Beads Task vs. Box Task: The specificity of the Jumping to Conclusions bias

Ryan P. Balzan[1*]

Rachel Ephraums[1]

Paul Delfabbro[2]

Christina Andreou[3]

[1]School of Psychology, Flinders University, South Australia, Australia

[2]School of Psychology, University of Adelaide, South Australia, Australia

[3]Center for Gender Research and Early Detection, University Psychiatric Clinics Basel, Switzerland

***Corresponding author:**
Dr Ryan Balzan
School of Psychology
Flinders University
GPO Box 2100
Adelaide SA Australia 5001
Email: drbalzan@gmail.com
Tel: +61 8 8201 3082

Highlights

- We investigated the Jumping to Conclusions (JTC) bias in high- and low-delusion-prone groups
- We used the ‘beads task’ and a conceptually similar ‘box task’ to assess JTC
- All participants requested significantly less information on the beads task relative to the box task.
- High-delusion-prone participants were significantly more *conservative* than low-delusion-prone group.

Abstract

Background and Objectives: Previous research involving the probabilistic reasoning ‘beads task’ has consistently demonstrated a *jumping-to-conclusions* (JTC) bias, where individuals with delusions make decisions based on limited evidence. However, recent studies have suggested that miscomprehension may be confounding the beads task. The current study aimed to test the conventional beads task against a conceptually simpler probabilistic reasoning "box task".

Methods: One hundred non-clinical participants completed both the beads task and the box task, and the Peters et al. Delusions Inventory (PDI) to assess for delusion-proneness. The number of ‘draws to decision’ was assessed for both tasks. Additionally, the total amount of on-screen evidence was manipulated for the box task, and two new box task measures were assessed (i.e., ‘proportion of evidence requested’ and ‘deviation from optimal solution’).

Results: Despite being conceptually similar, the two tasks did not correlate, and participants requested significantly less information on the beads task relative to the box task. High-delusion-prone participants did not demonstrate hastier decisions on either task; in fact, for box task, this group was observed to be significantly more *conservative* than low-delusion-prone group.

Limitations: Neither task was incentivised; results need replication with a clinical sample.

Conclusion: Participants, and particularly those identified as high-delusion-prone, displayed a more conservative style of responding on the novel box task, relative to the beads task. The two tasks, whilst conceptually similar, appear to be tapping different cognitive processes. The implications of these results are discussed in relation to the JTC bias and the theoretical mechanisms thought to underlie it.

Keywords: jumping-to-conclusions (JTC), beads task, box task, delusion-proneness

1. Introduction

Recent research has looked into the role that cognitive-reasoning biases play in the formation and maintenance of delusions. The jumping-to-conclusions (JTC) bias is the most extensively studied cognitive bias in this literature, and is defined as hasty decisions based on less evidence, when compared to people without delusions or low delusional ideation. The most common approach to elucidate the JTC bias is a probabilistic reasoning task called the ‘beads task’ (Huq, Garety, & Hemsley, 1988). In a typical beads task, participants are shown two transparent containers filled with colored beads in different but reciprocal proportions. Once the containers have been removed from the participants view, they are then told that beads, one at a time and with subsequent replacement, will be randomly drawn from one of the two containers and they need to decide which container the bead *sequence* is coming from (which is actually predetermined to favour one container). The most common finding is that participants with delusions will request fewer ‘draws to decision’ and exhibit higher rates of JTC (i.e., defined as a decision on 1 or 2 beads) compared to participants without delusions (for a recent meta-analysis see McLean, Mattiske, & Balzan, 2016). The bias has also been shown to be heightened among ‘delusion-prone’ but otherwise healthy samples (Ross, McKay, Coltheart, & Langdon, 2015).

Despite the apparent robustness of the beads task at elucidating the JTC bias, recent evidence suggests that the task may be confounded by poor task comprehension (Balzan, Delfabbro, & Galletly, 2012; Balzan, Delfabbro, Galletly, & Woodward, 2012). Specifically, both of these studies found that over half the sample failed to understand that all beads were drawn from the *same jar*, and that these ‘miscomprehending’ participants were significantly more likely to exhibit a JTC reasoning style. Against this high level of miscomprehension, it is possible that participants who ‘jump to conclusions’ on the *first* bead may have misinterpreted that the aim of the beads task is to determine which container this *particular* bead is coming from, rather than considering the *sequence* of beads (e.g., up to 10 beads, which these participants never see). Participants who make a decision on the first bead may therefore be simply answering the question ‘where does *this* bead come from?’, and assume that the initial bead has come from the container with the larger proportion of that color, despite being told from the outset that they should be considering the ‘bead sequence’ as a whole. Additionally, the beads task is typically only presented once or twice in an effort to reduce practice effects (i.e., becoming aware that the sequence is predetermined). In other words, it is possible that due to the small number of replications, the frequency of the bias reported in the literature may be overstated. In support of this, one recent study found that when a fMRI-adapted version of the beads task was presented multiple times to the participants, group differences in JTC among ‘at-risk mental state’ patients compared to healthy controls could not be demonstrated beyond the initial trial (Rausch et al., 2015). Of course, this is not to suggest that all previously reported JTC represents miscomprehension; this confound can be ruled out for participants who make a decision on the *second* bead (which is still classed as ‘JTC’), as this implies they are basing their decision on the bead *sequence*.

One of the aims of the current study is therefore to reduce this potential confound for those who JTC on the first bead, by using an alternative ‘draws to decision’ probabilistic reasoning task referred to as the 'box task'. The box task, adapted by the senior author from the Information Sampling Task of the CANTAB Battery (Clark, Robbins, Ersche, & Sahakian, 2006), is conceptually much simpler than the beads task. Participants are shown a number of grey boxes on screen (e.g., 25), each of which conceals one of two colors. Participants are told that one color is always in the majority (e.g., 80%), and that they must decide which color this is by clicking on as many of the grey boxes as they wish. Once they have decided which color is in the majority they are told to click on that color at the bottom of the screen. Importantly, the total amount of *potential evidence* is immediately available and obvious from the outset of the task (e.g., 25 boxes), reinforcing the notion that participants have the ability to consider a *sequence of evidence* should they choose to (as opposed to the beads task, where this is not as obvious from the outset). Therefore, should a participant make a decision on the first opened box, it is more likely to represent genuine JTC rather than a misunderstanding of the task's instructional set.

Moreover, the box task offers the opportunity to systematically manipulate factors affecting the salience of the evidence; for example, in addition to altering the ratio of evidence, as typically done in the beads task (e.g., 80/20 vs. 60/40 ratio), the box task can also vary the total amount of evidence that participants can choose from the outset of the task (e.g., 25 boxes, 49 boxes, 100 boxes). In this way, the box task can more effectively manipulate the *strength* of the evidence requested; that is, the salience of the evidence will be stronger when participants can choose from 25 on-screen boxes compared to when they have 49 boxes to choose from. This manipulation becomes useful when investigating the underlying mechanisms of the JTC bias. For example, several studies have posited that heightened JTC may be driven by a hypersalience of evidence-hypothesis matches, whereby people with delusions are more likely to make hasty decisions when faced with hypothesis-congruent evidence (e.g., Balzan, Delfabbro, Galletly, & Woodward, 2014; Speechley, Whitman, & Woodward, 2010). It follows that people with delusional tendencies should request *less evidence* when hypothesis-evidence matches are stronger, relative to non-delusional individuals.

Furthermore, the box task has an objective endpoint, which allows the requested 'proportion of evidence' to be determined in addition to just raw 'draws to decision'. This becomes important when comparing different on-screen box quantities (e.g., 25 with 49 boxes), as participants may request more raw 'draws to decision' with 49 boxes compared to

25 boxes, but this might simultaneously represent *proportionally less* evidence. Having an objectively correct number of 'draws' on the box task also allows a 'deviation from optimal solution' measure to be determined. For example, on an 80/20 ratio with 25 boxes, it would only take 6 draws of the same color to determine the solution with full confidence; the number of draws *below* this solution would represent *objective* JTC, while draws above this would represent *objective* conservatism.

In sum, due to some slight methodological differences, the box task may offer some advantages over the beads task (i.e., may reduce miscomprehension, easy to modify salience of evidence, has a definite solution, multiple trials). Despite these differences, the two tasks should still be consistent in the way they assess the 'draws to decision' construct. Accordingly, it is expected that participants who categorically JTC on the beads task (i.e., decision made ≤ 2 beads) will also request *less* evidence on the box task, consistent with a pilot study comparing the beads and box tasks (Chu, Sun, & So, 2015). Moreover, as JTC has been shown to be heightened in delusional groups, it is also expected that high-delusion-prone participants will request less evidence than the low-delusion-prone participants in both the beads and box task. Finally, throughout the box task, high-delusion-prone participants should request proportionally less evidence and exhibit objectively greater hastiness than low-delusion-prone participants, particularly for box scenarios where the salience of evidence is stronger.

2. Method

2.1. Participants

A total of 100 undergraduate students (76 females, 24 males) were recruited from Flinders University for partial course requirement or a small amount of reimbursement for their time. They were aged 18 – 62 years (*M* = 23.56, *SD* = 7.30). Participants were excluded if they had impaired color vision (relevant to both bead/box tasks).

2.2. Materials

2.2.1. The beads task. Participants were presented with an adapted computerized version of the original beads task (Huq et al., 1988), using the 'draws to decision' method and the standard conventional instructional set (for full details see McLean et al., 2016). They were shown a picture of two containers filled with 100 colored beads in reciprocal proportions (one trial with a bead ratio of 80/20 and one trial of ratio 60/40; ratio order was randomised between participants). They were told that the computer would randomly select beads from one container, and that the goal of the task was to determine which container the bead sequence was coming from. However, the task had a predetermined sequence of (up to) ten beads per trial, and ended once a container had been selected (note: if a decision was not made by 10 beads, the 'draws to decision' score was classified as 11). Pictures of the containers remained displayed during the task to ensure that participants remembered the proportion of beads in each container, and the sequence of beads was also shown at the top of the screen as it emerged. Participants were shown an example in an effort to reduce miscomprehension.

2.2.2. The box task. The box task was conceptually similar to the beads task. A number of grey squares ('boxes') were displayed on a screen (i.e., 25 boxes or 49 boxes), with each box concealing one of two colors, as shown at the bottom of the screen in two larger rectangular boxes (see Figure 1). Boxes were presented in certain ratios (e.g., ten trials of 80/20; ten trials of 60/40). The ratio for each particular trial was displayed in the top left of the screen. Participants were told that the goal of the task was to determine which color was represented the majority of the time (i.e., 80% or 60%), referred to as the 'dominant color'.

At the beginning of the trial all of the boxes were shown in grey (i.e., closed). Participants were required to mouse-click these grey boxes to reveal or 'open' the colour of the box, and were able to open as many as they wanted, in any order, before making a decision which was the dominant color by clicking on one of the two larger rectangles at the bottom of the screen. In this way, the box task was not pre-sequenced and participants thereby saw a random sequence of colors (unlike the beads task, which used a predetermined sequence). Participants were given two practice trials of 25 boxes to reduce miscomprehension (one per ratio). The number of boxes presented was manipulated to change the salience of the available evidence. There were 20 trials overall; 10 trials of 80/20 ratio (5 with 25 boxes; 5 with 49 boxes) and 10 trials of 60/40 ratio (5 with 25 boxes; 5 with 49 boxes).

2.2.3. The Peters et al. Delusional Inventory (21 Item). The Peters et al. Delusional Inventory (Peters, Joseph, Day, & Garety, 2004) is designed to assess delusion-proneness in the general population. The 21-item PDI has good psychometric properties, and had high internal consistency within the present sample (Cronbach's alpha = .93). The measure includes four components: a yes/no score, and measures of distress, preoccupation and conviction (each ranged from 0 to 5). Higher scores indicate increased tendency for delusion-proneness. Upper and lower quartiles were taken to classify participants into low-delusion-prone (scores ≤ 27; $n = 28$) and high-delusion-prone (scores ≥ 84; $n = 23$).

2.2.4. Wechsler Test of Adult Reading. To rule out the possibility delusion-prone and non-delusion-prone groups are different in regards to their intellectual functioning, both groups completed the Wechsler Test of Adult Reading (WTAR) (Wechsler, 2001), which is co-normed with the Wechsler Adult Intelligence and Memory Scales (WAIS-III). There were no significant differences in predicted full-scale intelligence scores between the low- (M = 106.79, SD = 8.07) and high- (M = 103.43, SD = 7.19) delusion-prone participants, $t(49) = 1.55$, $p = .128$.

2.2.5. Demographic information. Age, gender, and level of education were recorded. All participants were asked ‘which task did you prefer?’, which served to gauge the subjective feasibility of the box task.

2.3. Measures

2.3.1. Draws to decision. The number of draws to reach a definite decision on both the beads and box task.

2.3.2. Proportion of evidence requested. The raw number of ‘draws to decision’ for the box task was converted into a percentage to determine the proportion of evidence requested before reaching a definite decision (e.g., 10 boxes out of 25 is 40% of the evidence). This measure could not be calculated for the beads task because participants were unaware the task would terminate after 10 beads and beads were replaced after each trial.

2.3.3. Deviation from optimal solution. For the box task, a ‘deviation for optimal solution’ measure was created by first assessing the number of ‘dominant color’ boxes selected by participants. The objective optimal number of ‘dominant color’ boxes was then determined for each condition[1], and this was subtracted from the number of dominant color boxes selected by the participant. Positive numbers indicated that participants were demonstrating a conservative response style (i.e., requesting *more* evidence than required), while negative numbers indicated that participants were objectively jumping to conclusions (i.e., requesting *less* evidence than required). This variable could not be created for the beads task as beads were replaced after each trial (i.e., no objective ‘solution’).

2.3.4. Independent variables. The study used a mixed design. Within-subjects independent variables were task type (beads, box); ratio type (80/20, 60/40); and box quantity

[1] Optimal number of dominant color boxes: 80/20 ratio-25 boxes = 6 boxes; 80/20 ratio-49 boxes = 11 boxes; 60/40 ratio-25 boxes = 11 boxes; 60/40 ratio-49 boxes = 21 boxes.

(25 boxes, 49 boxes) The between-subjects independent variable was delusion proneness (low- and high-delusion-prone).

2.4. Procedure

The experiment took place in the Cognitive Neuropsychiatry Laboratory in the School of Psychology at Flinders University. Each session took approximately 20 minutes. After providing informed consent, participants completed the computer tasks, followed by the demographics questionnaire, the PDI, and the WTAR. The beads task was created using the program, SuperLab 4.5. The box task was adapted by the senior author at University of Hamburg, using Microsoft Visual Studio. The order that the box and the beads task were presented was counterbalanced, with equal numbers in each task type. The study was approved by the Social and Behavioural Science Research Ethics Committee (Flinders University).

3. Results

3.1. Beads vs. Box

3.1.1. Draws to decision. Beads and box 'draws to decision' totals were calculated by averaging the number of 'draws to decisions' across each ratio type (i.e., 80/20; 60/40). Participants requested significantly less evidence on the beads task than the box task, $t(99) = 17.47$, $p < .001$, $d = 2.55$. As shown in Table 1, when breaking down both tasks to ratio type, participants requested significantly less evidence on both the 80/20, $t(99) = 15.47$, $p < .001$, $d = 2.23$ and 60/40, $t(99) = 17.54$, $p < .001$, $d = 2.54$, versions of the beads task relative to the box task.

(Table 1 about here)

To ensure that the lower draws to decisions score on the beads task was not simply due to the beads task terminating after the presentation of 10 beads, a McNemar test was conducted. Participants were significantly less likely to request more than 10 pieces of evidence on the 80/20 ratio for the beads task (2%) compared to the box task (64%), $p < .001$. Similarly on the 60/40 ratio only 33% of participants on the beads task requested more than 10 pieces of evidence compared to 96% on the box task, ($p < .001$). Thus, despite the beads task not having a 'definite end point', these results indicate that the majority of participants would have made their decision *before* 10 beads even if the task had not terminated at this point.

Correlational analyses showed that there was no significant association between the draws to decision scores on either version of the beads task and the box tasks (see Table 2); although there was a trend towards a significant negative correlation between 60/40 beads task and the 80/20 box task ($r = -.188$, $p = .06$). Overall, this supports the idea that different factors might be affecting performance patterns on each task.

3.1.2. Jumping to conclusions. Independent samples *t*-tests assessed group differences on the box task (i.e., draws to decision, proportion of evidence requested and deviation from optimal) between participants who categorically jumped to conclusions (i.e., asked ≤ 2 beads) and those who requested more information during the beads task. Of note, the majority of these 'JTC participants' made their decision on the *first bead* ($n = 19/30$), and may have therefore miscomprehended the task; in contrast, no participant made a decision on the first trial of the box task. As shown in Table 3, JTC participants selected significantly *more* evidence and were objectively *more conservative* on the box task relative to those participants who did not demonstrate JTC during the beads task ($t(98) = 2.50, 2.47, 2.49$, $p < .05$, for box draws to decision, proportion of evidence, deviation from optimal measures, respectively[2]). This is also consistent the negative correlation trend between the 60/40 beads task and the 80/20 box task (i.e., less evidence for beads implies more evidence for box). This again suggests that participants may be using different strategies when completing each task (i.e., if JTC on beads, should request *less* evidence on the conceptually similar box task).

(Table 3 about here)

3.1.3. Preference. To gauge the subjective feasibility of the box task, participants were asked whether they preferred the beads task or the box task. Participants were more likely to prefer the box task (76%) over the beads task (24%), despite the fact that the box task had 20 trials (beads task only had 2). This preference for the longer box task suggests that it may be more intuitive and easier to comprehend relative to the beads task.

3.2. Delusion-proneness

3.2.1. Draws to Decision. Total Peters et al. Delusional Inventory (PDI) scores across the sample were correlated with the total number of draws to decision on the beads and the box task. A non-significant relationship was found between both tasks and the total PDI score (beads task: r = -.082, p = .424; box task: r = .161, p = .114); no PDI subscale score (i.e., distress, preoccupation, conviction) was significant either. This suggests that delusion-proneness, as assessed by the PDI, was not related to draws to decision in the current sample.

To test whether high-delusion-prone participants would request less evidence than the low-delusion-prone participants in both tasks, upper and lower quartiles of the PDI were taken to classify participants into low-delusion-prone (scores ≤ 27; n = 28) or high-delusion-prone (scores ≥ 84; n = 23). There were no group differences between low- and high-delusion-proneness in age (low: M = 24.43, SD = 9.27; high: M = 22.61, SD = 5.53; $t(49)$ <

[2] These analyses were repeated for the subset of participants who showed JTC on the first bead with similar results (i.e., $t(98)$ = 2.50, 2.47, 2.49, p < .05, for box draws to decision, proportion of evidence, deviation from optimal measures, respectively)

1); gender (low: female/male = 20/8; high: female/male = 18/5; χ^2 (1, *N* = 51) < 1); or years of education (low: *M* = 15.00, *SD* = 2.21; high: *M* = 14.09, *SD* = 2.57; *t*(49) = 1.36, *p* = 0.18).

An independent samples *t*-test was conducted to assess whether high- and low-delusion prone participants performed differently on the beads and box task. As seen in Table 4, there were no significant group differences for delusion-proneness on the beads task, *t*(49) < 1, *d* = .12 or the box task, *t*(49) = 1.49, p = .143, *d* = .42.

(Table 4 about here)

For the beads task, high-delusion-prone participants were requesting less evidence relative to low-delusion-prone participants which, although non-significant, was in the direction consistent with previous literature. Conversely on the box task, there was a trend for high-delusion-prone participants to request *more* evidence compared to low-delusion-prone participants, which is inconsistent with previous JTC research.

3.2.2. Proportion of Evidence Requested. To test the hypothesis that high-delusion-prone participants would request *proportionally* less evidence when evidence for the correct solution was more salient (i.e., 80/20 with 25 boxes vs. 60/40 with 49 boxes), a 2 Ratio (80/20, 60/40) x 2 Box quantity (25, 49) x 2 Delusion-proneness (high, low) mixed between-within subjects ANOVA was conducted. Table 5 summarises the proportion of evidence requested by ratio, box quantity and delusion-proneness, and suggests both groups were requesting *less* evidence for 49 than 25 boxes (*F*(1,49) = 60.06, *p* < .01, η_p^2 = .55), which is in contrast to the predicted direction. The 3-way interaction between these factors was non-significant, *F*(1, 49) < 1. There was a significant 2-way interaction between box quantity and delusion-proneness, *F*(1,49) = 4.10, *p* = .048, η_p^2 = .08, suggesting that the *low-delusion-prone* group were requesting relatively less evidence for 49-box condition compared to the high-delusion-prone-group (see Figure 2). Although pairwise comparisons between high- (M = 49.05, SD = 18.52) and low-delusion-prone groups (M = 39.99, SD = 17.87) suggest this difference for the 49-box condition overall was non-significant, $F(1,49) = 3.15$, $p = .08$, $\eta_p^2 = .06$, a pairwise comparison between groups for the 49-box condition under the 80/20 ratio revealed that the low-delusion-prone group was requesting significantly less evidence, $F(1,49) = 6.30$, $p = .02$, $\eta_p^2 = .11$ (see Table 5). There was also a significant correlation between delusion-proneness (total PDI across the sample) and the 49-box condition under the 80/20 ratio, $r = .24$, $p = .02$ (but not for the PDI subscales). These findings appear to be in contrast to previous JTC literature (i.e., low-delusion-prone group requesting *less* evidence) and the ‘hypersalience’ hypothesis (i.e., the high-delusion-prone group requested *less* evidence when each piece of evidence was ‘weaker’ in the 49-box condition). No other 2-way interaction was significant.

(Table 5 about here)

(Figure 2 about here)

3.2.3. Deviation from Optimal. A 2 Ratio (80/20, 60/40) x 2 Box quantity (25, 49) x 2 Delusion-proneness (high, low) mixed between-within subjects ANOVA was conducted to test the hypothesis that high-delusion-prone participants would deviate further (< 0) from the optimal solution when evidence was stronger (i.e., 80/20 with 25 boxes vs. 60/40 with 49 boxes). Mean deviation from optimal scores for both groups across these conditions are summarised in Table 6. The 3-way interaction was significant, $F(1,49) = 4.17$, $p = .047$, $\eta_p^2 = .08$. Figure 2 shows the 2-way interactions for box quantity and delusion-proneness separately by ratio (i.e., 80/20, 60/40).

For the 80/20 ratio, the 2-way interaction was significant $F(1,49) = 10.79$, $p = .002$, $\eta_p^2 = .18$, and suggested that the low-delusion-prone group shifted from a conservative to a more optimal response strategy as the number of on-screen boxes increased, whereas the high-delusion-prone group became even more conservative as the number of boxes increased (Figure 3A). The pairwise comparison between groups for the 49-box condition under the 80/20 ratio revealed that the high-delusion-prone group was significantly more conservative, $F(1,49) = 6.26$, $p = .02$, $\eta_p^2 = .11$ (see Table 6) and there was a significant correlation between delusion-proneness (across the sample) and the 49-box condition under the 80/20 ratio, ($r = .24$, $p = .02$), consistent with the ‘proportion of evidence requested’ findings above. These deviation from optimal results for the 80/20 ratio condition suggest that the box task is inconsistent with previous JTC findings in both a relative and an *absolute* sense (i.e., high-delusion-prone individuals are relatively and *objectively* more conservative than low-delusion-prone individuals); however, unlike the proportion of evidence findings, they were somewhat consistent with the hypersalience mechanism (i.e., high-delusion-prone individuals became more conservative with weaker evidence).

The results for the 60/40 condition were vastly different, in that both groups adopted a near-optimal strategy for 25 boxes, yet became very biased (i.e., < 0) in the 49 box condition (Figure 3B). Again, the low-delusion-prone group appeared to be more biased, but the 2-way box quantity and group interaction for the 60/40 ratio was non-significant $F(1,49) = 1.65$, $p = .21$, $\eta_p^2 = .03$.

(Table 6 about here)

(Figure 3A/B about here)

4. Discussion

The present study aimed to investigate the jumping to conclusions (JTC) bias across two conceptually similar probabilistic reasoning tasks. This study represents one of the first attempts at using the novel ‘box task’ to assess JTC.

4.1. Beads vs. Box

The findings clearly show differences between the tasks, whereby participants requested significantly fewer ‘draws to decision’ on the beads task compared to the box task. This is despite (i) the beads task having no definite ‘endpoint’ or solution (i.e., participants were unaware that the task would terminate after 10 beads); (ii) that beads were replaced after each trial (i.e., which should increase sampling relative to a condition where evidence is drawn without replacement, like the box task); and (iii) that the evidence is *weaker* for the beads task (i.e., 1 bead out of 100 total beads vs. 1 box out of 25 or 49 boxes). Accordingly, participants should have been more *conservative* on the beads task.

It is also clear from the current findings that participants showed different patterns of responding for both tasks. Specifically, participants who displayed JTC on the beads task (i.e., decision ≤ 2 beads) requested significantly *more* information on the box task and were objectively *more conservative* relative to non-JTC participants. Moreover, unlike a recent pilot study (Chu et al., 2015), the two tasks did not significantly correlate with one another (although there was trend towards a negative correlation between the 60/40 beads and 80/20 box tasks). These findings suggest that the beads and the box task may be assessing different cognitive mechanisms that influence the way participants respond on conceptually similar tasks.

The fact that beads task terminated after 10 beads cannot account for the hastier decision-making style, as the majority of participants decided before even reaching 10 beads. There are a few reasons could that account for these differences. First, the hastier decision-making on the beads task may actually reflect miscomprehension, at least for the 19 participants who made their decision on the first bead, and may have therefore erroneously assumed that the task was to judge where that first bead came from (rather than the bead sequence). On the box, it is much easier for participants to comprehend that there is a *sequence* of evidence to consider, as participants are presented with all the potential evidence from the outset, thereby reducing the likelihood that any miscomprehension could result. The box task is also conceptually simpler, and hypothetically easier to understand, which was indirectly supported by the observation that participants had a greater preference for the box task over the beads task (despite its longer length).

Nevertheless, this explanation does not seem to account for why the remaining 81 participants, who made decisions at two or more beads, still selected *less evidence* on the beads task compared to the box task. However, it could be the very fact that the box task makes it easier for participants to comprehend the nature of task (i.e., to consider a *sequence* of evidence) that encourages them into a ‘conservative’ responding style. For example, although it should only take 6 draws of the same color to determine the solution with full confidence on an 80/20 ratio with 25 boxes, this could leave up to 19 ‘unopened’ boxes onscreen, which might prompt doubt or uncertainty (‘second-guessing’). By this rationale, it is possible that if the beads task displayed the total number of drawn beads on the screen and allowed participants to self-select evidence from this line of ‘unknown’ beads (i.e., like the box task), the results may have be similar; future comparisons should test this proposition.

Importantly, the tasks also differ with regard to the randomness of the sequence (i.e., the color sequence is predetermined for the beads task but is random for the box task). The randomised color sequence in the box task implies that participants may not have been seeing the ‘dominant’ box color until later in the sequence, relative to the beads task where the ‘dominant’ color was obvious from the beginning of the sequence. While this is unlikely to be the main reason participants requested almost *four times* more information on the box task, future box task studies should ensure a predetermined color sequence between participants.

4.2. Delusion-proneness

The hypothesis that high-delusion-prone participants would request less evidence than the low-delusion-prone participants in both the beads and box task was unsupported. For the beads task, there was a small trend that high-delusion-prone individuals requested less evidence, yet this was non-significant. These results can be compared to those from a recent meta-analysis on delusional ideation and JTC, which reported a small but significant negative correlation between 'draws to decision' and delusion-proneness (r = -.10) with high-delusion-prone participants typically found to request less evidence (Ross et al., 2015). However, given that this meta-analysis only reported a small effect size from a pooled sample of 1754 participants from the general population, it is possible that the current results lacked sufficient power to detect any differences between high- and low-delusion-proneness.

The box task allowed for greater exploration of the relationship between JTC and delusion-proneness because the total evidence participants could request could be manipulated (i.e., 25 or 49 boxes), and it also included two additional measures of hasty decision-making (i.e., proportion of evidence requested; deviation from optimal solution). Specifically, it was predicted that high-delusion-prone participants would request proportionally less evidence and exhibit objectively greater JTC (i.e., greater negative deviation from the optimal solution), particularly for box scenarios where the salience of evidence was stronger (i.e., 80/20 ratio with 25 boxes). However, the results from the box task were inconsistent with these predictions, in that the high-delusion-prone group was more *conservative* than the low-delusion-prone group, whether observing 'draws to decision', 'proportion of evidence requested', or the 'deviation from optimal' measures. The high/low-delusion-proneness pairwise comparisons were for the most part non-significant across conditions. However, for the 80/20 ratio with 49 boxes, it was evident that the high-delusion-prone group were significantly more conservative, as they requested proportionally *more* evidence and *conservatively* deviated from the optimal solution (in contrast, the low-delusion-prone group approached the optimal solution for this condition). This directly contradicts the established set of findings within the JTC literature, in which high-delusional groups are usually found to request less evidence (e.g., Fine, Gardner, Craigie, & Gold, 2007; McLean et al., 2016). The hypersalience mechanism was also largely unsupported by the current findings, as both groups appeared to request less evidence when the salience of the evidence became weaker; although for the 80/20 ratio condition, the high-delusion-prone group did exhibit greater conservatism for the 49-box condition (i.e., greater conservatism for 'weak' evidence).

Given the current study represents one of the first to employ the box task within a probabilistic reasoning context, we can only speculate as to the causes of these discrepant findings. Overall, it is appears that all participants, irrespective of delusion-proneness, were more conservative on the box task relative to the beads task. We have argued that miscomprehension, as well as differences in the way evidence was presented may partially account for these differences (i.e., all the potential evidence is available from the beginning; predetermined vs. random sequences). However, when considering ratio type and box quantity variables, it becomes apparent that there may be an additional factor contributing to the more conservative approach observed for the box task. It is clear from the results that all participants demonstrated a hastier decision-making strategy when they were presented with 49-boxes. This suggests that the subjective 'cost' of responding for more evidence in the 49-box condition was higher than in the 25-box condition, due to the extra time and effort it would require to request the same proportion of boxes. Arguably, decision-making would become relatively hastier with more available evidence to choose from (e.g., 100 boxes, 200 boxes, etc). Although each piece of evidence is technically 'weaker' under such conditions (i.e., 1 of 25 'stronger' than 1 of 49), the threshold to keep requesting more evidence becomes higher as more effort/time is required. This could account for why the hypersalience mechanism was not supported (i.e., subjective 'cost' was too great as the number of boxes increased). Importantly, it appears the influence of this 'cost' factor differed depending on group and the ratio of evidence. For example, the high-delusion-prone group, in contrast to the low-delusion-prone-group, became surprisingly more conservative as the number of boxes increased under the 80:20 ratio (i.e., the cost of requesting more information was not strong enough to alert them to stop collecting superfluous evidence, as it was under 60/40 ratio). Assuming the hypersalience mechanism is heightened within the high-delusion-prone group, they may have 'enjoyed' collecting more evidence under the 49-box 80:20 ratio condition, as most of the self-selected boxes would have 'matched' the dominant color (not necessarily the case under a 60:40 ratio, with more 'misses').

At present, there is no clear explanation for why the heightened 'conservatism' might have occurred (i.e., the higher salience of evidence under the 80/20 ratio should have made the subjective cost of requesting more evidence *higher* not lower). Nevertheless, future replications could look at further increasing the 'cost' of responding under different ratios by adding a monetary incentive to the task or using a points system, similar to a recently incentivized version of the beads task (van der Leer, Hartig, Goldmanis, & McKay, 2014). Future studies should also aim to extend these findings to individuals with clinical delusions, to assess how delusional severity affects the interaction between salience of the evidence and cost of responding on measures of hasty decision-making. Finally, future replications comparing both tasks could utilize the "graded estimates" method, where participants provide probability estimates (e.g., 0 to 100%) that a particular bead (or box) is from one of the two containers (or is the dominant color); this additional measure could provide more specificity into how participants are comparatively viewing the evidence on each task.

4.3. Conclusions

The present study found that all participants were much more conservative on the box task relative to the beads task. It is possible that the hastier decision-making observed for the beads task is partially due to miscomprehension of the more complex instructional set of the task and the subtle methodological differences between the two conceptually similar tasks. Hasty decision-making was not related to delusion-proneness for either task; for the box task, high-delusion-proneness was more associated with heightened conservatism. Nevertheless, the box task may be an interesting platform by which to further investigate probabilistic reasoning, and more research is required before we can jump to any conclusions about its suitability in assessing JTC in delusional groups.

Declaration of interest

No authors do not declare any conflicts of interest.

Acknowledgements

The authors sincerely wish to thank Prof Steffen Moritz and the anonymous reviewers for their insightful comments and suggestions on earlier versions of this manuscript.

References

Balzan, R. P., Delfabbro, P. H., & Galletly, C. A. (2012). Delusion-proneness or miscomprehension? A re-examination of the jumping-to-conclusions bias. *Australian Journal of Psychology, 64*(2), 100-107. doi:10.1111/j.1742-9536.2011.00032.x

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

Balzan, R. P., Delfabbro, P. H., Galletly, C. A., & Woodward, T. S. (2014). Metacognitive training for patients with schizophrenia: Preliminary evidence for a targeted, single-module programme. *Australian and New Zealand Journal of Psychiatry, 48*, 1126-1136. doi:10.1177/0004867413508451

Chu, H., Sun, X., & So, S. (2015). The Beads, the Fish and the Box: Interrelationship between “jumping to conclusions” tasks and their links with cognitive abilities (Abstracts of the 23rd European Congress of Psychiatry). *European Psychiatry, 30*, 1706. doi:http://dx.doi.org/10.1016/S0924-9338(15)31309-2

Clark, L., Robbins, T. W., Ersche, K. D., & Sahakian, B. J. (2006). Reflection impulsivity in current and former substance users. *Biological Psychiatry, 60*(5), 515-522. doi:10.1016/j.biopsych.2005.11.007

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*, 46-77. doi:10.1080/13546800600750597

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

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-analytic approach. *Schizophrenia Bulletin*. doi:10.1093/schbul/sbw056

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

Rausch, F., Mier, D., Eifler, S., Fenske, S., Schirmbeck, F., Englisch, S., . . . Zink, M. (2015). Reduced activation in the ventral striatum during probabilistic decision-making in patients in an at-risk mental state. *Journal of Psychiatry and Neuroscience, 40*, 163-173.

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

Speechley, W., Whitman, J., & Woodward, T. S. (2010). The contribution of hypersalience to the “jumping to conclusions” bias associated with delusions in schizophrenia. *Journal of Psychiatry and Neuroscience*, *35*, 7-17.

van der Leer, L., Hartig, B., Goldmanis, M., & McKay, R. (2014). Delusion-proneness and 'jumping to conclusions': relative and absolute effects. *Psychological Medicine*. doi:10.1017/S0033291714002359

Wechsler, D. (2001). *Wechsler Test of Adult Reading*. San Antonio, TX: The Psychological Corporation.

Table 1: Mean (SD) draws to decision across condition on both the beads and the box tasks

	Task Type	
Ratio Type	*Beads Task*	*Box Task*
*80/20**	3.51 (2.06)	13.53 (6.02)
*60/40**	6.55 (3.59)	21.88 (7.73)
Total*	5.03 (2.57)	17.70 (6.55)

*$p < .001$

Table 2: Pearson correlations (p-values) between 80/20 and 60/40 versions of the beads and box tasks (n = 100)

	Beads Task	
Box Task	*80/20*	*60/40*
80/20	-.062 (.54)	-.188 (.06)
60/40	.017 (.86)	-.067 (.51)

Table 3: Mean (SD) draws to decision, proportion of evidence requested (%), and deviation from optimal[1] scores for the box task (averaged across conditions) by JTC (decision on ≤ 2 beads and on first bead) on the beads task.

	JTC on Beads	
Box Task Measure	*JTC ≤ 2 beads (n = 30)*	*No JTC (n = 70)*
*Draws to decision**	20.14 (7.07)	16.66 (6.07)
*Proportion of evidence**	55.90 (18.43)	46.84 (15.99)
*Deviation from optimal**	1.48 (4.92)	-0.90 (4.13)
	JTC on first bead (n = 19)	*No JTC (n = 81)*
*Draws to decision**	20.69 (6.77)	17.00 (6.34)
*Proportion of evidence**	57.35 (17.58)	47.37 (16.67)
*Deviation from optimal**	1.91 (4.72)	-0.68 (4.32)

[1]Positive deviation from optimal scores indicate an objective conservative response style (i.e., requesting *more* evidence than required); negative scores indicate objective jumping to conclusions (i.e., requesting *less* evidence than required)

*$p < .05$

Table 4: Mean (SD) draws to decision scores for low- and high-delusion-prone participants across the beads and box task (averaged across 80/20 and 60/40 conditions).

	Task Type	
Delusion-Proneness	*Beads Task*	*Box Task*
Low	4.98 (2.40)	16.67 (6.19)
High	4.67 (2.72)	19.25 (6.09)

Table 5: Mean (SD) proportion of evidence requested (%) for low- and high-delusion-prone participants across the box task.

		Delusion-proneness	
Ratio Type	Box Quantity	Low (*N* = 28)	High (*N* = 23)
80/20	25	42.03 (18.27)	46.12 (14.90)
	49	28.40 (14.87)*	39.57 (16.91)
60/40	25	67.97 (19.06)	69.57 (16.76)
	49	51.57 (22.61)	58.53 (22.62)

*p = .02, between low- and high-delusion-proneness

Table 6: Mean deviation from optimal scores[1] (SD) for low-and-high-delusion-prone participants across the box task.

		Delusion-proneness	
Ratio Type	Box Quantity	Low (*N* = 28)	High (*N* = 23)
80/20	25	2.82 (3.48)	3.15 (2.90)
	49	0.22 (5.79)*	4.58 (6.66)
60/40	25	-0.66 (2.82)	-0.04 (2.43)
	49	-6.01 (6.71)	-3.63 (6.58)

[1]Positive deviation from optimal scores indicate an objective conservative response style (i.e., requesting *more* evidence than required); negative scores indicate objective jumping to conclusions (i.e., requesting *less* evidence than required)

*p = .02, between low- and high-delusion-proneness

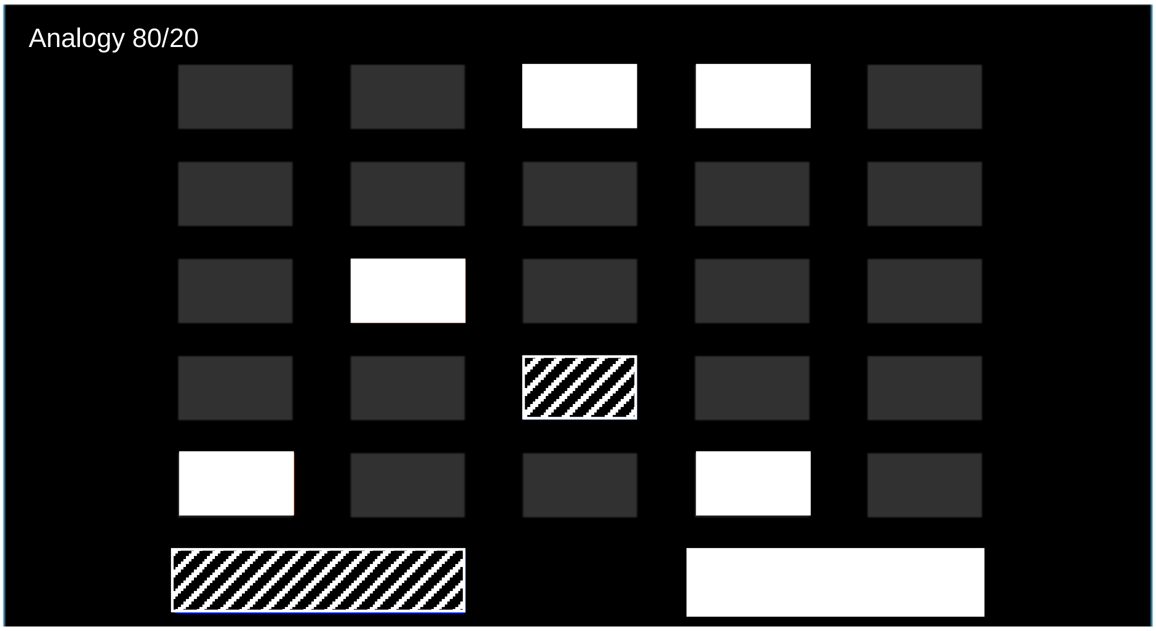

Figure 1: A screenshot of the box task with 25 boxes in the 80/20 ratio. Participants had ‘opened’ 6 boxes (modified for black and white)

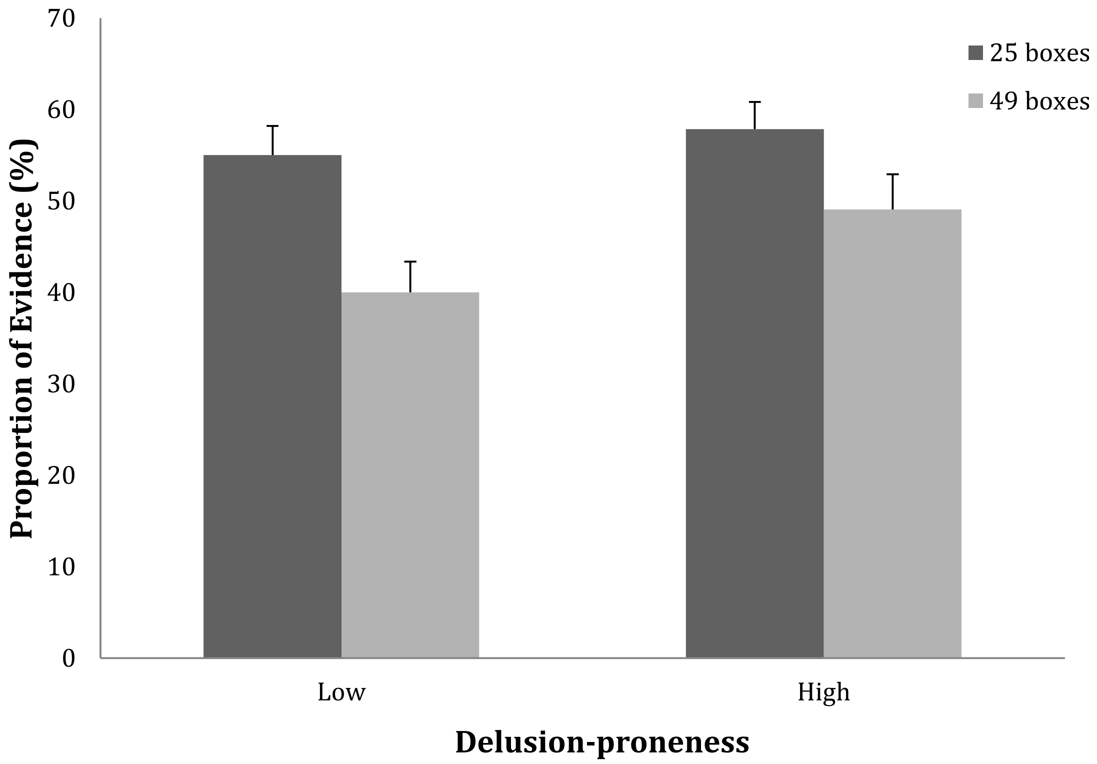

Figure 2: Proportion (%) of evidence (with standard error bars) requested for low-and-high-delusion-prone participants across the box task (averaged across 80/20 and 60/40 ratio conditions).

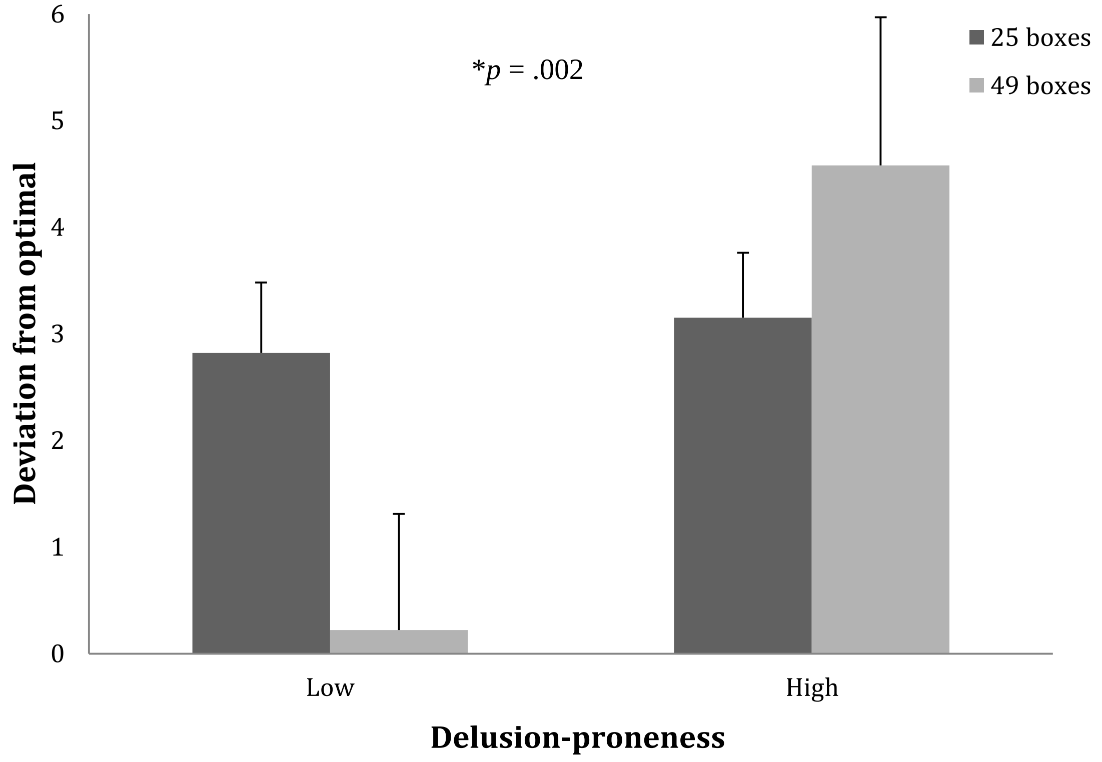

Figure 3A: Deviation from optimal scores (with standard error bars) for low- and-high-delusion-prone participants across 25 and 49 box task conditions for the 80/20 ratio (> 0 indicates conservatism)

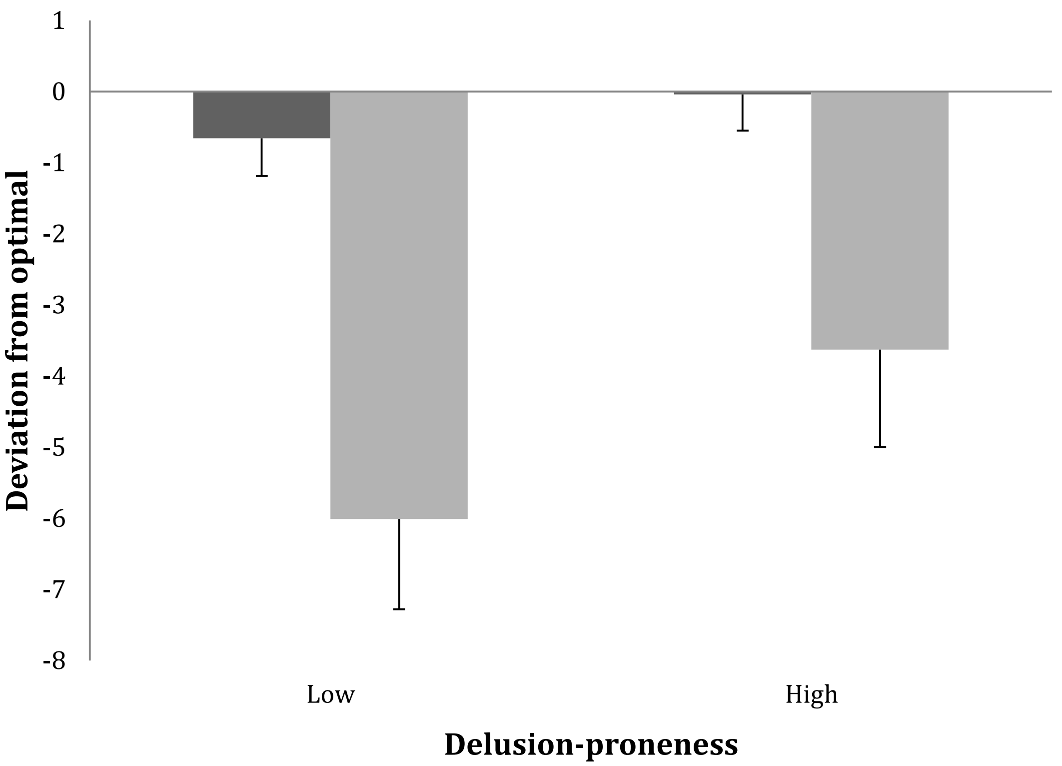

Figure 3B: Deviation from optimal scores (with standard error bars) for low-and-high-delusion-prone participants across 25 and 49 box task conditions for the 60/40 ratio (< 0 indicates JTC)