Prolonged rather than hasty decision-making in schizophrenia using the box task. Must we rethink the jumping to conclusions account of paranoia?

Steffen Moritz [a,*], Jakob Scheunemann [a], Thies Lüdtke [a,b], Stefan Westermann [a], Gerit Pfuhl [b], Ryan P. Balzan [c,e,1], Christina Andreou [d,1]

[a] *Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[b] *Department of Psychology, UiT - The Arctic University of Norway, Tromsø, Norway*

[c] *College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia*

[d] *Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany*

[e] *ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia*

Abstract

Jumping to conclusions (JTC) is the best established cognitive bias in schizophrenia and is increasingly targeted in interventions aimed to improve positive symptoms. To address shortcomings of the standard measure to capture JTC, the beads task, we developed a new variant—the box task—which was subsequently validated in people with elevated psychotic-like experiences. For the first time, the box task was administered in a sample of individuals with manifest schizophrenia. We hypothesized that patients with schizophrenia would display an elevated JTC bias relative to controls.

Method: We recruited a large sample of 101 patients with schizophrenia and matched them to an online sample recruited from the general population. In the box task, participants must decide which of two kinds of colored balls are presented more often. Participants are told that the task may end prematurely, and that task performance will be counted as an error if no decision had been made before that point. The primary measure was the number of draws to decision (DTD), where fewer DTD corresponds to greater JTC.

Results: In contrast to expectations, participants with schizophrenia showed significantly higher DTD (i.e., reduced JTC). Consistent with our previous findings, patients also displayed a lowered decision threshold compared to controls. Response confidence for the final decision was lower in patients and correlated with self-esteem and positive symptoms. While there was evidence that previous knowledge of the box task lowered DTD, exclusion of participants with experience on the box task did not substantially change results.

Discussion: The study fits a growing body of experiments casting doubt on the generalizability of the JTC effect in schizophrenia across different tasks. While the study tentatively supports a liberal acceptance account of psychosis, caution is warranted and we recommend that research should explore and control for potentially important mediators (e.g., task difficulty, stress, test-taking attitudes).

1. Introduction

1.1. Reasoning in schizophrenia

Cognitive biases—for example, deviances in attributional style (Kinderman and Bentall, 1996; Moritz et al., 2018), overconfidence in errors (for overviews see Hoven et al., 2019; Moritz and Woodward, 2006a), and a bias against disconfirmatory evidence (BADE) (McLean et al., 2017; Sanford et al., 2014; Woodward et al., 2008)—are implicated in the formation and maintenance of positive symptoms in schizophrenia. While the idea that reasoning/logical thinking is impaired in schizophrenia is quite old (see Jaspers, 1913, p. 97), systematic research into processes of reasoning did not begin in earnest until 30 years ago. Recent meta-analyses (Dudley et al., 2016; Fine et al., 2007; McLean et al., 2017; Ross et al., 2015) suggest that patients with schizophrenia, as well as participants scoring high on measures of schizotypy/psychosis proneness, gather less information before making a judgement. This response pattern has been termed the jumping to conclusions (JTC) bias, which is usually defined as a decision after gathering only one or two pieces of information. JTC is widely regarded an important factor

in the pathogenesis of delusions specifically, although its correlational association to JTC is at best weak (Dudley et al., 2016; Ross et al., 2015).

1.2. Experimental paradigms to capture jumping to conclusions

The tasks used to capture JTC are usually neutral in content and thus unlikely reflect a mere consequence of acute paranoia (e.g., black car in front of the house is taken as evidence for surveillance). Traditionally, JTC is measured by the beads task that requires participants to decide from which of two containers a sequence of beads has been drawn; typically the sequence favors one of the two containers (Garety et al., 1991). JTC is indicated by relatively fewer draws to decision (DTD). However, the original task faces several limitations. In particular, the task is often administered as a single trial thereby decreasing reliability, experience with the task likely impacts subsequent performance (Peters and Garety, 2006) as the outcome is often foreseeable and its parameters are also influenced by cognitive deficits.

Some researchers have argued that liberal acceptance, which refers to the tendency to assign greater meaning to weakly supported evidence, may be the primary mechanism of both DTD and JTC, respectively (Moritz et al., 2016; Moritz and Woodward, 2004; Prike et al., 2018; Reininghaus et al., 2019). Under certain conditions, liberal acceptance may occur even without concurrent reduced DTD (Moritz et al., 2016, 2007).

To address shortcomings of the beads task and to more accurately assess liberal acceptance, our group developed the box task (Andreou et al., 2015; Balzan et al., 2017). In the box task, participants are asked to open boxes containing balls of two different colors and should decide which of the two colors is more frequent. Subsequently, we developed a variant of the box task which includes a pressure condition. Here, participants are told that the task will be terminated prematurely and that performance will be counted as erroneous if they do not decide before the task is aborted. For the pressure condition, we found that people with elevated psychotic-like experiences were overconfident in their judgements, hastier, and also showed a lowered decision threshold (Moritz et al., 2017a). The box task has several advantages over the beads task. For example, the objective is more simplistic, making it easier to understand. Thus, it may pose less cognitive demands than the beads task (for the importance of working memory deficits for bead task performance see Freeman et al., 2014). Another possible advantage of the box task is that unlike the beads task many different parallel versions can be set up making it useful for both basic research and interventional programs which require multiple points of assessment.

Recent studies have cast doubt on the notion that persons with increased scores on schizotypy/psychosis proneness measures are actually at higher risk for developing psychosis (Moritz et al., 2019a). That is, results obtained in people with elevated ratings on self-report measures of psychotic-like experiences cannot be fully extrapolated to schizophrenia and vice versa (David, 2010). For this reason, in this study we examined whether we could replicate our previously reported finding using the box task (i.e., reduced DTD among those scoring higher on psychosis-like experiences) in a clinical sample of patients with schizophrenia. The present study used the 'pressure' condition that best discriminated between groups high and low on psychosis-proneness, in line with studies showing that JTC is exaggerated when patients are put under stress (Moritz et al., 2015b). We expected that patients with schizophrenia would show more JTC than controls. Like the forerunner trial, the study was conducted online.

2. Method

Patients were recruited online in the framework of a randomized controlled trial comparing an active intervention group (Internet-based cognitive behavioral therapy for psychosis; iCBTp) with a waitlist control group in a sample of individuals with schizophrenia-spectrum disorders (for further details on recruitment see Rüegg et al., 2018).

The registration website informed potential participants about the study aims and provided a link to an online survey platform.

Inclusion criteria were age of 18 or older, sufficient command of the German language, access to the Internet, provision of informed consent, acute and/or lifetime diagnosis of a schizophrenia-spectrum disorder according to DSM-5, simultaneous pharmacological and/or regular psychiatric or psychological care, and at least mild severity in one or more of the following three symptoms of the Positive and Negative Syndrome Scale (PANSS): delusions (P1), hallucinations (P3), and suspiciousness/ persecution (P6). Criteria for exclusion were acute suicidality, posing an acute threat to others, a neurological disease of the central nervous system requiring treatment, and refusal to formulate an emergency plan.

Following receipt of informed consent, questionnaires requesting sociodemographic and psychopathological data were administered online for the baseline assessment. If criteria for participation were met, study staff contacted interested persons to conduct a clinician administered telephone interview as second part of the baseline assessment. Interviewers conducted the Mini International Neuropsychiatric Interview (MINI) (Lecrubier et al., 1997) adopted for DSM-5 criteria. For the PANSS (Kay et al., 1987), the interviewers rated all PANSS items except for G5 'mannerisms and posturing', which could not be assessed via telephone. The interviewers were trained in extensive workshops and continuously supervised by experienced study staff. As a final step in the enrollment process, staff members used self-report and interviewer data to check inclusion and exclusion criteria.

Recruitment was undertaken from December 2016 to May 2018. A total of 7,237 persons accessed the study website, of whom 746 gave informed consent to participate. Eventually, 101 participants proceeded with the assessment and fulfilled all study criteria.

We registered the trial (clinicaltrials.gov, NCT02974400, 28.11.2016) and received approval from the Ethics Committee of the Canton of Bern, Switzerland, (KEK 03/14) and the German Psychological Society (SM052015).

Controls from the general population were recruited via WiSo-Panel. This German online platform provides researchers with the opportunity to conduct anonymous studies on large participant samples from the general population (Göritz, 2014). We derived data on the box task from a published study (Moritz et al., 2019b). From the 13,451 participants contacted by WiSo-Panel, 1,091 started the study; 763 completers fulfilled all criteria blind to results. From these, 101 participants were individually matched to the patient sample based on parallel consideration of gender, age and years of school education (a deviation by one year was tolerated) but blind to results on the box task. In case of multiple matches, participants were picked randomly. The study has a much larger sample size than most previous trials examining JTC in schizophrenia. However, our sample still represents a convenience sample because participants were not primarily recruited for the purpose of investigating JTC. Sample size was determined by other considerations.

2.1. Positive and Negative Syndrome Scale (PANSS)

The PANSS is a semi-structured interview that measures 30 items capturing symptoms typical of schizophrenia as well as more general psychopathology (Kay et al., 1987). Items are scored on a Likert scale ranging from 1 to 7. The scale is considered the gold standard for the assessment of symptom severity in schizophrenia (Suzuki, 2011). The PANSS was administered in patients only. The interview was conducted over the telephone (see above). Interviewers had a bachelor's degree in psychology or higher, were extensively trained in workshops, and were continuously supervised by experienced study staff. To establish inter-rater reliability, interviewers were asked to rate an audiotaped interview after the workshop. The mean correlation between trainers' and individual raters' PANSS item scores was $M_r = 0.82$, CI = [0.77, 0.85] and thus adequate. For flat affect, we mainly rated modulation of speech and expressive aspects of the language. Other negative symptoms like passive/apathetic social withdrawal, difficulty in abstract thinking,

lack of spontaneity and flow of conversation, stereotyped thinking and poor rapport could well be rated via telephone. Importantly, positive rather than negative symptoms represented the primary syndrome of interest.

2.2. Delusion and Voices Self-Assessment scale (DV-SA)

We administered the German version of the DV-SA (Pinto et al., 2007), which assesses delusions and auditory hallucinations from the patients' perspective. The DV-SA demonstrates satisfactory to good internal consistency and test-retest reliability according to its authors. The DV-SA was administered in patients only.

2.3. Rosenberg Self-Esteem Scale (R-SES)

The R-SES (Rosenberg, 1965) is the most widely used instrument for assessing self-esteem. The reliability and validity of the German version are considered good (von Collani and Herzberg, 2003). The R-SES was administered in patients only.

2.4. Patient Health Questionnaire – Depression Module (PHQ-9)

Depression severity was assessed with the PHQ-9 (Kroenke et al., 2001). This self-report measure captures the nine DSM-IV criteria for depression. Items are scored on an ordinal scale from "0" (not at all) to "3" (nearly every day; total score 0 to 27). Internal consistency is high (Cronbach's $\alpha = 0.86$–0.89). The PHQ-9 was administered in patients only.

2.5. Box task

The box task was originally developed by Andreou et al. (2015) based on the Information Sampling Task (IST) (Clark et al., 2006). We further revised the task and added probability and confidence ratings to allow assessment of the decision threshold and post-decisional confidence. For the present study, we used subtask D from the version published by Moritz et al. (2017a), which best discriminated individuals who scored low versus high on the Community Assessment of Psychic

Experiences positive scale in terms of core data gathering parameters. After a short practice trial, participants were presented with a panel comprised of 12 gray boxes (3 rows $\times$ 4 columns; see Fig. 1). Whenever participants pressed the "continue" button, a ball in red or green color was shown in one of the boxes. Unlike in the original version by Andreou et al. (2015; see also Balzan et al., 2017) participants could not indicate where the ball would appear. The ratio of the two different colors was also unknown to the participant. The participants' task was to decide which ball color would be more prevalent over time. After each step, participants were asked to endorse on a 11-point Likert scale the probability that one color would be displayed more often than the other (i.e., 100% sure red (=1), 90%, 80%, 70%, 60%, 50% (= 6), 60%, 70%, 80%, 90%, 100% sure green (=11); a rating of, for example, 8 indicated that the participant thought the likelihood that green balls would be shown more often than red balls is 70%).

Following each new ball, participants could either decide for color A or color B, or else delay their decision. If a decision was made, the participant was asked to provide their respective confidence rating on a 4-point Likert scale (1 = highly confident, 2 = quite confident, 3 = not very confident, 4 = guessed), after which the trial was terminated. Participants were informed that the trial would be terminated at some point before all balls were shown and that, in case no decision was made until this point, the trial would be counted as incorrect.

The task allows computation of the mean number of DTD (i.e., mean number of boxes opened before making a decision for one color), the decision threshold (i.e., the subjective probability a participant deemed sufficient for a decision), the initial probability (i.e., probability rating for the first ball of each trial) and decision confidence as well as accuracy. In this version of the box task, originally termed "difficult long," the termination of the trial came relatively late after the ninth ball; the ratio of the colors was almost balanced and the sequence was the same for all participants (rgrgrgrgr, r = red, g = green).

3. Results

As can be seen in Table 1, the two samples were matched according to gender, age and years of education. Typical for online studies, there were more female than male participants. The patient sample was
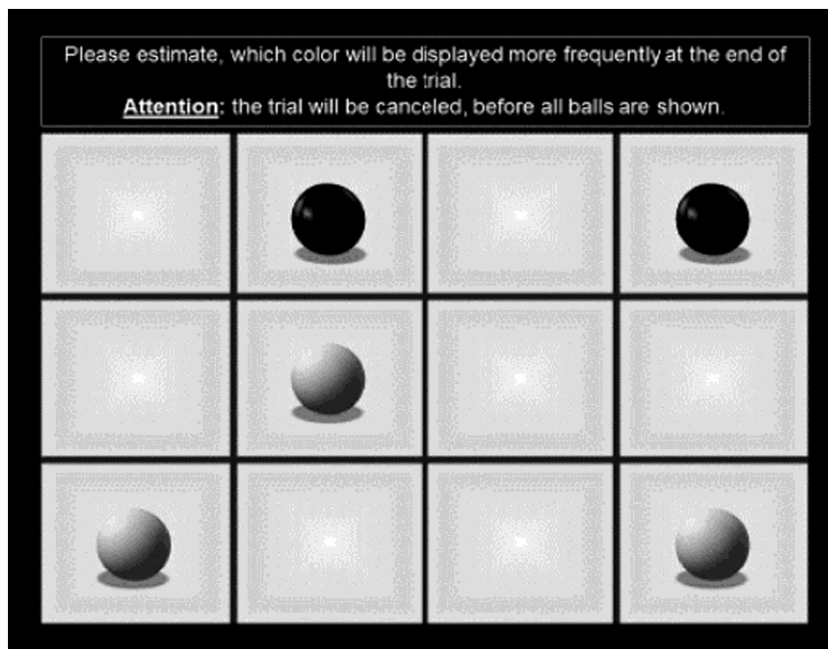

Fig. 1. Box task. Participants have to estimate, which color is shown more frequently in the end. Participants were told that the trial would be terminated before all balls were displayed (originally, red and green balls are shown).

Table 1
Demographic, psychopathological and box task performance of the sample frequently, means and standard devisions.

Variable	Schizophrenia ($n = 101$)	Control ($n = 101$)	Statistics
Demographic characteristics			Matched sample
Gender (male/female)	42/59	42/59	$\chi^2(1) = 0.00, p$ N .99
Age in years	40.04 (9.59)	40.04 (9.61)	$t(200) = 0.00, p$ N .99, $d = 0.00$
Years of formal school education	11.69 (1.62)	11.71 (1.58)	$t(200) = 0.09, p = .93, d = 0.01$
Psychopathology and treatment related variables			
PANSS total	59.98 (14.93)	–	n.a.
PHQ-9	12.26 (6.29)	–	n.a.
Rosenberg Self-Esteem Scale	25.82 (8.07)	–	n.a.
First admission inpatient treatment in years	25.64 (7.67)	–	n.a.
% maximum antipsychotic dosage	41.84 (40.36)	–	n.a.
Draws to decision	5.83 (3.74)	2.90 (3.29)	$t(200) = 5.91. p$ b .001, $d = 0.83$
Decision threshold in %	71.01 (16.99)	81.33 (13.76)	$t(157)^* = 4.12, p$ b .001, $d = 0.67$
Post-decisional confidence	2.61 (0.86)	3.64 (0.59)	$t(157)^* = 8.58, p$ b .001, $d = 1.40$
Initial probability in % (first box)	61.78 (16.76)	77.13 (17.40)	$t(200) = 6.38, p$ b .001, $d = 0.90$
Accuracy (0 = incorrect, 1 = correct)	0.41 (0.72)	0.34 (0.89)	$t(200) = 0.61, p = .543, d = 0.09$

Notes. * For the decision threshold and the confidence parameters data was not available from all participants as these parameters cannot be determined if no decision is made.

mildly ill according to PANSS scores (Leucht et al., 2005) and showed rather few psychotic symptoms accordingly but had a high length of illness and moderate severity of depression (PHQ-9).

As expected, patients had a lower decision threshold, but unexpectedly patients showed increased DTD and lower confidence compared to controls (see Table 1; please note that all statistics were not corrected for multiple comparisons). Accuracy did not differ between groups. In a next step, we explored the dynamics for this inconsistent pattern of results. Table 2 shows that a higher decision threshold was associated with higher confidence and higher initial probability rating. A higher decision threshold was also associated with better accuracy. The more beads were drawn the lower the confidence in the decision and the lower the initial probability. The former result seems counter-intuitive as more information should strengthen confidence. Yet, this finding might have occurred because in our set-up the chain of events (switching of red and green balls) ambiguates rather than disambiguates the solution; one may also argue that people with low initial confidence will require more information.

Subsequently, a number of exploratory correlations were run between the box task parameters with demographic and psychopathological variables (see Table 3). School education, gender, antipsychotic dosage, length of illness were not associated with any of the parameters.

Self-esteem ($r = 0.251, p = .037$) and the DV-SA delusion subscale ($r = 0.261, p = .030$) were correlated with enhanced confidence. The conventional PANSS positive score was negatively correlated with DTD ($r = -0.211, p = .034$). Notwithstanding that patients showed increased DTD relative to controls, we would like to report that within the patient group unusual thought content (i.e., bizarre delusions) was negatively correlated with DTD ($r = -0.203, p = .042$) and (at trend level) positively with confidence ($r = 0.231, p = .056$). At a weak-to-medium effect size, draws to decision was somewhat increased in the group ($n = 28$) who had undergone metacognitive training (MCT, $M = 6.75$) compared to those with no experience in MCT ($M = 5.47$), but the difference did not achieve significance, $t(99) = 1.61, p = .111, d = 0.35$.

JTC was counted as a decision after the first step, which was observed more often in controls (70.3%) than patients (25.7%), $\chi^2(1) =$

Table 2
Intercorrelations of box task parameters (full sample).

	Decision threshold	Confidence	Accuracy	% first ball
Draws to decision	−0.652***	−0.519***	−0.151*	−0.664***
Decision threshold	1	0.594***	0.205**	0.780***
Confidence	0.594***	1	−0.013	0.577***
Accuracy	0.205**	−0.013	1	0.106

Notes. * p b .05; ** p b .005; *** p b .001.

40.16, p b .001. Likewise, more controls (89.1%) than patients (68.3%) completed the task with a decision, $\chi^2(1) = 13.03, p$ b .001.

3.1. Effects of repeated administration on the box task

A total of 22 non-clinical individuals had participated earlier in the study on the original box task (Moritz et al., 2017a). The correlation of the two DTD parameters (subtask D as part of the entire box task and subtask D as sole trial) was moderate but there was evidence for less DTD upon repetition ($M_{prior} = 4.91$ and $M_{current} = 3.32$, $t(21) = 2.02, p = .056$). We therefore recalculated the box task parameters for those participants only who had no experience with the box task before. All results remained essentially the same (all previously significant results remained at p b .001; again, a nonsignificant effect for accuracy).

4. Discussion

This study replicated previous research showing an elevated liberal acceptance bias in schizophrenia, but—inconsistent with most other studies—found no evidence of an elevated JTC bias (or a reduced DTD) in this population relative to controls. This is consistent with a small, but growing number of studies showing no enhanced JTC bias in schizophrenia patients (e.g., Lincoln et al., 2010; Moritz et al., 2016), or even heightened conservatism (i.e., more DTD) in people scoring high on measures of delusional ideation (Baker et al., 2019; McLean et al., 2020). Unlike recent studies conducted with the beads task or its variants,

the fish task (in the study by Moritz et al., 2017a, the fish task was run in parallel to the box task) or sheep task (Moritz et al., 2016), we were unable to find fewer DTD in patients in the present study. On the contrary, patients with schizophrenia showed relatively delayed decision-making in a newly designed task, which had previously yielded the expected behavioral pattern (i.e., fewer DTD, overconfidence, lowered decision threshold) in participants with elevated psychotic-like experiences. Yet, the previous study was conducted in a sample derived from the general population split according to the severity of psychosis-like symptoms and had been awaiting confirmation in a patient sample. In line with our two stage theory of delusions a lowered decision threshold was found in psychosis (i.e., a lower probability was deemed sufficient to decide for one of the colors) suggesting that patients are more liberal to consider a hypothesis/to make a decision. While this result was indeed expected, it should be interpreted cautiously in light of the uneven course of probability judgements between groups we will

turn to later.

Against the background of meta-analytic evidence and several different paradigms that confirmed JTC in patients, we would be jumping to conclusions ourselves if we fully dismissed the JTC account of

Table 3
Correlations between box task parameters with external variables (patients only).

	Draws to decision	Decision threshold in %	Post-decisional confidence	Initial probability in %	Accuracy
Age in years	−0.050	−0.079	0.091	−0.008	−0.133
Years of formal school education	0.145	−0.205 (+)	−0.049	−0.112	0.142
PANSS Sum score	−0.101	0.025	0.004	0.104	0.011
PANSS Positive	−0.211*	0.064	0.133	0.063	0.091
DV-SA delusion	0.043	0.288*	0.261*	0.048	0.149
PHQ	−0.108	0.292*	−0.008	0.197*	0.131
Rosenberg Self-Esteem Scale	−0.007	−0.056	0.251*	−0.004	0.030
First admission inpatient treatment	−0.033	−0.014	0.015	0.005	−0.141
% maximum antipsychotic dosage	0.016	0.131	0.037	−0.047	0.022

Notes. Accuracy, 0 = incorrect, 1 = correct.
Notes. * p b.05.

schizophrenia. Yet, we may need to rethink our theoretical framework and investigate under which conditions the bias does and does not occur. A presumed advantage of the box task over other paradigms is that it collects several parameters beyond DTD, which may inform us about underpinnings of hasty vs. delayed decision-making. Looking at these parameters indeed allows some insight into the at first counter-intuitive results of this study. Patients displayed lower initial probabilities than controls; the high correlation of this parameter with DTD (see Table 2) renders it likely that it contributed substantially to the delayed DTD. Such a response pattern has been found before; in one of our studies using a variant of the beads task, patients showed lower initial probabilities and a lower decision threshold but no deviant DTD (Moritz et al., 2016). This abnormality, which is not found across all studies (Moritz et al., 2007), may explain similar or even delayed DTD. We do not think, however, that this explanation is sufficient to account for the entire pattern of results and recommend that researchers in the field should continue to seek further mediators of the phenomenon. We think it is
fruitful to further explore the contribution of the following factors:

1. *Motivation*: Poor motivation may result in premature termination of a task in order to save time. This mediator has been rendered unlikely in prior studies using graded estimates procedures (Moritz and Woodward, 2005), simulated decisions (Moritz et al., 2007), a gambling condition rewarding delayed decision-making (Lincoln et al., 2010), or the non-serial data-gathering paradigm (van der Leer et al., 2017). Yet, we need to remember that initial probability, as mentioned before, also differs across studies. Thus, ruling out a mediator in one study does not preclude its contribution in others. In the current study, the controls from the general population may have been hastier simply because these were "hired for work" (i.e., participants from a large participant pool who are usually paid for their services), compared to patients who underwent the task as part of an intervention trial and whose motivation was thus likely higher and more intrinsic. In line with this, JTC was exceptionally high in the control sample.

2. *Intelligence, task difficulty and comprehension problems*: The beads task and possibly also its variants, like the fish and the sheep task, are posing high cognitive demands for some participants. A subgroup of participant makes illogical responses and seemingly misunderstands the task (Balzan et al., 2012; Klein and Pinkham, 2018; Moritz and Woodward, 2005). The box task was constructed to put less cognitive burden on participants. Yet, it needs to be tested whether it is indeed easier to grasp than the original bead task. Sub-task D is perhaps especially difficult as the ratio of red and green balls is even and one may well argue that any decision is premature (in contrast, in the beads task, increasing information disambiguates the task). The effect of subjective difficulty on task performance is hard to predict. Some studies point to the possibility that low intelligence and neurocognitive deficits foster hasty decision-making (Takeda et al., 2018; Tripoli et al., 2019), especially in more difficult tasks (Gaweda et al., 2017), while other research from the field of

memory suggests that difficult tasks may lead to more cautious responses in patients; the lowered confidence found here speaks for the latter (Moritz et al., 2015a). In this case, the high ambiguity (i.e., the sequence did not clearly favor one color) may have persuaded controls to decide prematurely. Pressure may be a double-edged sword in that it may induce stress in some participants which has been associated with hastier decision-making in patients (Moritz et al., 2015b) or may trigger conservatism.

3. *Prior experience and battery*: At least a subgroup of the control participants knew the box task, as the WiSo Panel sample had been invited before to participate in the beads task (Moritz et al., 2017a). We looked in how far this distorted results. At trend level, there was evidence that DTD was shorter upon repetition. Exclusion of those participants who underwent the box task before did not change group differences, however. Similarly, patients who have been treated with interventions aimed to ameliorate cognitive biases such as JTC and overconfidence may show normalized performance. We have checked for this and found some evidence (small to medium but nonsignificant effect size in the expected direction), which deserves to be investigated further. Finally, the design of the test battery may also play a role. Ideally, test batteries are similar across samples but this is the exception rather than the norm. For example, the PANSS and other schizophrenia-relevant instruments are usually not administered in controls, which is also true for the present study. Time of assessment until a decision-making task is administered may be relevant as a short attention span may compromise logical thinking and influence decision-making in unpredicted ways. Moreover, the type of task preceding the assessment may well influence performance. For example, a preceding task favoring speed such as tasks sensitive to take the best heuristics (Gigerenzer and Goldstein, 1996), may shorten decision-making while a task favoring accuracy may delay decision-making. In our study, the data of controls was derived from a study, where we tested the assumption that those high on schizotypy would show overconfidence, JTC as well as more subclinical positive symptoms under conditions that induce a dilemma (McGregor et al., 2001). This hypothesis was not verified, that is, participants exposed to the dilemma showed similar results as those on the control condition. However, it cannot be excluded that both the experimental and the control condition speeded performance. Speaking for this argument, the rate of JTC (decision after one box opened) was exceptionally high in controls.

4. *The role of psychopathology*: For exploratory purposes, we looked at correlates of box task performance. While none of these would have withstood a Bonferroni correction, some interesting results emerged that should be followed up. In line with the literature, there was a small association between positive symptoms and fewer DTD. More interestingly, psychopathology and self-esteem were correlated with confidence. Both aspects may be very central to our understanding of psychosis but have not been linked much before. Conviction, an extreme form of confidence, is a core criterion of delusions (Jaspers, 1913) while self-esteem is usually low in patients

(Freeman et al., 1998) but volatile and associated with the delusional content (Sundag et al., 2015). So far, the evidence that low self-esteem is a risk factor for psychosis is not fully conclusive, perhaps owing to methodological problems (Moritz et al., 2017c). Yet, there are some suggestions that fluctuations in self-esteem may be involved in the formation of paranoia (Udachina et al., 2014) and it has been recommended that confidence should be considered in this context (Thewissen et al., 2008). Such a study will likely refute any simple trait-like differences of patients and controls in terms confidence level but perhaps highlight contextual variables that may exaggerate confidence and subsequently paranoia. Such *sparks* may already be sufficient to trigger psychological processes (e.g., confirmation bias) that can foster the onset of delusional ideas. The study faces several limitations. In addition to the issues already discussed, it is worth noting that the schizophrenia sample displayed few core symptoms. Further, we only administered a single trial of the box task, which had best discriminated individuals high and low on psychotic-like experiences in our forerunner study. This unexpected result was not replicated which could mean different things apart from the confirmation of the null hypothesis. For example, it could be that excessive JTC in patients needs time to evolve (the present study administered subtask D only, while it was preceded by other subtasks in the forerunner study). One may argue, however, that repeated presentation of the beads task, the conventional task to capture JTC, rather washes out group differences (Rausch et al., 2014). We have no explanation as to why this could be the case but it deserves to be tested. Otherwise, a single trial is clearly unreliable and we should have administered other JTC paradigms concurrently, for example, the "Judge Advisor System" (Scheunemann et al., 2020) or the "What is this?" task (Serrano-Guerrero et al., 2020). Unfortunately, different tests show inconsistent external validity; for example, variants of the beads task are modestly correlated with the box task (Moritz et al., 2017a) and there was no relationship with a subjective (self-report) measure of JTC (Peters et al., 2014). Finally, our design was unbalanced, in the sense that the assessment battery was not identical in the two participant groups. For example, the self-esteem scale was used in patients only.

To conclude, the study found evidence for a lowered decision threshold but not less JTC in participants with psychosis. This challenges the ubiquity of the JTC in psychosis. Researchers may explore why some successors of the beads task have found JTC in patients (Moritz et al., 2006; Moritz and Woodward, 2006b; Serrano-Guerrero et al., 2020; Speechley et al., 2010; Veckenstedt et al., 2011; Ziegler et al., 2008) while others have not.

We encourage researchers to further examine possible mediators but also moderators that may explain the waxing and waning nature of JTC in psychosis.

Contributors

Christina Andreou, Ryan P. Balzan and Steffen Moritz developed the research idea. Stefan Westermann and Thies Lüdtke conducted the trial and data collection. Jakob Scheunemann and Gerit Pfuhl were important for the theoretical conceptualization. All authors contributed substantially and proof-read the manuscript. Stefan Westermann was PI of the study.

Declaration of competing interest

None.

References

Andreou, C., Schneider, B.C., Braun, V., Kolbeck, K., Gallinat, J., Moritz, S., 2015. Dopamine effects on evidence gathering and integration. J. Psychiatry Neurosci. 40, 140306.

Baker, S.C., Konova, A.B., Daw, N.D., Horga, G., 2019. A distinct inferential mechanism for delusions in schizophrenia. Brain 142, 1797–1812. https://doi.org/10.1093/brain/ awz051.

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. Aust. N. Z. J. Psychiatry 46, 532–540. https://doi.org/10.1177/0004867411435291.

Balzan, R.P., Ephraums, R., Delfabbro, P., Andreou, C., 2017. Beads task vs. box task: the specificity of the jumping to conclusions bias. J. Behav. Ther. Exp. Psychiatry 56, 42–50. https://doi.org/10.1016/j.jbtep.2016.07.017.

Clark, L., Robbins, T.W., Ersche, K.D., Sahakian, B.J., 2006. Reflection impulsivity in current and former substance users. Biol. Psychiatry 60, 515–522. https://doi.org/10.1016/j. biopsych.2005.11.007.

David, A.S., 2010. Why we need more debate on whether psychotic symptoms lie on a continuum with normality. Psychol. Med. 40, 1935–1942. https://doi.org/10.1017/ S0033291710000188.

Dudley, R., Taylor, P., Wickham, S., Hutton, P., 2016. Psychosis, delusions and the "jumping to conclusions" reasoning bias: a systematic review and meta-analysis. Schizophr. Bull. 42, 652–665. https://doi.org/10.1093/schbul/sbv150.

Fine, C., Gardner, M., Craigie, J., Gold, I., 2007. Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. Cogn. Neuropsychiatry 12, 46–77. https://doi.org/10.1080/13546800600750597.

Freeman, D., Garety, P., Fowler, D., Kuipers, E., Dunn, G., Bebbington, P., Hadley, C., 1998. The London-East Anglia randomized controlled trial of cognitive-behaviour therapy for psychosis. IV: self-esteem and persecutory delusions. Br. J. Clin. Psychol. 37 (Pt 4), 415–430. https://doi.org/10.1111/j.2044-8260.1998.tb01399.x.

Freeman, D., Startup, H., Dunn, G., Cernis, E., Wingham, G., Pugh, K., Cordwell, J., Mander, H., Kingdon, D., 2014. Understanding jumping to conclusions in patients with persecutory delusions: working memory and intolerance of uncertainty. Psychol. Med. 44, 1–8. https://doi.org/10.1017/S0033291714000592.

Garety, P.A., Hemsley, D.R., Wessely, S., 1991. Reasoning in deluded schizophrenic and paranoid patients. Biases in performance on a probabilistic inference task. J. Nerv. Ment. Dis. 179, 194–201. https://doi.org/10.1097/00005053-199104000-00003.

Gawęda, Ł., Prochwicz, K., Krężołek, M., Kłosowska, J., Staszkiewicz, M., Moritz, S., 2017. Self-reported cognitive distortions in the psychosis continuum: a Polish 18-item version of the Davos Assessment of Cognitive Biases Scale (DACOBS-18). Schizophr. Res. https://doi.org/10.1016/j.schres.2017.05.042.

Gigerenzer, G., Goldstein, D.G., 1996. Reasoning the fast and frugal way: models of bounded rationality. Psychol. Rev. 103, 650–669. https://doi.org/10.1037/0033295X.103.4.650.

Göritz, A., 2014. Determinants of the starting rate and the completion rate in online panel studies. In: Callegaro, M., Baker, R., Bethlehem, J., Göritz, A.S., Krosnick, J.A., Lavrakas, P.J. (Eds.), Online Panel Research: A Data Quality Perspective. Wiley, Chichester, pp. 154–170. https://doi.org/10.1002/9781118763520.ch7.

Hoven, M., Lebreton, M., Engelmann, J.B., Denys, D., Luigjes, J., van Holst, R.J., 2019. Abnormalities of confidence in psychiatry: an overview and future perspectives. Transl. Psychiatry 9, 268. https://doi.org/10.1038/s41398-019-0602-7.

Jaspers, K., 1913. Allgemeine Psychopathologie [General Psychopathology]. 4. Auflage. Springer, Berlin (Germany) https://doi.org/10.1007/978-3-642-52895-8.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophr. Bull. 13, 261–276. https://doi.org/10.1093/ schbul/13.2.261.

Kinderman, P., Bentall, R.P., 1996. A new measure of causal locus: the internal, personal and situational attributions questionnaire. Pers. Individ. Dif. 20, 261–264. https:// doi.org/10.1016/0191-8869(95)00186-7.

Klein, H.S., Pinkham, A.E., 2018. Examining reasoning biases in schizophrenia using a modified "Jumping to Conclusions" probabilistic reasoning task. Psychiatry Res. 270, 180–186. https://doi.org/10.1016/j.psychres.2018.09.020.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Gen. Intern. Med. 16, 605–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Lecrubier, Y., Sheehan, D.V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K.H., Janavs, J., Dunbar, G.C., 1997. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur. Psychiatry 12, 224–231. https://doi.org/10.1016/S0924-9338(97)83296-8.

Leucht, S., Kane, J.M., Kissling, W., Hamann, J., Etschel, E., Engel, R.R., 2005. What does the PANSS mean? Schizophr. Res. 79, 231–238. https://doi.org/10.1016/j. schres.2005.04.008.

Lincoln, T.M., Ziegler, M., Mehl, S., Rief, W., 2010. The jumping to conclusions bias in delusions: specificity and changeability. J. Abnorm. Psychol. 119, 40–49. https://doi. org/10.1037/a0018118.

McGregor, I., Zanna, M.P., Holmes, J.G., Spencer, S.J., 2001. Compensatory conviction in the face of personal uncertainty: going to extremes and being oneself. J. Pers. Soc. Psychol. 80, 472–488. https://doi.org/10.1037/0022-3514.80.3.472.

McLean, B.F., Mattiske, J.K., Balzan, R.P., 2017. Association of the jumping to conclusions and evidence integration biases with delusions in psychosis: a detailed meta-analysis. Schizophr. Bull. 43, 344–354. https://doi.org/10.1093/schbul/sbw056.

McLean, B.F., Mattiske, J.K., Balzan, R.P., 2020. Jumping to conclusions in the less-delusion-prone? Preliminary evidence from a more reliable beads task. J. Behav. Ther. Exp. Psychiatry 68, 101562. https://doi.org/10.1016/j.jbtep.2020.101562.

Moritz, S., Woodward, T., 2004. Plausibility judgment in schizophrenic patients: evidence for a liberal acceptance bias. Ger. J. Psychiatry 7, 66–74.

Moritz, S., Woodward, T.S., 2005. Jumping to conclusions in delusional and non-delusional schizophrenic patients. Br. J. Clin. Psychol. 44, 193–207. https://doi.org/ 10.1348/014466505X35678.

Moritz, S., Woodward, T.S., 2006a. Metacognitive control over false memories: a key determinant of delusional thinking. Curr. Psychiatry Rep. 8, 184–190. https://doi.org/ 10.1007/s11920-006-0022-2.

Moritz, S., Woodward, T.S., 2006b. A generalized bias against disconfirmatory evidence in schizophrenia. Psychiatry Res. 142, 157–165. https://doi.org/10.1016/ j.psychres.2005.08.016.

Moritz, S., Woodward, T.S., Hausmann, D., 2006. Incautious reasoning as a pathogenetic factor for the development of psychotic symptoms in schizophrenia. Schizophr. Bull. 32. https://doi.org/10.1093/schbul/sbj034.

Moritz, S., Woodward, T.S., Lambert, M., 2007. Under what circumstances do patients with schizophrenia jump to conclusions? A liberal acceptance account. Br. J. Clin. Psychol. 46, 127–137. https://doi.org/10.1348/014466506X129862.

Moritz, S., Göritz, A.S., Gallinat, J., Schafschetzy, M., Van Quaquebeke, N., Peters, M.J.V., Andreou, C., 2015a. Subjective competence breeds overconfidence in errors in psychosis. A hubris account of paranoia. J. Behav. Ther. Exp. Psychiatry 48, 118–124. https://doi.org/10.1016/j.jbtep.2015.02.011.

Moritz, S., Köther, U., Hartmann, M., Lincoln, T.M., 2015b. Stress is a bad advisor. Stress primes poor decision making in deluded psychotic patients. Eur. Arch. Psychiatry Clin. Neurosci. 265, 461–469. https://doi.org/10.1007/s00406-015-0585-1.

Moritz, S., Scheu, F., Andreou, C., Pfueller, U., Weisbrod, M., Roesch-Ely, D., 2016. Reasoning in psychosis: risky but not necessarily hasty. Cogn. Neuropsychiatry 21, 91–106. https://doi.org/10.1080/13546805.2015.1136611.

Moritz, S., Göritz, A.S., Balzan, R.P., Gawęda, Ł., Kulagin, S.C., Andreou, C., 2017a. A new paradigm to measure probabilistic reasoning and a possible answer to the question why psychosis-prone individuals jump to conclusions. J. Abnorm. Psychol. 126, 406–415. https://doi.org/10.1037/abn0000262.

Moritz, S., Pfuhl, G., Lüdtke, T., Menon, M., Balzan, R.P., Andreou, C., 2017c. A two-stage cognitive theory of the positive symptoms of psychosis. Highlighting the role of lowered decision thresholds. J. Behav. Ther. Exp. Psychiatry 56, 12–20. https://doi. org/10.1016/j.jbtep.2016.07.004.

Moritz, S., Bentall, R.P., Kolbeck, K., Roesch-Ely, D., 2018. Monocausal attribution and its relationship with reasoning biases in schizophrenia. Schizophr. Res. 193, 77–82. https://doi.org/10.1016/j.schres.2017.06.057.

Moritz, S., Gawęda, Ł., Heinz, A., Gallinat, J., 2019a. Four reasons why early detection centers for psychosis should be renamed and their treatment targets reconsidered: we should not catastrophize a future we can neither reliably predict nor change. Psychol. Med. 49, 2134–2140. https://doi.org/10.1017/S0033291719001740.

Moritz, S., Stojisavlevic, M., Göritz, A.S., Riehle, M., Scheunemann, J., 2019b. Does uncertainty breed conviction? On the possible role of compensatory conviction in jumping to conclusions and overconfidence in psychosis. Cogn. Neuropsychiatry 24, 284–299. https://doi.org/10.1080/13546805.2019.1642863.

Peters, E., Garety, P., 2006. Cognitive functioning in delusions: a longitudinal analysis. Behav. Res. Ther. 44, 481–514. https://doi.org/10.1016/j.brat.2005.03.008.

Peters, E.R., Moritz, S., Schwannauer, M., Wiseman, Z., Greenwood, K.E., Scott, J., Beck, A.T., Donaldson, C., Hagen, R., Ross, K., Veckenstedt, R., Ison, R., Williams, S., Kuipers, E., Garety, P.A., 2014. Cognitive biases questionnaire for psychosis. Schizophr. Bull. 40, 300–313. https://doi.org/10.1093/schbul/sbs199.

Pinto, A., Gigantesco, A., Morosini, P., La Pia, S., 2007. Development, reliability and validity of a self-administered questionnaire on subjective opinion about delusions and voices. Psychopathology 40, 312–320. https://doi.org/10.1159/000105529.

Prike, T., Arnold, M.M., Williamson, P., 2018. The relationship between anomalistic belief and biases of evidence integration and jumping to conclusions. Acta Psychol. 190, 217–227. https://doi.org/10.1016/j.actpsy.2018.08.006.

Rausch, F., Mier, D., Eifler, S., Esslinger, C., Schilling, C., Schirmbeck, F., Englisch, S., Meyer-Lindenberg, A., Kirsch, P., Zink, M., 2014. Reduced activation in ventral striatum and ventral tegmental area during probabilistic decision-making in schizophrenia. Schizophr. Res. 156, 143–149. https://doi.org/10.1016/j.schres.2014.04.020.

Reininghaus, U., Oorschot, M., Moritz, S., Gayer-Anderson, C., Kempton, M.J., Valmaggia, L., McGuire, P., Murray, R., Garety, P., Wykes, T., Morgan, C., Myin-Germeys, I., 2019. Liberal acceptance bias, momentary aberrant salience, and psychosis: an experimental experience sampling study. Schizophr. Bull. 45, 871–882. https://doi.org/10.1093/ schbul/sby116.

Rosenberg, M., 1965. Society and the Adolescent Self Image. Princeton University Press, Princeton, NJ.

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. Schizophr. Bull. 41, 1183–1191. https://doi.org/10.1093/schbul/sbu187.

Rüegg, N., Moritz, S., Berger, T., Lüdtke, T., Westermann, S., 2018. An internet-based intervention for people with psychosis (EviBaS): study protocol for a randomized cottrolled trial. BMC Psychiatry 18. https://doi.org/10.1186/s12888-018-1644-8.

Sanford, N., Veckenstedt, R., Moritz, S., Balzan, R.P., Woodward, T.S., 2014. Impaired integration of disambiguating evidence in delusional schizophrenia patients. Psychol. Med. 44. https://doi.org/10.1017/S0033291714000397.

Scheunemann, J., Gawęda, Ł., Reininger, K.M., Jelinek, L., Hildebrandt, H., Moritz, S., 2020. Advice weighting as a novel measure for belief flexibility in people with psychotic-like experiences. Schizophr. Res. 216, 129–137. https://doi.org/10.1016/j. schres.2019.12.016.

Serrano-Guerrero, E., Ruiz-Veguilla, M., Martín-Rodríguez, A., Rodríguez-Testal, J.F., 2020. Inflexibility of beliefs and jumping to conclusions in active schizophrenia. Psychiatry Res. 284, 112776. https://doi.org/10.1016/j.psychres.2020.112776.

Speechley, W.J., Whitman, J.C., Woodward, T.S., 2010. The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. J. Psychiatry Neurosci. 35, 7–17. https://doi.org/10.1503/jpn.090025.

Sundag, J., Lincoln, T.M., Hartmann, M.M., Moritz, S., 2015. Is the content of persecutory delusions relevant to self-esteem? Psychosis 7, 237–248. https://doi.org/10.1080/ 17522439.2014.947616.

Suzuki, T., 2011. Which rating scales are regarded as "the standard" in clinical trials for schizophrenia? A critical review. Psychopharmacol. Bull. 44, 18–31.

Takeda, T., Nakataki, M., Ohta, M., Hamatani, S., Matsuura, K., Ohmori, T., 2018. Effect of cognitive function on jumping to conclusion in patients with schizophrenia. Schizophr. Res. Cogn. 12, 50–55. https://doi.org/10.1016/j.scog.2018.04.002.

Thewissen, V., Bentall, R.P., Lecomte, T., van Os, J., Myin-Germeys, I., 2008. Fluctuations in self-esteem and paranoia in the context of daily life. J. Abnorm. Psychol. 117, 143–153. https://doi.org/10.1037/0021-843X.117.1.143.

Tripoli, G., Quattrone, D., Murray, G., Gayer-Anderson, C., Rodriguez, V., Ferraro, L., Cascia, C. La, Sartorio, C., Seminerio, F., Barbera, D. La, Morgan, C., Sham, P., Forti, M. Di, Murray, R., 2019. Jumping to conclusions, general intelligence, and psychosis liability: findings from the multi-centre EU-GEI case-control study. Schizophr. Bull. 45, S219–S220. https://doi.org/10.1093/schbul/sbz019.322.

Udachina, A., Varese, F., Myin-Germeys, I., Bentall, R.P., 2014. The role of experiential avoidance in paranoid delusions: an experience sampling study. Br. J. Clin. Psychol. 53, 422–432. https://doi.org/10.1111/bjc.12054.

van der Leer, L., Hartig, B., Goldmanis, M., McKay, R., 2017. Why do delusion-prone individuals "jump to conclusions"? An investigation using a nonserial data-gathering paradigm. Clin. Psychol. Sci. 5, 718–725. https://doi.org/10.1177/2167702617698811.

Veckenstedt, R., Randjbar, S., Vitzthum, F., Hottenrott, B., Woodward, T.S., Moritz, S., 2011. Incorrigibility, jumping to conclusions, and decision threshold in schizophrenia. Cogn. Neuropsychiatry 16. https://doi.org/10.1080/13546805.2010.536084.

von Collani, G., Herzberg, P.Y., 2003. Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. Z. Dif. Diagn. Psychol. 24, 3–7.

Woodward, T.S., Moritz, S., Menon, M., Klinge, R., 2008. Belief inflexibility in schizophrenia. Cogn. Neuropsychiatry 13, 267–277. https://doi.org/10.1080/ 13546800802099033.

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? Psychol. Psychother. 81, 237–245. https://doi.org/10.1348/147608308X297104.