Importance of controlling the frequency of lure foils in nback tasks

Deniz Mahmut Gün

University of Osnabrück

Importance of controlling the frequency of lure foils in nback tasks

Abstract

The nback task is a popular measure of working memory. Contrary to expectation it is only weakly correlated to another established measure of working memory. This report tries to draw attention to the effects of luring stimuli in the sequence, which is an often ignored and therefore uncontrolled parameter during experiments that deploy the nback task. Furthermore it tries to recreate the effects that have been demonstrated in previous studies using a different experimental design and an open source implementation of an nback task with controllable probabilities for the occurence of lure foils and targets.

Theoretical Background

With the vast array of available tasks and tests to assess psychometric constructs such as working memory capacity (WMC) it is difficult to choose the right task for the right hypotheses. (Wilhelm, Hildebrandt & Oberauer, 2013) The nback task as presented by Kirchner (Kirchner, 1958) has ever since remained a popular method to assess working memory related performance. Performance in different nback tasks correlate with performance in other tasks that are widely used as a measurement of fluid intelligence such as raven progressive matrices or the stroop task. An assessment by Jaeggie et. al even suggests that improvements in the nback task are transferable to fluid intelligence tasks. (Jaeggi et al., 2010)

nback

The nback task comes in different variants. What is common among all of them is that the participants are shown a sequence of stimuli and they need to indicate whether the current stimulus is the same as the one n trials back. It requires the participants to continually hold a specific sequence of stimuli in their memory, update it quickly and retrieve certain stimuli and their position accurately. For this reason, it is generally considered to have face validity as a measure for working memory. However, it only has a weak correlation with complex span tasks, which are an already established measure of working memory. This raises questions about the construct validity of nback tasks as working memory measure. (Kane, 2007)

The nback task is not without some rectifiable methodological drawbacks either that this report tries to draw attention to. Kane et al. have observed that 'lure' foils elicited more false alarms than control foils. (Kane, 2007) A lure is a stimulus that occurs at n-1 or n+1 steps ahead in an nback task. This makes it harder for the participants to reject it as false and thus baits them into making errors of commission. Although the successful computational model of nback task performance developed by Harbison et al. already predicts that accuracy is worse for lure foils than for other stimuli, it is not yet common custom to differentiate between the error probability of lure foils and non-targets.

WMC component of nback performance

Harbison et al. have pointed out that due to the low correlation with complex span tasks, the component of working memory that the nback task measures is not the capacity component. They observed that lures at n+1 had higher probabilities of being false hits, indicating that participants were unable to remove the stimulus from the partial sequence of stimuli that they have to maintain during the task.

Research Question

The results of the experiment presented in this paper are supposed to emphasize the importance of controlling the probability of lures when deploying the nback task as a measure of any psychometric construct. The question the experiment tries to answer is 'Does the frequency of lure foils in nback tasks negatively influence the overall performance measured in accuracy?' In other words, 'Do participants commit more errors on lure foils than on non-target stimuli?' Furthermore, some light shall be shed on the question, whether nback task performance depends on the executive control component of working memory rather than its overall capacity. This would be done by showing that the accuracy of lure foils at n+1 is not significantly lower than the accuracy for lure foils at n-1.

Hypothesis

The aim of this report is to test the hypotheses that accuracy is significantly worse for lure foils than other non-targets in a straight-forward manner. Formally.

The first hypothesis to test is whether accuracy for lures in general differs significantly from non lures. The second hypothesis is more specific. Here the hypothesis to test is that specifically the accuracy for stimuli at n+1 is significantly worse than for other non-target stimuli.

Whereas hypothesis H_a intends to recreate the effect observed by Kane (Kane, 2007), hypothesis H_b is a test of whether this experiment would yield results that support the conclusion that the nback task performance largely depends on the executive control component of working memory that was made by Harbison et al. (Harbison, Edu, Atkins & Dougherty, 2011)

Sampling Plan

The effect size estimate is a conservative guess due to the fact that no studies were found that allow for a good estimate of the effect size with the exact experimental setup of this study. Due to the large number of stimuli (200 per participant), 50 participants should yield about 10.000 data points, which is a conservative guess. The results can be used as reference for future sampling for a similar experimental setup.

Materials

A simple browser based implementation of the nback task with variable n and variable frequencies of targets and lure foils within the babe framework for browser based experiments will be used. Link to the implementation: https://github.com/denizmguen/xplab-project-nback

Procedure

First, participants who are well trained in the nback task are asked not to participate. Following the general instructions, every participant absolves practice trials with increasing difficulty (from n=3 to n=4). The participants are asked to repeat the practice until they feel confident. The practice trials are followed by two blocks with task difficulties n=3 and n=4 in random order. The first n stimuli are discarded for analysis. For both blocks $\frac{1}{9}$ of all stimuli are targets. Groups differ with regards to the frequency of lure foils. For the *test* block $\frac{1}{4}$ and for the *control* block $\frac{1}{9}$ of all stimuli are lures. Lure foils at n+1 and n-1 appear evenly often. Before each block, the participants are shown 20 practice stimuli to get accustomed to the new condition.

Variables

The dependent variable is accuracy. Accuracy is indirectly measured through the directly measured variable ’correctness’ and the total number of stimuli within each group. The independent variables are split into fixed and random effects. The fixed effects are the categorical variable ’stimulus type’ which has four levels: [’non-target’,’target’, ’lure n+1’, ’lure n-1’] and ’n’ which has two levels: [’3’, ’4’]. The random effects are ’subject’ since we expect that each subject performs differently and ’group’ to account for differences like exhaustion that could be tied to a higher frequency of lure foils.

Confirmatory Hypothesis Testing

Below α_N denotes the mean accuracy for non-target stimuli and α_L denotes the mean accuracy for lure stimuli. The first set of hypotheses posits that the mean accuracy for lure stimuli at positions n-1 and n+1 will be significantly lower than for other non-targets.

$$H_{0a} : \alpha_N = \alpha_{Ln-1},\ H_{1a} : \alpha_N > \alpha_{Ln+1}$$

$$H_{0b} : \alpha_N = \alpha_{Ln-1},\ H_{1b} : \alpha_N > \alpha_{Ln+1}$$

For the second part of the analysis we want to test the hypotheses that there is no significant difference in accuracy between lures at n-1 and n+1.

$$H_{0c} : \alpha_{Ln-1} = \alpha_{Ln+1},\ H_{1c} : \alpha_{Ln-1} \neq \alpha_{Ln+1}$$

Analysis Plan

For H_{1a} and H_{1b} a bayesian regression model will be fitted using the brms (Bürkner, 2017) package in R. Since we cannot simply presume equal variances for all stimulus types, the method of choice is Bayesian Robust Estimation. To be precise, Kruschke's robust model will be fitted with the following regression formula:

accuracy~stimulus type + n + (1|subject), sigma ~stimulus type

The Static Hamiltonian-MCMC method provided by the brms package is used to sample from the posterior distribution. An additive model is chosen since n only has two levels and there would likely be only minor negligible differences between the additive and multiplicative model. The additive model is preferred for its simplicity. In order to asses whether the participants really commit more errors on lure foils than on non-targets the probability that the coefficients for the parameters 'lure-n-1' and 'lure-n+1' are lower than the coefficient for 'non-targets' is calculated. If that should be the case for at least 95% of the samples from the posterior distribution, we'll accept H_1 and reject H_0.

For H_{1c} a two sided Kolmogorov Smirnov test will be performed on the samples of the posterior distribution.

Exclusion Criteria

A Rosner test will be applied with respect to the accuracy and reaction times to remove outliers from the data set. Other than that participants who have trained on the nback task are asked not to participate but due to the browser based nature of the experiment there are no reliable controls for exclusion prior to absolving the experiment.

References

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi:10.18637/jss.v080.i01

Harbison, I., Edu, J., Atkins, S. & Dougherty, M. (2011). N-back training task performance: Analysis and model.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J. & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, *38*(6), 625–635. doi:https://doi.org/10.1016/j.intell.2010.09.001

Kane. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3). doi:10.1037/0278-7393.33.3.615

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352. doi:10.1037/h0043688

Wilhelm, O., Hildebrandt, A. & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. doi:10.3389/fpsyg.2013.00433