

A Formal Framework for Structured N-Back Stimuli Sequences

Morteza Ansarinia (morteza.ansarinia@uni.lu)

University of Luxembourg,
Research Unit Education, Culture, Cognition & Society (ECCS)
L-4366 Esch-sur-Alzette, Luxembourg

Dominic Mussack (dominic.mussack@uni.lu)

University of Luxembourg,
Research Unit Education, Culture, Cognition & Society (ECCS)
L-4366 Esch-sur-Alzette, Luxembourg

Paul Schrater (schrater@umn.edu)

Depts. of Computer Science and Psychology,
University of Minnesota, Twin Cities
Minneapolis, MN, USA

Pedro Cardoso-Leite (pedro.cardosoleite@uni.lu)

University of Luxembourg,
Research Unit Education, Culture, Cognition & Society (ECCS)
L-4366 Esch-sur-Alzette, Luxembourg

Abstract

Numerous cognitive tasks, like the n-back, employ sequences of stimuli to target particular cognitive functions. These sequences are generated to satisfy specific criteria but the generation process typically induces unintentional statistical structure in the sequences which may not only affect performance but also alter the strategies participants use to complete the task.

Here we propose that the generation of stimulus sequences can be conceptualized as a soft constraint satisfaction problem and offer experimental evidence demonstrating the impact of local sequence features on human behavior. Our approach to sequence generation provides a means to better control and assess sequence structures, which in turn could help clarify the cognitive and neural processes involved in cognitive tasks.

Keywords: structural features; cognitive strategies; n-back; constraints satisfaction

Introduction

With more than 1600 hits on PubMed, the n-back task is one of the most popular tasks in cognitive psychology today. It is widely used not only to evaluate working memory capacity but also as a training protocol to improve working memory and possibly fluid intelligence (Au et al., 2015; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008). In the n-back task, participants are presented a sequence of stimuli and have to determine for each stimulus if it matches or not the stimulus presented n -steps ago. Stimuli that match are called "targets", those that don't match are called "distractors", with close misses (i.e., distractors that would be targets under a slightly different N) are called "lures". While the task is widely considered a working memory task, it does not correlate well with other "gold-standard" working memory tasks, such as the complex span

task (Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Miller, Price, Okun, Montijo, & Bowers, 2009).

Previous studies have raised concerns that the n-back task may be solved using multiple strategies, not all of which rely purely on working memory processes (Ralph, 2014). There are numerous variants of the n-back task, but even within a variant participants could use various strategies. One source of variation in the n-back task that is potentially biasing participants' strategies are the statistical properties of the sequences of stimuli used for n-back task, which are typically uncontrolled for and differ across studies (Braver, 2012). For example, (Ralph, 2014) showed that various statistical properties of n-back sequences may favor a reactive cognitive control strategy whereby people's performance relies on detecting stimulus familiarity rather than on active information updating in working memory. Because statistical properties of stimulus sequences seem to bias cognitive control strategies and hence cause heterogeneous behavioral and neurophysiological outcomes it is necessary to characterize those statistical properties and develop methods to generate adequate sequences.

Here, we propose an approach that allows researchers to parameterize interesting features of the n-back sequences which may affect behavior. We then evaluate the predictive effect of such uncontrolled parameters on behavioral outcomes. Results from this research may have implications on the way the n-back task is put into practice to study working memory or improve cognitive skills. While our focus here is on the n-back, the principles presented below apply to a broader range of cognitive paradigms.

Parameterizing the N-Back Sequences

While n-back sequences are usually thought of as an ordered set of i.i.d. generated and sequentially independent stimuli, in



practice the sequences of stimuli are neither objectively nor subjectively independent. Objective local structure is introduced by design constraints like a fixed number of target or stimulus set size, while subjectively people are highly sensitive to local sample structure in sequences. For example, unconstrained sampling from a uniform distribution to generate sequences may lead to frequent local repetitions of stimuli (i.e., "lumpiness"; Abelson (1995)). In the n-back task, such local patterns could encourage people to identify targets solely based on stimulus familiarity rather than to use their working memory, as this strategy may in this case lead to high performance at low cognitive cost. Here we define a few basic measures known to be important for the perception of local structure in sequences, and show how to use these measures to parameterize families of sequences.

N-Back sequences are typically generated by randomly sampling M stimuli from a vocabulary set V (e.g., a set of 8 letters) with the constraint of having a specific number of targets (T) in the sequence, given the fixed value of N for the intended n-back version. Researchers typically manipulate N and T to study behavioral and neural correlates of working memory; other parameters are treated as nuisance variables.

A common procedure to generate n-back sequences involves two steps: first a sequence of stimulus-role placeholders (e.g., D =distractor, T =target) is generated; then particular stimuli are sampled from the vocabulary to fulfill those roles. For example, the first step might generate the sequence "DDDDTDTDT" while the second step would instantiate particular stimuli (e.g., "ABCEDEA"). Generating n-back sequences using this procedure is problematic however because the resulting sequences are typically highly skewed with some stimuli being presented much more frequently than others and frequently presented stimuli having a higher probability of being targets (Ralph, 2014). Moreover, lures are more likely to trigger false alarm responses and to require proactive control processes.

The lack of control for parameters such as lures and lumpiness may compromise results interpretations and generate scientific confusion because such parameter may affect cognitive strategies and consequently increase behavioral and neurophysiological data heterogeneity (Juvina & Taatgen, 2007). Ralph (2014) urged researchers to carefully control frequency distribution of stimuli, stimulus repetition, the fraction of targets and the fraction of lures, and the number of different stimuli in the vocabulary set in order to have a better handle on cognitive strategies. However, generating sequences that fulfill multiple criteria may not always be possible or practical using standard, brute-force approaches; there might for instance be cases where no such sequence exists. Furthermore, future research may require the addition or removal of criteria and such changes would typically require rewriting new sequence generators.

In the following section we conceptualize the generation of structured sequences for the n-back as a constraint satisfaction problem. This approach has several key advantages: a)

it provides an implementation blueprint that accommodates a wide range of use cases b) it supports the softening of constraints to ensure approximate solutions can be found within a practical timespan; c) it supports compositional control of constraints that is well suited for hypothesis testing and d) by taking advantage of the Maximum Entropy optimization framework and Conditional Random Fields model, it is possible to move from an intuitive definition of constraints to the space of probability distributions that are invaluable for modeling and data analysis (Batou & Soize, 2013).

Structured Sequences

A sequence is an ordered set of M stimuli sampled from a vocabulary of V stimuli that satisfies specific criteria. A sequence of stimuli that (approximately) satisfies a set of specific constraints on parameters or features is a qualified sequence.

The problem of generating a qualified sequence can be reduced to a soft constraint satisfaction problem, P :

$$P = \langle X, D, C, W \rangle$$

where X is a set of structural variables to be controlled (see Table 1), D is the set of distributions over the variables, C is the set of constraints expressed as expected values for X (see Table 2), and W is a cost function that uses the constraints to map a sampled sequence to a real value (Table 2); it represents the degree to which a particular sequence violates the constraints in C . Generating a qualified sequence for the n-back task can be formulated as minimizing the aggregated cost of violating the constraints. Note that some constraints in the n-back task cannot be relaxed; for example, constraints which includes the expected value of the N , must be fully satisfied for the sequences to be valid.

We have argued that sequence structure may affect cognitive performance and that consequently such features need to be controlled. We argued for the use of the constraint satisfaction framework as a principled approach to evaluate and generate qualified sequences. This approach operates on structural variables which may or may not affect human behavior and thus may or may not require stringent control.

To evaluate the relevance of the structural variables highlighted above for the n-back task we will analyze an existing dataset which did not explicitly manipulate or control for these structural variables. If these structural variables are informative about participants' n-back performance it follows that they are scientifically relevant and should be explicitly listed and constrained for both sequence generation and performance evaluation.

Evaluating Behavioral Impacts of Structural Features

Dataset

We used a previously published n-back dataset from Cardoso-Leite et al. (2016). This dataset contains n-back data from 60 healthy adults ($M=20.68$, $SEM=0.42$) completing both the 2-back and 3-back versions of the n-back paradigm. For each

Table 1: Structural Variables (X)

Structural Variables	
x_N	N , number of trials to look back for a target.
x_t	Targets ratio describes the number of target trials in a sequence regardless of the stimulus.
x_s	Skewness is maximum deviation of stimuli frequency from uniform distribution.
x_l	Lures ratio represents the number of distractors which would be targets for $N - 1$ or $N + 1$.
x_v	Vocabulary size is the number of all unique stimuli to be presented.
x_{tl}	Recent targets ratio represents the number of targets in recent trials.
x_{ll}	Local lures ratio describes the number of lures in recent trials.
x_{vl}	Local vocabulary size is the number of unique stimuli presented in recent trials.
x_{ul}	Lumpiness is maximum number of repetitions in a sequence.
x_{sl}	Local skewness is the number of unique items shown in recent trials.
x_g	Gap is the number of trials since the last time same stimulus was appeared.

Table 2: Constraints (C)

Constraint	Violation Cost (W)
$E[x_n] = N$	$W_n \sim \begin{cases} 0 & x_n = N \\ \infty & x_n \neq N \end{cases}$
$E[x_t] = T \times trials$	$W_t \sim 1 - \mathcal{N}(T \times trials, 1)$
$E[x_{tl}] = \frac{T \times w}{trials}$	$W_{tl} \sim 1 - \mathcal{N}(\frac{T \times w}{trials}, 1)$
$E[x_l] = L \times trials$	$W_l \sim 1 - \mathcal{N}(L \times trials, 1)$
$E[x_{ll}] = \frac{L \times w}{trials}$	$W_{ll} \sim 1 - \mathcal{N}(\frac{L \times w}{trials}, 1)$
$E[x_v] = V $	$W_v \sim 1 - \mathcal{N}(V , 1)$
$E[x_{vl}] = \min(V , w)$	$W_{vl} \sim 1 - \mathcal{N}(\min(V , w), 1)$
$E[x_{ul}] = w$	$W_{ul} \sim 1 - \mathcal{N}(w, 1)$
$E[x_s] = \frac{trials}{ V }$	$W_s \sim 1 - \mathcal{N}(\frac{trials}{ V }, 1)$
$E[x_{sl}] = \max(1, \frac{w}{ V })$	$W_{sl} \sim 1 - \mathcal{N}(\max(1, \frac{w}{ V }), 1)$
$E[x_g] = \frac{trials}{w}$	$W_g \sim 1 - \mathcal{N}(\frac{trials}{w}, 1)$

version participants completed 3 sequences of 30 trials each which results in a grand total of 360 n-back sequences and 10800 trials. On each trial, stimulus identity, reaction time and accuracy were recorded. For more details about this dataset, see Cardoso-Leite et al. (2016).

Data Analysis

To evaluate the need to control for structural variables we fit and contrast two nested models that predict participants accuracy on a trial-by-trial basis, using a different set of predictor variables.

The *base model* uses the common approach of relating performance to descriptors of the sequence as a whole (i.e., x_n , x_v , and x_t) as well as the current stimulus (i.e., target or distractor) to predict the accuracy of the response to the current stimulus.

The *extended model* includes in addition all the structural variables listed in Table 1 (e.g., x_l , x_{ll} , x_s). These structural variables are computed not on the sequence as a whole but rather on the recent stimulus history (8 previous stimuli, excluding the current stimulus). This approach exploits local variation along the dimensions of the structural variables to evaluate the impact of those variables on accuracy.

The data was subdivided into a training (80%) and a test set (20%). Both models were fit to the same training set using the imbalanced Partial Least Squares (PLS) method; this method was chosen because most responses were correct (92%) and the predictor variables are not mutually independent. Both models were then evaluated by their ability to account for test data using the area under the curve (AUC) as the model performance metric. The reliability of the AUC was further characterized using bootstrapping (1000 repetitions).

Two main conclusions can be drawn if the extended model outperforms the base model: a) structural variables affect behavior and hence need to be controlled by the sequence generator, b) even when they are controlled at the level of a sequence as a whole, local variations in structural variables may already be enough to affect behavior and it might be necessary to use trial-by-trial estimates of local properties to analyze human behavior and brain activity.

Results

Figure 1 shows the ROC curves for the two fitted models. The base model predicts response accuracy above chance level ($AUC=59.51$; $CI_{95\%} = [54.81, 64.21]$). The addition of structural variables as predictors in the extended model improves model performance substantially ($AUC=68.56$; $CI_{95\%} = [65.76, 71.36]$).

To determine which variables drive the performance accuracy of the extended model, we ran a model-based variable importance analysis using the *Boruta* package in R (Kursa & Rudnicki, 2010). These importance scores were calculated using random forest method alongside shadow features, which are copies of original features but with randomly replaced values; this serves to remove the importance of a fea-

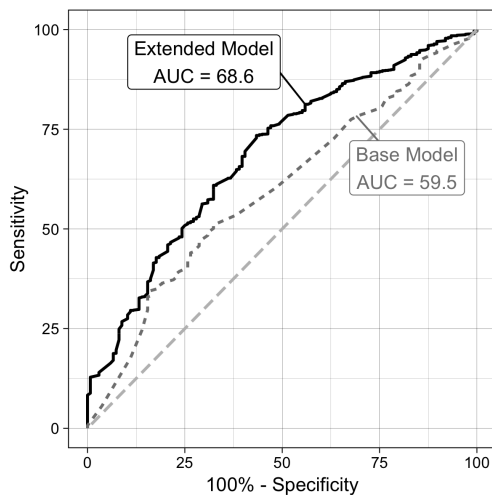


Figure 1: Accuracy classification performance for the base and extended models. AUC = Area Under the Curve

ture while nevertheless maintaining their distribution of values unchanged.

This analysis shows that the structural features computed on the recent history contributes most to the predictability of participants accuracy. Figure 2 shows the relative importance of the predictor variables used by the extended model.

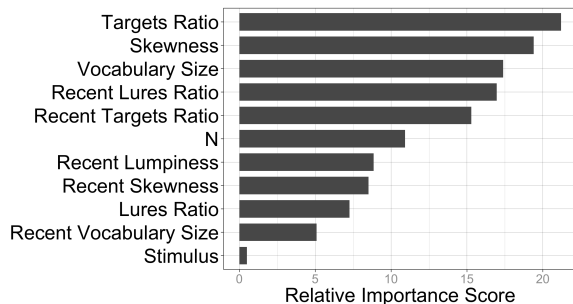


Figure 2: Relative importance of structural variables (V) on the prediction of participants response accuracy.

Although a direct causal relationship cannot be inferred from the results, higher contribution of recent trials in the extended model (i.e., higher relative importance of x_{vI} , x_{lI} , x_{lII} , and x_{sI} than their global counterparts, x_v , x_l , x_{lI} , and x_s) suggests that behavioral responses are partially guided by a more fine-grained set of structural features.

Conclusion

In sum, we propose a compositional framework to parameterize and exploit interesting features of the n-back sequences and evaluated behavioral effects of the features of random sequences. We developed two predictive models to compare the importance of these structural features.

Methods that are commonly used to generate n-back sequences use independent random sampling for each trial and cannot control all the influential features. Instead of an independent random sampling process, we proposed a framework to reformulate generating the n-back sequences as a soft constraint satisfaction problem. This approach can be used to formalize the effect of structural patterns in other cognitive tasks that present random sequences of stimuli.

Acknowledgments

This research was supported by the Luxembourg National Research Fund (ATTRACT/2016/ID/11242114/DIGILEARN) and (INTER Mobility/2017-2/ID/ 11765868/ULALA).

References

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 22(2), 366-377.

Batou, A., & Soize, C. (2013). Calculation of Lagrange Multipliers in the Construction of Maximum Entropy Distributions in High Stochastic Dimension. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1), 431-451.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106-113.

Cardoso-Leite, P., Kludt, R., Vignola, G., Ma, W. J., Green, C. S., & Bavelier, D. (2016). Technology consumption and cognitive control: Contrasting action video game experience with media multitasking. *Attention, Perception, & Psychophysics*, 78(1), 218-241.

Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829-6833.

Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394-412.

Juvina, I., & Taatgen, N. (2007). Modeling control strategies in the N-back task. In *Proceedings of the eight International Conference on Cognitive Modeling* (p. 73-78). Psychology Press.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11).

Miller, K., Price, C., Okun, M., Montijo, H., & Bowers, D. (2009). Is the N-Back Task a Valid Neuropsychological Measure for Assessing Working Memory? *Archives of Clinical Neuropsychology*, 24(7), 711-717.

Ralph, J. (2014). *Statistical manipulation and control strategies of the n-back task*. (PhD Thesis). Rensselaer Polytechnic Institute, Troy, New York.