# Bayesian modeling of human performance in a visual processing training software

Julien Diard, Muriel Lobier, Sylviane Valdois

## HAL Id: hal-00530368
## https://hal.archives-ouvertes.fr/hal-00530368

# Bayesian modeling of human performance in a visual processing training software

*Julien Diard, Muriel Lobier, Sylviane Valdois*

Laboratoire de Psychologie et NeuroCognition - CNRS
Université Pierre-Mendès-France, BSHM, BP 47, 38040 Grenoble, France

## ABSTRACT

Dyslexia is a deficit of the identification of words, which is thought to be a consequence of different possible cognitive impairments. Recent data suggest that one of these might be a specific deficit of the visual attention span (VAS). We are developing a remediation software for dyslexic children that focuses on the VAS and its training. A central component of this software is the estimation of the performance of a given participant for all possible exercises.

We describe a preliminary probabilistic model of participant performance, based on Bayesian modeling and inference. We mathematically define the model, making explicit underlying generalization hypotheses. The model yields a computation of the most probable predicted performance space, and, as a direct extension, an exercise selection strategy.

**Keywords**: Bayesian modeling, dyslexia, user modeling, human performance.

# INTRODUCTION

Developmental dyslexia is a specific learning disability characterized by a deficit in word identification. A dyslexic child is unable to acquire basic reading skills despite adequate intelligence, education, and sensory abilities. Developmental dyslexia is considered to be a cognitive disorder, this cognitive disorder being a consequence of an underlying neurobiological dysfunction.

For many years, the only recognized theory explaining developmental dyslexia was the phonological theory. It states that dyslexic children have a specific impairment of phonological skills: because of such a deficit, dyslexic children are unable to appropriately segment a word into single sounds and to link these sounds to the appropriate letters (Vellutino et al. 2004). However, some cases of developmental dyslexia are clearly not phonological and there is emerging evidence that visual attention might play a critical role in this disorder (Boden & Giaschi, 2007; Vidyasagar & Pammer, in press).

A key cognitive skill to fluent reading is the ability to recognize and process several letters in the same fixation. Data has shown that a sub-group of dyslexic children has significant difficulties in identifying a sufficient amount of letters in the short time frame of a fixation. This finding, together with the Multi-Trace Memory (MTM) reading model (Ans et al., 1998), has led to the visual attention span deficit theory of dyslexia (Bosse et al., 2007; Valdois et al., 2004). The visual attention span (VAS) is defined as the number of visual elements that can be processed simultaneously. It is measured using a global report task. In this task, a 5-letter consonant string is displayed during 200 ms. Subjects need to verbally report all the letters they have identified. Performance in this task is both reliably correlated with reading performance across all primary grades (Bosse et al., 2009) and significantly lower for a sub-group of dyslexic children than for age-matched controls (Bosse et al., 2007). Based on Bundesen (1990)'s theory of visual attention, a recent case study (Dubois et al., in press) of two dyslexic children has linked two potential cognitive sources to this VAS deficit: a reduced visual information processing rate or a limited number of items that can be stored in visual short term memory.

This insight on the specific cognitive components that are linked to the VAS is central in order to develop a targeted training regimen. We are developing evidence-based software that aims to train the deficit in visual information processing rate. The goal is to improve the reading performance of dyslexic children.

In this paper, we describe a preliminary probabilistic model of participant performance using Bayesian modeling and Bayesian model comparison. We provide the mathematical definition of the model that makes explicit all the underlying generalization hypotheses, contrary to previous approaches. We show

that the model, for a given set of experimental observations, yields a computation of the most probable predicted performance space. It also yields, as a direct extension, a computation of the probability distribution over exercises to propose to participants. In other words, it naturally provides an exercise selection strategy that is optimal with respect to current observations.

# VISUAL ATTENTION PROCESSING TRAINING SOFTWARE

The training software is designed to be used daily on a home computer. The typical training regimen calls for six training sessions per week, each session lasting 20 min. During these sessions, the subject has to perform visual processing tasks on multi-element arrays. Up to 150 "trials" will thus be performed in each training session. The two main characteristics of this software are the specificity of its tasks and the adaptability of the presented exercises. The tasks were designed to tap specifically visual processing of multi-element arrays. Three different exercise parameters are used to modulate the difficulty of each trial presented to the subject. For the purpose of this paper, we do not delve in depth on task specificity but concentrate on the topic of trial adaptability.
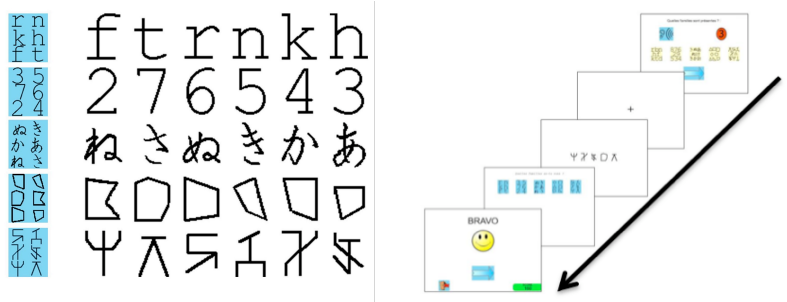


Figure 1. Left: families of characters used in the software. Right: succession of screens presented in any given trial, from the instruction screen to the feedback screen.

The training tasks all involve the visual categorization of both alphanumeric and non-alphanumeric character arrays. Five different character categories or "families" are thus defined: letters, digits, hiragana characters, polygons and pseudo-letters (see Fig. 1, left). All tasks involve the identification of the visual category of one or more characters. We now describe the sequence of events of a single trial.

The following screens are successively displayed during a single trial:

instruction screen, fixation screen, stimulus screen, answer screen and feedback screen (see Fig. 1, right). The instruction screen displays the specific categorization task to be carried out in the current trial. On the stimulus screen an array of 2 to 7 characters belonging to one or two visual categories is displayed during 120 to 420 ms.

Three difficulty parameters characterize exercises, each of which can take 6 values. The character array can hold 2, 3, 4, 5, 6 or 7 elements. The display duration can be of 420, 360, 300, 240, 180 or 120 ms. The 6 different tasks are: 1- Is there an element of this family in the display? 2- How many families are there in the display? 3- How many elements of this family are there in the display? 4- Which two families are present in this display? 5- How many elements of these two families are present in the display? 6- Which two families are present in the display and how many elements of each are there? These difficulty parameters are ordered by increasing difficulty, i.e. task number 3 is easier than task number 6 but harder than task number 1.

These three difficulty parameters, numbered from 1 to 6, are the three dimensions of a matrix (number of elements, display duration, task), which we call the performance space. We call "exercise" a particular triplet in this space: for example, the coordinates (2, 2, 3) denote an exercise in which the subject will be asked how many elements of a given category are displayed (task = 3) and 3 characters will be displayed (number of elements = 2) during 360 ms (duration = 2). Exercises with small coordinates are easy whilst exercises with large coordinates are harder.

## BAYESIAN MODELING OF HUMAN PERFORMANCE

A central component of the remediation software is the estimation of the performance of a given participant for all possible exercises. Its use is two-fold. Firstly, it allows us to define the exercise selection strategy, that is to say, the strategy used to automatically select the next exercise to propose to the participant. An exercise with a predicted success rate of 75% is assumed to be optimal. Indeed, it is both easy enough to maintain motivation and hard enough to drive learning effectively. Secondly, a correct estimation of the performance of a subject allows us to track, over time, the overall increase of performance and thus to assess the quality of the remediation process.

A previous approach, in the context of dyscalculia (Wilson et al., 2006), was based on heuristic estimation of human performance. In other words, an algorithmic solution was developed that tracked the subjects' performance. This solution was unable to correctly converge to the assumed representation: in simulation, the actual performance space was a cuboid but the estimation would not converge to a cuboid.

We now present a probabilistic model, based on Bayesian Programming (Lebeltel et al., 2004; Bessière et al., 2008), that solves this issue. More precisely, the model includes an explicit hypothesis about the performance space of participants. This hypothesis constrains the recognition algorithm: if the model computes the most probable cuboid, the output is guaranteed to be a cuboid.

## BAYESIAN MODEL DEFINITION

The model relates the performance space of a participant at a point in time with his/her successes or failures for presented exercises.
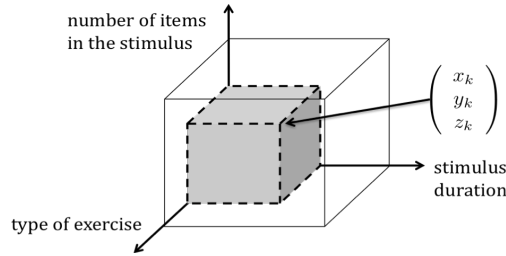


Figure 2. The performance space of a participant is a cuboid inscribed in the cube of all possible exercises, described by the point $K = (x_k, y_k, z_k)$.

We firstly assume that performance spaces are represented by a single point of the performance matrix $M$ (see Fig. 2). At this point $K = (x_k, y_k, z_k)$, the success rate is midway between 100% and chance level (denoted $c$). We further assume that exercises that are easier than $K$ (inside the cuboid) have higher success rates, and exercises harder than $K$ have lower success rates: more precisely, we assume that success rates vary continuously, following 3 sigmoid functions of (unique) parameter $\alpha$ that are successively applied to each dimension (following Wilson et al. (2006)):

$$M_{x_k,y_k,z_k}(x,y,z) = 1/\sqrt[3]{e^{\alpha(x-x_k)}e^{\alpha(y-y_k)}e^{\alpha(z-z_k)}}$$

In probabilistic terms, this translates into the $\pi_P$ model. Let $(x^{0:T}, y^{0:T}, z^{0:T})$ be the trials from time 0 to time $T$, and $S^{0:T}$ be a set of boolean variables that describe whether these trials resulted in successes or failures. Assuming that trials are independent and identically distributed over time, the resulting decomposition of the joint probability distribution is:

$$P(x_k \ y_k \ z_k \ x^{0:T} \ y^{0:T} \ z^{0:T} \ S^{0:T} \mid \pi_P)$$

$$= \ P(x_k \ y_k \ z_k \mid \pi_P) \prod_{i=0}^{T} P(x^i \ y^i \ z^i \mid \pi_P) P(S^i \mid x^i \ y^i \ z^i \ x_k \ y_k \ z_k \ \pi_P)$$

In this decomposition, $P(x_k \ y_k \ z_k \mid \pi_P)$ and $P(x^i \ y^i \ z^i \mid \pi_P)$ are assumed to follow uniform probability distributions. The last term, $P(S^i \mid x^i \ y^i \ z^i \ x_k \ y_k \ z_k \ \pi_P)$ is the prediction term, in the sense that given a supposed performance space and an exercise, it predicts the corresponding success rate. It is computed by applying the 3-dimensional sigmoid function of parameter $\alpha$, according to the distance between $K = (x_k \ y_k \ z_k)$ and the presented exercise $x^i \ y^i \ z^i$.

## HUMAN PERFORMANCE EVALUATION

The model being defined, it can now be used to recognize the performance space of a participant given observations of trial results. This is done by computing:

$$P(x_k \ y_k \ z_k \mid x^{0:T} \ y^{0:T} \ z^{0:T} \ S^{0:T} \ \pi_P)$$

$$\propto \ P(x_k \ y_k \ z_k \mid \pi_P) \prod_{i=0}^{T} P(x^i \ y^i \ z^i \mid \pi_P) P(S^i \mid x^i \ y^i \ z^i \ x_k \ y_k \ z_k \ \pi_P)$$

$$\propto \ \prod_{i=0}^{T} P(S^i \mid x^i \ y^i \ z^i \ x_k \ y_k \ z_k \ \pi_P)$$

In other words, maximizing the probability over performance spaces $K$ is reduced to maximizing the likelihood of the observed data $S^{0:T}$. In order to avoid that the product of probabilities degenerates to numerical zeroes, the log of the likelihood is evaluated and maximized. This transforms the product of probabilities into a sum of log probabilities.

## EXPERIMENTAL RESULTS

We evaluate our recognition algorithm in simulation: we define a "true" performance space $K$ and use it to simulate the results $S^{0:T}$ of $T$ trials. We then use these observations first to compute the probability distribution over $K$, as described above, and afterwards to maximize this probability in order to output an estimate $\hat{K}$ of $K$. To measure the quality of our algorithm, we compute the recognition error as the Manhattan distance between $\hat{K}$ and $K$. Fig. 3 (left) shows an example of recognition errors as the number of simulated trials $T$ increases: obviously, adding observations reduces the recognition error. Using the actual software and dyslexic children, in a typical twenty-minute remediation session, around 100 exercises can be presented. This would yield, on average, and assuming that $\pi_P$ is an adequate model, an error around 1. This is fairly acceptable, as it amounts to correct recognition along two dimensions and an error of 1 in the last dimension.
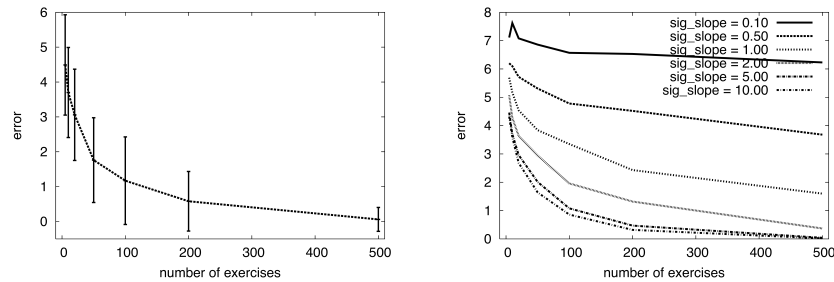
Figure 3. Left: error in the performance space recognition as a function of the number of trials (with sigmoid slope $\alpha$ = 5.0; mean and standard-deviations over 100 simulations). Right: errors for different values of the sigmoid slope $\alpha$.

However, Fig. 3 (right) shows the impact of the $\alpha$ parameter on the difficulty of the problem: when $\alpha$ is high, predicted success rates vary quickly. The problem is easy and the recognition algorithm's error is low with few trials. On the other hand, when $\alpha$ is low, predicted success rates vary slowly, which makes the problem harder and limits the convergence of the algorithm.

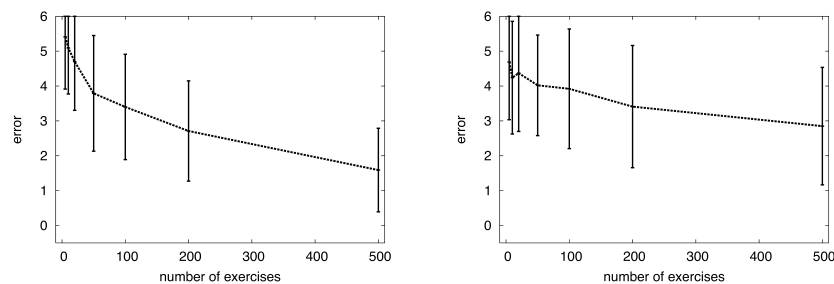## USING PERFORMANCE SPACE RECOGNITION FOR EXERCISE SELECTION



Figure 4. Left: error in the participant recognition as a function of the number of trials (sigmoid slope $\alpha$ = 1.0; 100 simulations), with exercises chosen at random. Right: Same configuration, but exercises are chosen, at each time step, to be at the current most probable $(x_k, y_k, z_k)$.

Another major aspect of the simulation has an impact on the quality of the recognition algorithm. It is the strategy for selecting exercises. Indeed, whereas

previous simulations used purely random selection, the goal is to use the recognition algorithm in order to present exercises with a predicted success rate around the one at $K$ (easy enough to maintain motivation and hard enough to drive learning effectively). However, it can be experimentally observed that such an exercise selection makes the recognition more difficult (see Fig. 4). Indeed, a random selection strategy "explores" the space of possible exercises, whereas using exercises around the estimated $K$ concentrates the trials in a narrow portion of the space, possibly slowing convergence in case of erroneous initial estimates.

This effect is most obvious for $\alpha$ parameter values that correspond to difficult configurations. Indeed, in easy cases ($\alpha = 5.0$ for instance), the difference in convergence rate is only marginal.

There is a Bayesian answer to this issue. The exercise selection strategy can be added to the probabilistic model, in the form of a $P(x^i\ y^i\ z^i \mid x_k\ y_k\ z_k\ \pi_P)$ term. Then, instead of explicitly having to compute an estimated $K$ value, the uncertainties of $P(x_k\ y_k\ z_k \mid x^{0:T}\ y^{0:T}\ z^{0:T}\ S^{0:T}\ \pi_P)$ could be propagated by summing over $K$:

$$
P(x^{T+1}\ y^{T+1}\ z^{T+1} \mid x^{0:T}\ y^{0:T}\ z^{0:T}\ S^{0:T}\ \pi_P)
$$
$$
= \sum_{x_k, y_k, z_k} \left( \begin{array}{l} P(x^{T+1}\ y^{T+1}\ z^{T+1} \mid x_k\ y_k\ z_k\ \pi_P) \\ \times P(x_k\ y_k\ z_k \mid x^{0:T}\ y^{0:T}\ z^{0:T}\ S^{0:T}\ \pi_P) \end{array} \right)
$$

Initially, after few trials ($T$ small), the recognition over $K$ would still yield high uncertainties in $P(x_k\ y_k\ z_k \mid x^{0:T}\ y^{0:T}\ z^{0:T}\ S^{0:T}\ \pi_P)$; the position of $K$ would be uncertain and therefore the selected trial would be as if drawn at random. After more observations are gathered ($T$ large), the recognition error would be low and $P(x_k\ y_k\ z_k \mid x^{0:T}\ y^{0:T}\ z^{0:T}\ S^{0:T}\ \pi_P)$ would be fairly peaked, leading to selecting exercises in the close neighborhood of this peak. As a consequence, this model would gradually shift from a random exploration of possible exercises to a selection of exercises as described by the $P(x^i\ y^i\ z^i \mid x_k\ y_k\ z_k\ \pi_P)$ term. In other words, it would automatically shift from initial calibration to an adequate training and remediation program.

## MODEL EXTENSIONS

The model we have presented is a basis for possible extensions. For instance, we have so far considered the $\alpha$ parameter as an internal parameter to the model but it could be explicitly handled in a probabilistic manner. We could estimate its value, or, in a more Bayesian fashion, propagate the uncertainties about it in the computations in a principled manner (as shown previously for the exercise selection strategy).

We detail a similar idea in another context. Instead of the $\alpha$ parameter, consider the shape of the performance space we have assumed. The founding hypothesis of the $\pi_P$ model is a cuboid shape describing the way trials are correctly or incorrectly answered. In order to experimentally assess the validity of this assumption, a
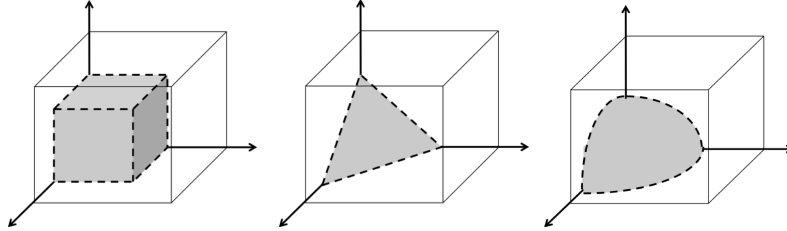
hierarchical Bayesian model can be defined.



Figure 5. Alternative models of human performance that can be formally studied using Bayesian model comparison.

The first step is to define alternative models, based on different assumptions. For instance, instead of a cuboid, assume a tetrahedron or spherical section shape in the space of exercises (see Fig. 5). This is easily translated into corresponding $\pi_T$ and $\pi_S$ models, with different prediction terms $P(S^i \mid x^i\, y^i\, z^i\, x_k\, y_k\, z_k\, \pi_T)$ and $P(S^i \mid x^i\, y^i\, z^i\, x_k\, y_k\, z_k\, \pi_S)$ and joint probability distributions. Next, create a meta-variable, that considers all possible models: $M = \{\pi_P,\, \pi_T,\, \pi_S\}$, and encapsulate the three models in:

$$P(x_k\, y_k\, z_k\, x^{0:T}\, y^{0:T}\, z^{0:T}\, S^{0:T}\, M)$$
$$=\quad P(M)P(x_k\, y_k\, z_k\, x^{0:T}\, y^{0:T}\, z^{0:T}\, S^{0:T} \mid M)$$

The term $P(M)$ is a prior over models, which can be assumed uniform. Then, this model can be used to compute a probability distribution over models given experimental observations $P(M \mid x^{0:T}\, y^{0:T}\, z^{0:T}\, S^{0:T})$ in order to select the most probable model. This model would select the hypothesis that best describes the performance space of a given participant. Alternatively, as previously, uncertainties over the unknown variable $M$ could be propagated with computations that would involve a summation over $M$.

## CONCLUSION

We have presented a dyslexia remediation software and the component that tracks participants' performance and selects exercises. It is based on Bayesian Programming and inference. We have presented and discussed the design of the exercise selection strategy: Bayesian inference and summation over unknown variables theoretically yields a gradual shift from initial random exploration to exploitation of the optimal exercise to practice. We have outlined the extension of our model to hierarchical model comparison, in order to assess the quality of underlying assumptions.

Another natural extension would be to expand the model so as to track, over time, the displacement of the point $K$ that describes participant performance. Technically, it would involve transforming the current model into a dynamic Bayesian filter. Instead of $K$, the algorithm would need to estimate the time series $K^{0:T}$. As the number of data sample is inherently limited, regularization assumptions would be required to overcome this computational challenge.

## ACKNOWLEDGMENTS

## REFERENCES

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. Psychological Review, 105(4), 678-723.

Bessière, P., Laugier, C., and Siegwart, R., editors (2008). Probabilistic Reasoning and Decision Making in Sensory-Motor Systems, volume 46 of Springer Tracts in Advanced Robotics.

Boden C & Giaschi D. (2007). M-stream and reading-related visual processes in developmental dyslexia. Psychological Bulletin, 133, 346-366.

Bosse, M.-L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: a cross-sectional study. Journal of Research in Reading, 32(2), 230-253.

Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. Cognition, 104(2), 198-230.

Bundesen C. (1990) A theory of visual attention. Psychological Review, 97:523–47.

Dubois, M., Kyllingsbæk, S., Prado, C., Musca, S. C., Peiffer, E., Lassus-Sangosse, D., Valdois, S. (in press) Fractionating the multi-character processing deficit in developmental dyslexia: Evidence from two case studies. Cortex.

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). Biological Psychiatry, 57(11), 1301-1309.

Lebeltel, O., Bessière, P., Diard, J., and Mazer, E. (2004). Bayesian robot programming. Autonomous Robots, 16(1):49–79.

Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: review of evidence for a selective visual attentional disorder. Dyslexia, 10(4), 339-363.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? Journal of Child Psychology and Psychiatry, 45(1), 2-40.

Vidyasagar, T.R. & Pammer, K  (in press). Dyslexia: a deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Science*.

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., and Cohen, D. (2006). Principles underlying the design of "the number race", an adaptive computer game for remediation of dyscalculia. Behavioral and Brain Functions, 2(19).