheart, M. (2006). Need for closure, jumping to conclusions, and decisiveness in delusion-prone individuals. *Journal of Nervous and Mental Disease*, *194*(6), 422–426.

Meehl, P. E. (1962). Schizotaxia, schizotypy and schizophrenia. *American Psychologist*, *17*, 827–833.

Meehl, P. E. (1990). Toward an integrated theory of schizotaxia, schizotypy and schizophrenia. *Journal of Personality Disorders*, *4*, 1–99.

Moritz, S., & Woodward, T. S. (2005). Jumping to conclusions in delusional and non-delusional schizophrenic patients. *British Journal of Clinical Psychology*, *44*, 193–207. doi:10.1348/014466505X35678

Mujica-Parodi, L. R., Malaspina, D., & Sackeim, H. A. (2000). Logical processing, affect, and delusional thought in schizophrenia. *Harvard Review of Psychiatry*, *8*, 73–83.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, *32*, 932–943.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Fong, G. T. (1982). Improving inductive inference. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 445–459). Cambridge: Cambridge University Press.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339–363.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Peters, E. R., Day, S., & Garety, P. A. (1996). The Peters *et al.* Delusions Inventory (PDI): New norms for the 21-item version. *Schizophrenia Research*, *18*, 118.

Peters, E. R., Joseph, S., Day, S., & Garety, P. A. (2004). Measuring delusional ideation: The 21-item Peters *et al.* Delusions Inventory (PDI). *Schizophrenia Bulletin*, *30*, 1005–1022.

Peters, E. R., Thornton, P., Siksou, L., Linney, Y., & MacCabe, J. H. (2008). Specificity of the 'jump-to-conclusions' bias in deluded patients. *British Journal of Clinical Psychology*, *47*, 239–244. doi:10.1348/014466507X255294

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision making*, *10*, 33–51.

van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological Medicine*, *39*(2), 179–195.

Warman, D. M., Lysaker, P. H., Martin, J. M., Davis, L., & Haudenschield, S. L. (2007). Jumping to conclusions and the continuum of delusional beliefs. *Behaviour Research and Therapy*, *45*, 1255–1269.

Young, H. F., & Bentall, R. P. (1997). Probabilistic reasoning difficulty in deluded, depressed and normal subjects: Effects of task difficulty and meaningful versus non-meaningful material. *Psychological Medicine*, *27*, 455–465.

Ziegler, M., Rief, W., Werner, S.-M., Mehl, S., & Lincoln, T. M. (2008). Hasty decision-making in a variety of tasks: Does it contribute to the development of delusions? *Psychology and Psychotherapy: Theory, Research and Practice*, *81*, 237–245. doi:10.1348/147608308X297104

Received 24 June 2008; revised version received 6 October 2009