Very prolonged practice in block of trials: Scaling of fitness, universality and persistence

A.S. Itami *, F.J. Antonio, R.S. Mendes

Departamento de Física, Universidade Estadual de Maringá, Avenida Colombo 5790, 87020-900, Maringá – PR, Brazil
National Institute of Science and Technology for Complex Systems, Avenida Colombo 5790, 87020-900, Maringá – PR, Brazil

**Highlights**

- Statistical analysis of the reaction times in a psychomotor experiment.
- Very prolonged blocks of trials without rest breaks.
- Performance fitness, persistent behavior and universal patterns.

**Abstract**

In this study, we analyze the reaction times obtained from participants in a psychomotor activity composed of a large number of trials without breaks. We numerically evaluate the learning factor directly obtained as the interevent time in the subsequent trials comparing two different blocks of trials. We investigate the learning in terms of average values and their respective variability. In a broader scenario, we show that the learning can be associated with a scale factor acting over the reaction times. Aside from the fitness improvement, we identify that the reaction times have a positively skewed distribution, while their differences are distributed symmetrically as a Laplace distribution whose width diminishes with practice. We found that the differences of the reaction times after practicing become smaller obeying a linear rule. In addition to these universal patterns, we verify that the performance fitness does not exhibit persistence, but their differences do exhibit persistent behavior on the absolute values and anti-persistent behavior for the signs.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

It is possible to identify the positive effect of practice in everyday situations. Noticeable examples are the process of acquiring new abilities like learning how to read and to write, how to walk, to ride a bicycle, and to play any kind of sport or even a game. As it is well known, the time interval needed to complete a task diminishes with practice at a lower and lower rate while the practice goes on Refs. [1–3]. Traditionally, research in the field of skills acquisition looks to quantify the effect of practice mainly in terms of a reduction in the time interval spent to complete a given task [4–7]. Scientists have done many experiments throughout the years in order to investigate the learning process; the focus has been tasks on blocks with short time of execution divided into short blocks, punctuated with rest breaks. A non-extensive but representative list of examples is pressing buttons, ski-simulator [8,9], tracing geometric figures in a mirror [10], 1023 choice reaction time task [11], alphabet arithmetic task in the study of learning curve [12], memory search task in study of age-related learning.
Fig. 1. Setup of the psychomotor activity (left). In the right there is a picture illustrating the dimensions of the beads and their five colors used during the activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

differences [13], easy 4 choice reaction time task in the investigation of very prolonged practice [14], movement of a finger or pointer from one place to another in the study of Fitts’ Law [15], human stick balancing [16], speed stack bimanually with pyramid of cups in the study of motor learning in children with unilateral cerebral palsy [17], and reading text in the study of text reading fluency in adult readers [18].

Scientist have often performed trials in blocks with the objective of removing some presumed transient randomlike changes from trial to trial while emphasizing the persistent changes or the global trend of learning over the trials [19]. In contrast, a widely accepted perspective on psychomotor ability is that it is not a single general factor, but rather made up of an independent set of subfactors [20]. In this framework, the procedure of blocking data as groups of trials can, eventually, modify or mask intrinsic characteristics such as persistent trends as well as those of the transient changes [19,21].

The aim of this study was to investigate the performance trend on a very prolonged practice composed of blocks of trials, without breaks, in order to evaluate at what level the learning factor is related to the data. Asking the question: is it just a scaling factor or does it act more deeply, disturbing the shape of the probability distribution and altering the presence of correlations? We have used very prolonged blocks (1001 trials) and obtained the interevent times, aka Reaction Times ($T$) from trial to trial, as an alternative approach to the vast literature on shorter blocks of trials [4,5,7,11–13]. We have used commonly employed methods in statistical physics of complex systems to analyze the data. More precisely, we have focused on the distribution and the persistent trend (or memory effect) in the time series of $T$ and their first differences in time series. The persistent behavior in the time series were characterized by Detrended Fluctuation Analysis (DFA), proposed by Peng et al. [22] in their study about DNA sequences time series. The DFA method has been applied in the study of many contexts like synchronization and coordination processes in human movement [23], activity in a social network [24], music [25], chess game [26,27], and stock market data [28–30].

2. The data

The data are from a psychomotor activity performed by participants using a setup as depicted in Fig. 1. A total of 14 healthy adults (8 males and 6 females) with ages varying from 21 up to 38 years (average 25.9 ± 1.4 years) took part in the activity. All the participants are right-handed, have normal or corrected vision, and had no preliminary training. They sit down in front of the setup holding a container with 1001 acrylic flat beads in their lap. The beads are symmetric in shape, as depicted in the setup (see Fig. 1). There was a color sequence for the containers, as shown in Fig. 1; then, the participants had to separate the beads respecting that sequence. Participants should deposit the beads one by one in the container of the corresponding color. The participants had to be fast, but also as accurate as possible. Participants completed one block per day in two different days. Each block consisted of 1001 trials and the only feedback provided for the participants was the time spent on the whole block. It is valid to point out that the exercise employed non-repeated beads to avoid any induced bias due to the proportion of colors. Besides, since when the number of containers is increased, the movement of the participants’ hand would be longer to reach some containers than others, only 5 colors were used.

The data collected are the noise of each bead being deposited in the proper container. It was captured by a standard microphone working at a sampling frequency of 44.1 kHz and later converted to numerical data, from where the time
Fig. 2. (a) Moving average of the $T$ of one specific participant in the first (upper curve; blue) and second (lower curve; red) turns using sliding windows of size 100. (b) The average of the $T$ ($\mu$) of each person in the first (blue circles) and second (red squares) turns, indicating a reduction of the mean time. (c) The ratio of the standard deviation and the average of $T$ of each person in the first (blue circles) and second (red squares) turns, showing that this ratio is constant for each participant when the error bars are taken into consideration. In (b) and (c), the participants were ordered following the crescent difference $\mu_1 - \mu_2$, and the error bars correspond to 99% bootstrapped confidence interval.

3. Data analysis and results

Initially, we analyzed the time $T$ spent between putting two consecutive beads in the proper recipients. Since the activity involved training factors, it is instructive always to compare the time series of each participant from his/her first turn with his/her second turn, representing the $T$ of such times series as $T_1$ and $T_2$, respectively.

3.1. Scaling of fitness

Fig. 2(a) illustrates the reaction times of a specific participant. Since the reaction time data have large fluctuations, we applied a moving average technique [31] to soften the curves and to prove the difference between them. Notice that the upper curve ($T_1$; first turn) is always over the lower curve ($T_2$; second turn). In other words, the mean reaction times become smaller with practice. Naturally, peaks and valleys occur frequently in a very prolonged sequence of trials and they can be associated to transient randomlike changes and/or factors like fatigue [32], to momentary loss of attention [33] and to the process of choice of each bead [34]. In the context of time series, the presence of such peaks and valleys can be viewed as non-stationarities.

As pointed out previously, when compared to the first turn of the activity, the participants followed the tendency of spending less time in the second one, as displayed in Fig. 2(a). Quantitatively, Fig. 2(b) shows $\mu = \langle T \rangle$, the arithmetic mean time in seconds spent for each bead and the error bars correspond to bootstrapped 99% confidence interval, the interval for the variable corresponding to the given confidence level on the measurement [35]. The bootstrap is a resampling technique recognized for providing trustable confidence intervals for variables even when the fluctuations have non-Gaussian profiles [36]. The blue curve (circles) [red curve (squares)] is a measurement of $\mu$ computed from $T_1$ [$T_2$] for all the participants. 86% of the participants spent less time in the second turn. 67% of them had a relative reduction of more than 10%. It is evident from the error bars that at least 79% of the participants improved their ability despite any biased error in estimates.

Another quantity related to motor learning is the standard deviation $\sigma$ of the time series of $T$. This quantity provides a measure of how the time series is spread out, which is very suitable to analyze learning curves as an ability/accuracy measure. By construction, a small standard deviation indicates that the data points tend to be very close to the mean value and hence to each other, while a large standard deviation reveals that the data points are very spread out from the mean and from each other. Although there was a systematic decreasing trend for 86% of the participants, the error bars did not allow conclusive results for most participants.

A natural extension to the previous results is to investigate at which proportion the training influences not only the quantities $\mu$ and $\sigma$, but also their proportion (i.e., the ratio $\sigma/\mu$), the variability. In fact, we have found that the variability follows a universal pattern in the form:

$$\frac{\sigma}{\mu} = \text{constant}. \quad (1)$$

It is evident that the variability is dimensionless and, on average, the same constant held for all the participants; regardless of the turn of the activity. Fig. 2(c) illustrates a comparison between the first and second turns for each participant as well as their respective average values and 99% bootstrapped confidence intervals. It is evident that the averages collapse in almost all the confidence intervals. In particular, a T-test indicated that the null hypothesis of the curves being equal could not be rejected at a level of confidence of 99% ($p$-value 0.77). This result is consistent with the literature for some activities like human nonverbal counting [37] and numerical cognition experiments on indigenous people [38]. Beyond pointing in the direction of a universal mechanism, this scenery reveals that the learning factor can also be viewed as a scaling factor in the time series of the reaction times. A concern at this point is to distinguish if the rescaling learning process deeply disturbs
The time series of $T$ are always composed of positive values whose distributions are rather asymmetric (positively skewed) with an evident peak around the mode, i.e., the most observed value of $T$. The distributions of $T_1$ and $T_2$ of a specific participant are depicted in Fig. 3(a). Similar curves can be found in the literature and are, typically, fitted with a convolution of an exponential distribution to a Gaussian one [39,40]. This approach is intuitively plausible since it represents a mixture of random fluctuation and a learning curve. Nonetheless, this family of curves is not universal and cannot adjust to all of our data. It is valid to notice that all the empirical curves seem to belong to the same distribution, but with different parameters, even when comparing one participant’s two turns (see Fig. 3(a)). A statistical test like the Cramér–von Mises test supports this assumption at a 99% confidence level for 98% of the combinations. The Cramér–von Mises test is a convenient tool for judging the goodness of the fit of one distribution function to another one [41]. Typically, it is used to compare the comparability of an empirical distribution to a theoretical probability distribution. After rescaling the data by multiples of its mean, more than 80% of the inter-participant curves seemed not to belong to the same distribution. In contrast, when comparing the two distributions of the same participant, at least 50% of the curves could not be rejected of obeying the same distribution (see Fig. 3(a) and (b)). In this direction, a straightforward conclusion is that each participant has its own distribution, whose parameter is more sensitive to the learning and seems to be related to the mean reaction time.
Unlike for \( T \), it was possible to identify a symmetric distribution for the first differences of \( T \), defined as
\[
\Delta T(t) = T(t) - T(t - 1).
\]
As an example, the distribution of the first differences of one participant is depicted in Fig. 3(c). In order to get a better approach for a stable region, the time intervals corresponding to the first 100 beads were dropped, as well as the time intervals corresponding to the last 100 beads. It is evident that the distribution of \( \Delta T \) is wider than the distribution of \( \Delta T_2 \), which can be viewed as another manifestation of the learning fitness. In addition, both satisfied a Laplace distribution \( \mathcal{L}(\alpha, \beta) = \exp(-|x - \alpha|/\beta)/2\beta \), centered in \( \alpha \) and with scale parameter \( \beta \). The best fit parameters were obtained by the maximum-likelihood method and the Cramér–von Mises test indicates that the Laplace hypothesis cannot be rejected at a confidence level of 99% (the \( p \)-values laid in the range [0.11, 0.98]).

The two parameters of each Laplace distribution were different (see Fig. 3(d) and (e)). Particularly, the scale parameters \( \beta \) are connected to the learning patterns identified in Fig. 2(b). In other words, all the distributions tended to be wider in the first turn of each participant. It is worthy to note that the peak of the distributions is symmetrically centered around zero (see the scatter plot of \( \alpha_2 \) as a function of \( \alpha_1 \) in Fig. 3(d)). On the other hand, a scatter plot of the width of the distributions for each participant (see Fig. 3(e)) reveals a growth law for the scale parameter:
\[
\beta_2 = a + b \beta_1,
\]
where the intercept \( a = 0.09 \pm 0.04 \) is given in seconds and the slope \( b = 0.70 \pm 0.11 \) is dimensionless. This result suggests that besides always symmetrical, the fluctuations of each \( \Delta T \) after a very prolonged practice reduce linearly and depending on its initial value.

Laplace distributions can be reduced to a distribution independent of any free parameter by normalizing the data as
\[
\tilde{\Delta T} = \frac{\Delta T - \bar{\mu}}{\sigma},
\]
where \( \bar{\mu} \) is the average of \( \Delta T \) and \( \sigma \) its standard deviation. Using this standardized construction, all the time series \( \Delta T \) became zero-mean and exhibited unitary standard deviation. Their probability density function of all participants are depicted in Fig. 3(f) in comparison with a Laplace distribution with the form \( \mathcal{L}(0, 1/\sqrt{2}) \). It was noticeable that all the curves collapsed, unveiling the presence of a universal behavior that can be found, for instance, as the distribution of reading times [18] and in the same–different judgment [39].

### 3.3. Memory effect

Regarding probability distributions, we already pointed out that the learning effect behaved as a scaling factor over the time series. Now, we tackle another aspect concerning how the training deeply altered the time series structure. To do so, we investigated the time series of each participant searching the presence of persistent behavior or not, quantifying the findings via a fractal analysis. For analyzing the memory effect generated by the temporal dynamics of a time series \( u(t) \), it is very common to use DFA, which has been employed extensively in the literature to investigate persistence in time series via the Hurst exponent \( (h) \) even on non-stationary times series [42,43].

Basically, DFA-1 consists of four steps: (i) reduce the fluctuations accumulating the time series: \( y(t) = \sum_{j=1}^{t} u(j); \) (ii) divide \( y(t) \) in windows; (iii) for each window, remove non-stationary local trends in the form of a polynomial function of degree \( F \); (iv) compute the standard deviation over all the windows of size \( s \), denoting their average values by \( F(s) \). Typically, time series exhibiting correlations are characterized by a power-law behavior like
\[
F(s) \propto s^{h},
\]
where \( h \) is a scaling exponent which usually provides the value of the Hurst exponent. The numerical value of \( h \) quantifies the presence of autocorrelations of the time series. It is long-range correlated (anti-correlated) when \( h > 0.5 \) \((0 < h < 0.5)\). Otherwise, if \( h = 0.5 \), the time series is has no memory or has only short-range correlation.

The output obtained through the DFA-2 method over the time series of the reaction times for a participant is given in Fig. 4(a), where there is a comparison of the first and second turns of the experiment. In Fig. 4(b) the values of \( h \) for all the participants are depicted, also comparing the two turns of the experiment. All the results were around \( h = 0.5 \), indicating that there were no temporal correlations. In addition, there was no meaningful difference on the Hurst exponent when comparing the first turn with the second turn of each participant. It is known in the literature that the scaling range chosen for fitting the power law can disturb the numerical results [44]. We have used the range [6, 192] in log scale for all the fits, and a brute-force algorithm [45] corroborates our approximations for \( h \) and the general profile of the fluctuations.

Applying the DFA method to the time series of \( \Delta T \), we found the critical result \( h \approx 0 \). This outcome is consistent with many examples in the literature, where the DFA method has been used instead to investigate correlations in magnitude and sign of first difference time series [46–50]. In this scenario, we wrote the first difference of the reaction times of each participant as
\[
\Delta T = \text{abs}(\Delta T) \; \text{sign}(\Delta T),
\]
generating two components: (i) the time series of absolute values: \( \text{abs}(\Delta T) \); and (ii) the time series of signs: \( \text{sign}(\Delta T) \), composed of the elements \(-1 \) for \( \Delta T < 0 \), \( 0 \) for \( \Delta T = 0 \) or \( 1 \) when \( \Delta T > 0 \). The output obtained through the DFA-2
method over the time series \(\text{abs}(\Delta T)\) and \(\text{sign}(\Delta T)\) are available in Fig. 4(c). Again, there was no remarkable difference between the first and the second turn of each participant. It is noteworthy that by using computer-generated uncorrelated random data, the approach of splitting the time series into two components still leads to \(h \sim 0.5\) for both the components. In contrast, data obtained from nature have the ubiquitous tendency of persistent absolute values and anti-persistent signs. A few examples are the study of heartbeats [46,49] and investigations of earthquakes [50]. In a broader scenario, this indicates that our data are closer to other natural situations than to completely random numbers. Nonetheless, all the participants exhibited the same persistent or anti-persistent behavior. In a general context, these findings concerning the Hurst exponent indicate another universal characteristic of individuals that is not affected by training.

4. Discussion and conclusions

In this work, we analyzed statistically the response of a psychomotor learning activity characterized by very prolonged blocks of trials without any breaks. In general, psychomotor learning is connected with physical skills, relating cognitive functions to physical movement directed at achieving goals [51]. In the activity that we analyzed, the cognitive functions were associated to the decision-making process of following the color sequence and then searching for the respective bead in the ensemble of beads inside the initial container. On the other hand, the physical movement comprehended to select the right bead (due to the sequence of colors) and to deposit it in the proper container. During the analysis, we identified quantitative measures of the fitness profile by using the reaction times between trials in the learning practice activity. In that case, the psychomotor learning should be connected with the psychomotor efficiency measured through the profile of the reaction times.

In summary, we showed in Fig. 2(a) and (b) that the average time spent with each bead was significantly smaller in the second turn of the activity for 79% of the participants, which can be associated to an improvement in the efficiency of the participants for the proposed activity. This result is consistent with the analysis of the other activities, such as alphabet arithmetic task [12] and bimanual coordination [1–3,23]. Naturally, these results lead to an overall increase in the effective velocity interpreted as beads per second. With Fig. 2(c) we illustrated that practice not only reduces the time interval to complete a task but also reduces the fluctuations proportionally.

We identified that the reaction times follow positively skewed distributions whose parameters vary accordingly to intrinsic characteristics of participants and their level of training. We also found that the first differences in time series have
symmetric shapes that follow Laplace distributions. For this reason, the time series of the first differences of the reaction times ought to have an underlying universal mechanism. As expected, without using the normalization procedure proposed in Eq. (4), the Laplace distribution characteristic to the first differences of the reaction times had different scale parameters besides being centered in the neighborhood of zero. The learning effect identified in Fig. 2 governs the tail of the distributions. In other words, the learning feature tends to reduce the width of the distribution. We have found that the width of the Laplace distributions is intriguingly connected to very prolonged practice.

In the second part of the results, we have analyzed the profile of correlations in the time series of the reaction times (T) and in their first differences (ΔT). We have obtained that the time series of T does not exhibit memory effect. In contrast, the time series of ΔT presented two properties: (i) their absolute values were persistent; and (ii) their signs were anti-persistent. Similar results for the absolute and sign time series have also been found, for instance, in the increments between the time intervals of successive heartbeats [46], and in a connection between the pattern of earthquakes with the acoustic emission from crumpled plastic sheets [50]. In our results, it is also noticeable that in all the cases there was no relevant difference between the first and the second turn of the activity for each participant. A distinctive factor distinguishing our empirical observations from random uncorrelated noise is that the latter does not exhibit any kind of persistence, even in the magnitudes and signs of their first difference. It is worth noting that the learning effect due to practice did not introduce any kind of persistence present in the time series of reaction times and that it did not alter the family of the probability distributions.

It is important to notice that our main objective when looking to many sequential trials without breaks, was to investigate the properties of learning and of the persistent trend from trial to trial, which can be modified or masked when the data are blocked into smaller groups of trials. The activity that we considered deals with very prolonged local behaviors and not with a global behavior in the same way as the typical tapping activities [52,53]. As a consequence, the traditional learning curve, which has been extensively explored in the literature [54,55], could not be identified. However, the psychomotor learning due to the activity became evident when considering the decrease in the average reaction time of different blocks and the respective fluctuations around these average values. Another remarkable piece of evidence was found in the width of the Laplace distributions, whose value was found to be smaller after practicing. In contrast with standard deviation and mean value, the patterns of the variability, the family of the probability distributions and the persistence measured with the Hurst exponent were universal, that is, they did not depend on the training. Thus, the practice can be associated with a scale factor in the time series.

Acknowledgments

The authors are thankful for the financial support of the Brazilian agencies CNPq, CAPES, and the National Institute of Science and Technology for Complex Systems. F.J.A. is especially grateful to Capes/Fundação Aracuária for the grant number 87/2014.

References
