Psychiatry Research 265 (2018) 200–207

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Towards a reliable repeated-measures beads task for assessing the jumping to conclusions bias

Benjamin F. McLean[a,b], Julie K. Mattiske[a,*], Ryan P. Balzan[a,b]

[a] College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia
[b] ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia

ARTICLE INFO

Keywords:
Beads task
Jumping to conclusions
Repeated measures
Repeatability
Reliability
Psychosis

ABSTRACT

The jumping to conclusions bias (JTC), in which some people gather less information than others before making a decision, has been linked to delusions in psychosis. JTC is usually identified via the beads task, in which a sequence of beads (the "target" sequence) is used to measure the amount of evidence participants require before making a decision. Yet, despite its common use, the reliability of the task has never been properly investigated. We investigated its reliability, and tested an alternate version which used distractor sequences to obfuscate the target sequence. Healthy participants ($N = 212$) were randomised into two groups. One group completed ten trials using the target sequence, while the other completed ten trials of the target sequence and three distractor sequences. Our data indicated the standard task may not be reliable over repeated measures, but that by including distractor sequences, the task becomes more believable, repeatable, and reliable. Additionally, excluding first-trial data (a "silent" practice trial) also improves repeatability. These improvements to the task are relevant to single trial studies, and will be especially useful to repeated-measures longitudinal, experimental, and treatment studies. Such repeated-measures studies are important for investigating the causal link between JTC and delusions.

1. Introduction

The beads task (Huq et al., 1988) is used to measure the *jumping to conclusions* bias (JTC), in which some people gather less information than others might do before making a decision. Typically, participants are presented with two jars containing beads of two colours in equal but opposite ratios. One jar is chosen at random, and beads are drawn one at a time (purportedly at random, but actually in a predetermined sequence) from the jar. At each draw, participants either make a decision as to which jar beads were being drawn from, or request to see more beads. The amount of data gathered is most often measured as the number of beads drawn before a decision is made ("draws to decision" – DTD), though other criteria are possible, such as Liberal Acceptance (in which participants make a decision at a low level of confidence; Moritz et al., 2007). The beads task has been key in the large body of research linking JTC with delusions in psychosis (Dudley et al., 2016; McLean et al., 2017; Ross et al., 2015). However, despite the sizeable beads-task literature (a search of PubMed Central for the terms "*JTC*", "*jump(ing) to conclusions*", or "*beads task*" returned 1390 full-text journal hits), the reliability of the beads task has never been seriously investigated.

In the large majority of beads-task studies only a single trial of the beads task is presented – and when more than one trial is presented, the same sequence of beads is generally used repeatedly. Yet it is unclear whether DTD captured over single or multiple trials is a genuine reflection of a participant's true DTD. Two studies have found acceptable test-retest reliability over two and three trials (Lincoln et al., 2010; Moritz et al., 2015). However, and concerningly, two recent meta-analyses have shown that responses on the beads task vary with the number of trials provided (Dudley et al., 2016; Ross et al., 2015). There is also some evidence that between-group differences in JTC may only be observable on a single trial, and may disappear over multiple trials (Krug et al., 2014; Rausch et al., 2014). That the beads task may not be consistent across repeated measures signals possible problems with the psychometric properties of the task both in single-trial studies, and in longitudinal, treatment, and experimental studies that employ repeated administrations. Thus, our first objective was to assess consistency over multiple trials.

Changes in response to the beads-task over repeated measures might occur for a variety of reasons. Perhaps participants' responses change when they discover the predetermined nature of the bead sequence with which they are presented. Participants repeatedly exposed to the

* Corresponding author at: College of Education, Psychology and Social Work, Flinders University, Adelaide, Australia.
E-mail addresses: ben.mclean@flinders.edu.au (B.F. McLean), julie.mattiske@flinders.edu.au (J.K. Mattiske).

https://doi.org/10.1016/j.psychres.2018.04.043
Received 8 November 2017; Received in revised form 29 March 2018; Accepted 12 April 2018
Available online 13 April 2018
0165-1781/

AAABAAAABA sequence in common use (the “target” sequence) may soon observe that the first few beads are always of one colour, and that the first colour is always the majority colour, and begin to make earlier decisions than previously. Bead colours are usually changed from trial to trial to try and prevent this (see for example Ross et al., 2011; Waller et al., 2011), however simply changing bead colours may not effectively hide the single repeating target sequence.

Alternatively, perhaps early beads-task responses are influenced by the range of unknowns which participants undoubtedly face when completing this task for the first time. Supporting this idea, evidence from the broader area of cognitive psychology has found that practice effects over the first two trials of a range of cognitive assessment tools can produce significant changes, while responses over later trials are more consistent (Collie et al., 2003). If this is the case, it is of particular concern that the beads task is usually only administered once, and with no practice trial (Dudley et al., 2016).

Other measures of JTC have the potential to avoid the problem of repeating identical sequences, such as the box task (Balzan et al., 2017; Moritz et al., 2017), however the task is relatively new and its validity has not yet been established. The non-serial data-gathering paradigm of van der Leer et al. (2017), in which participants request the full number of fish they wish to see drawn from a lake in one go, could also avoid the issues with repeating sequences. Again however, this task is a significant divergence from the standard beads task, and requires further validation. In the meantime, the beads task is in many ways the standard method of identifying JTC, and its limitations over repeated measures needs to be established and improved.

To address the potential problem of participants recognising the key characteristics of the beads-task target sequence over repeated measures (which may result in changes in response), we proposed interleaving this sequence between distractor sequences. We expected distractor sequences to reduce the salience of the target sequence and, therefore, participants' recognition of its main features. We tested this solution using a two-group (target-sequence-only and distractor-sequences groups) repeated-measures design, in which each group completed 10 trials of their respective beads task. In addition, to assess the efficacy of a practice trial to address the potential problem of participants responding less reliably on initial trials, we compared the consistency of two consecutive trials conducted with and without a practice trial.

In order to assess the efficacy of distractor sequences in disguising the fact that the beads sequences were non-random, we recorded participants' self-reported belief that they were random. To assess the consistency over repeated trials of the target-sequence-only and distractor-sequences tasks, we measured the repeatability and reliability of the beads task over repeated measures. Here repeatability means that a measure returns stable or consistent values over trials under conditions in which a construct's true level should not change (e.g., measurements taken close in time, under identical conditions, with no intervention between measurements). This was operationalised as the mean squared error (MSE) over repeated measurements, reflecting intra-individual stability. Reliability on the other hand is the ability to discriminate between levels of a variable in the presence of noise (Portney and Watkins, 2015). This was operationalised as the Intraclass Correlation Coefficient (ICC(1,1), Shrout and Fleiss, 1979), which is the proportion of total variance explained by the true variance due to genuine differences between people.

The impact of having a reliable repeated-measures beads task would be significant. Such a task is needed for longitudinal, experimental, and treatment studies, and whenever aggregate measures are to be calculated for increased accuracy and precision.

2. Methods

2.1. Participants

240 participants on the Prolific online crowdsourcing platform completed our experiment. Participants were paid £2.20 for participation.

2.2. Beads task

Every participant completed ten trials of either the target-sequence-only or distractor-sequences beads task. The target-sequence-only group completed ten trials using only the target sequence (AAABAAAABA), while the distractor-sequences group completed ten trials using the target sequence plus three additional distractor sequences (i.e., the target sequence was presented ten times to this group also, but along with three distractor sequences each time it was presented). Comparing the two groups on an equal number of target sequences necessitated the distractor-sequences group completing 40 beads-task sequences compared with the 10 completed by the target-sequence-only group. The distractor sequences included one sequence for which the first bead colour was the minority colour, as this would occur from time to time under truly random conditions. The distractor sequences were AABAAABAAA, BAAAABAAAA, and AAAABAABAA, with the target sequence being presented between the first and second distractor sequences. New trials were identified via on-screen text, and each trial used a unique pair of bead colours.

For each sequence, the identity of the majority colour “A” and its association with the left or right-hand jar were pseudo-randomised. Participants were initially presented with a single bead, along with the query “Would you like to make a decision regarding which jar beads are being drawn from?”. Participants could either select “no, I would like to see another bead”, or “yes, I have made a decision”. Whenever participants requested another bead, the next bead in the sequence (to a maximum of ten beads) was displayed on screen, along with any previous beads to ensure responses were not affected by memory capacity (Freeman et al., 2014). When participants elected to make a decision, the jar they chose along with the DTD were recorded, and the sequence was stopped.[1] If no jar was chosen after the 10th bead, a DTD of 11 was recorded and the participant was progressed to the next sequence. This continued until participants completed ten trials of one sequence (target-sequence-only group) or ten trials of four sequences (distractor-sequences group).

2.3. Procedure

Participants accessed the online experiment via their own computer device, and were randomised to either the target-sequence-only condition or the distractor-sequences condition. After informed consent was established and basic demographic data were collected, participants were presented with detailed instructions explaining the beads task appropriate to their group, and a comprehension check consisting of a two-item test of their understanding of the task (see Experiment 10, Crump et al., 2013). If a participant failed the comprehension check it was repeated until a correct answer was recorded. Participants then completed ten beads-task trials.

Following completion of the beads task, the Peters Delusions Inventory (PDI, Peters et al., 2004) was administered (PDI data are not reported in this study), along with an attention check part-way through the 21 items. Participants then completed Likert scales (from 0 = definite disbelief to 4 = definite belief) indicating their belief in the non-swapping of beads-task jars mid-trial and in the randomness of

[1] Participants were not asked for their degree of confidence once they had made their decision.