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, on the overall accuracy. Furthermore it tries to recreate the effects that have been demonstrated in previous studies using a different experimental design and a open source implementation of a nback task with controllable lure and target parameter.

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 transfer 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 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 methodological drawbacks 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 stimuli ahead in a nback task. Making it harder for the participants to reject it as false and thus baiting them into making false hits. Although the successful computational model of nback task performance developed by Harbison et al. is predicated on the observation that the accuracy of is worse for lure foils than for other stimuli, it is not yet common custom to control the probability of lure foils.

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 inable to succesfully remove the stimulus from the partial sequence that they need 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 probability of lures in nback tasks significantly influence the performance measured by the accuracy of the participant's answers?' Furthermore more light shall be lit on the question wether nback task performance depends on the executive control component of working memory rather than its overall capacity.

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

In order to determine the sample size, the power-cohen's d table (**cohen1977**) is consulted.

An α value of .8 is chosen and the effect size is estimated to be around .5.

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.

Thus the sample size is 64 for the test-group and control group.

Materials

A simple browser based implementation of the nback task with variable n within the babe framework for browser based experiments will be used. The stimuli will be integers from 1 to 9. In the control group, the probability for lure foils and targets is just at $p = \frac{1}{9}$. For the test group it is 0.1 for targets and 0.25 for lures. Thus a fourth of all stimuli are lures. Link to the implementation:
https://github.com/denizmguen/xplab-project-nback

Procedure

First participants who are well trained in the nback task will be asked not to participate. Following the general instructions, every participant absolves practice trials with increasing difficulty (from n=1 to n=4). The participants are asked to repeat the practice until they feel that they understood how the task works.

Now follow two blocks with task difficulties n=3 and n=4 in random order. The first n stimuli are discarded for analysis. For both blocks 25% of all stimuli are targets. The two blocks differ with regards to the frequency of luring stimuli. For the ıtest block 0.25% and for the ıcontrol block, 0.11% of all stimuli are lures.

Before each block, the participants are shown 20 practice trials with the respective n to get accustomed to the condition.

Variables

The directly measured variables are the correctness of an answer and reaction time. The relevant independent variable is the stimulus type.

The reaction times are used for later exclusion of participants. Using the stimulus type and correctness the accuracy for different kinds of stimuli can be calculated.

Analysis Plan

The test group has an increased probability for lure foils to make the detection of an effect on the accuracy easier. The control group serves the purpose to show that the difference in overall accuracy in the test group is not random or due to unobserved effects like exhaustion.

The analysis has two parts: First the difference in accuracy between normal stimuli (non-lures) and lure foils is assessed using a t-test. Rows are first grouped by stimulus type, then averaged for each participant to obtain a table where there are 2 rows for each participant:accuracy for lures and accuracy for non lures.

In the second part the hypotheses is tested that lures at n+1 are significantly less accurate than non-lure stimuli. Again the stimuli are grouped by stimulus type, such that for each participant remain two rows: accuracy for n+1 lures and accuracy of non lures. A t-test is applied to assess whether the mean accuracy differs significantly for the stimulus types.

Exclusion Criteria

A rosner test will be applied with respect to the accuracy and reaction times to remove outliers from the data set.

Confirmatory Hypothesis Testing

$$H_{0a} : \alpha_N = \alpha_L$$

$$H_{1a} : \alpha_N < \alpha_L$$

where α_L and α_N denote the mean accuracy of lure and non-target stimuli respectively. But there is also a second, more specific hypothesis to be tested.

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

$$H_{1b} : \alpha_N < \alpha_{Ln+1}$$

α_{Ln+1} denotes the accuracy of lures at n+1.

References

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