Journal of Mathematical Psychology 84 (2018) 74–88

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

The geometry of learning

Gianluca Calcagni

Instituto de Estructura de la Materia, CSIC, Serrano 121, 28006 Madrid, Spain

HIGHLIGHTS

- A correspondence between Pavlovian conditioning processes and fractals is proposed.
- This duality is applied to many associative theories and conditioning programs.
- $1/f$ scaling in human cognition and a random fractal model are compared.
- Slow learning can be interpreted as an excitatory process contaminated by inhibition.
- Individual response is characterized by progressively damped fluctuations.

ARTICLE INFO

Article history:
Received 4 January 2018
Received in revised form 23 March 2018

Keywords:
Pavlovian conditioning
Associative models
Fractal geometry

ABSTRACT

We establish a correspondence between Pavlovian conditioning processes and fractals. The association strength at a training trial corresponds to a point in a disconnected set at a given iteration level. In this way, one can represent a training process as a hopping on a fractal set, instead of the traditional learning curve as a function of the trial. The main advantage of this novel perspective is to provide an elegant classification of associative theories in terms of the geometric features of fractal sets. In particular, the dimension of fractals can measure the efficiency of conditioning models. We illustrate the correspondence with the examples of the Hull, Rescorla–Wagner, and Mackintosh models and show that they are equivalent to a Cantor set. More generally, conditioning programs are described by the geometry of their associated fractal, which gives much more information than just its dimension. We show this in several examples of random fractals and also comment on a possible relation between our formalism and other "fractal" findings in the cognitive literature.

1. Introduction

Making psychology quantitative has been a difficult but feasible challenge since the first laboratory of psychophysiology established by Wundt. Making it a mathematical theory under analytic control has been, and possibly will always be, a utopia. Nevertheless, there are a plethora of analytic models which are able to fit and explain data coming from the observation of subjects in specific experiments. For instance, in the context of behavioral theories of Pavlovian conditioning, one can study the interplay between a conditioned stimulus (CS) and the subsequent occurrence of an unconditioned stimulus (US) of typically high relevance for the subject, such as food or an electric discharge. Despite their limited range of applicability, associative conditioning models are useful for several reasons. First, they express in a compact and economic way concepts that took time and many pages to be formulated. For example, the fact that "the prior activity influences the value of" the stimulus, recognized since the early stages of functionalism (Dewey, 1896), translates effectively in a description of conditioning as an iterative process progressively modifying the strength of the association and the salience of the stimulus. Second, they offer novel insights that can be easily checked and falsified quantitatively as new data and experimental designs become available. The question we would like to pose in this paper, limited to animal and human behavior, is: How much can we expand our toolbox of mathematical models in order to extract valuable information on learning processes?

The classic 1950s theoretical approaches to simple cases of conditioning are cast in the language of probability theory [see, e.g., the works by Bush and Mosteller (1951a, b, 1953), Estes (1950), Estes and Burke, 1953, and the reviews by Bower (1994) and Mosteller (1958)]. In these models, one considers the probability p of a given conditioned response (CR) as a function of the trial number n. The increment Δp_n at each trial is linear in p_n; by evaluating p_n iteratively, one obtains a learning curve. Alternatively, the probability p can be replaced by the strength of association V. This change of variable is useful for phenomenological applications because V, although mediated by internal variables such as the organism's motivational state or attention, can directly be measured

E-mail address: g.calcagni@csic.es.

https://doi.org/10.1016/j.jmp.2018.03.007

by several performance indicators, *in primis* the subject response. For instance, the quality of surprise in the US as a function of the appearance of the CS was first suggested by Kamin in relation with cue competition (Kamin, 1968, 1969). For a single CS, the evolution of the novelty (or "surprisingness") of the US along the learning curve had been made quantitative already by Hull in his linear model of Pavlovian conditioning (Hull, 1943). Recast in modern terminology by Rescorla and Wagner (1972) and Wagner and Rescorla (1972), this model states that the change ΔV_n in the strength of the association at the nth trial is

$$\Delta V_n = \alpha\beta(\lambda - V_{n-1}), \qquad n = 1, 2, 3, \ldots, \tag{1}$$

where $0 \leqslant \alpha \leqslant 1$ is the salience of the CS, $0 \leqslant \beta \leqslant 1$ is the salience of the US, and $0 \leqslant \lambda \leqslant 1$ is the magnitude or intensity of the US (i.e., the asymptote of learning). The term $\lambda - V_{n-1}$ indicates the surprisingness of the US, which decreases as the associative strength increases. The association strength gained up to the start of the nth trial can be found iteratively:

$$V_n = V_{n-1} + \Delta V_n = (1 - \alpha\beta)V_{n-1} + \alpha\beta\lambda\,. \tag{2}$$

The solution of this equation is

$$V_n = \lambda[1 - (1 - \alpha\beta)^{n-1}]\,. \tag{3}$$

When no association has been made yet, at the beginning of the first trial $V_1 = 0$, which fixes the unphysical constant $V_0 = -\alpha\beta\lambda/(1 - \alpha\beta)$. When $\lambda \neq 0$, the conditioning is excitatory and the US always occurs after the CS. Maximum learning is achieved when $V = \lambda$. If $\lambda = 0$, the US does not show up after the CS and the conditioning is inhibitory or of extinction. Rescorla and Wagner extended the linear model to the case of the presentation of multiple CSs (Rescorla & Wagner, 1972; Wagner & Rescorla, 1972), as we will discuss later.

The main contribution of this paper is to propose a geometric interpretation of learning processes which carries several advantages. First, it is useful at the time of assessing the efficiency of these processes quantitatively, both within a given model (how efficiency is affected by the salience of the stimuli for the subject) and when comparing different models. The *efficiency* of an excitatory conditioning can be roughly defined as the inverse of the number of trials necessary to increase the associative strength from 0 to, say, 0.9 λ. This concept is subject-dependent and may be used either to compare the learning of different individuals within the same program or, when averaging over individuals within the same experimental group, to compare different learning programs.

Specifically, we obtain the following results. (i) We recognize Eq. (2) as one of the similarity maps defining Cantor sets, which are an example of peculiar, totally disconnected sets known as deterministic fractals. (ii) We calculate the Hausdorff dimension $d_{\rm H}$ of the set for Hull's model and show that it depends on the parameters α and β in such a way that *the smaller the dimension, the more efficient the conditioning*. (iii) This picture can be generalized to any other conditioning described by iterative equations, giving explicit multidimensional examples that include Rescorla–Wagner, Mackintosh, and Pearce–Hall models. As a further application to nonlinear sets, (iv) we approximate Mackintosh theory (in the case of a single cue) with a new model where the recursive equation describes slow learning at intermediate trials; the dimension of this conditioning process is calculated and shown to be greater than in the Hull model for the same asymptotic value of the parameters, in agreement with (ii). Note that, in the presence of a single cue, the learning rate is already enough to compare different individuals or programs. One can see this by noting that the Hausdorff dimension (8) only depends on the product of the saliences and the smaller the salience, the larger the dimension. Nevertheless, when one goes beyond single-cue configurations and considers more complicated settings (Section 4), it may become progressively difficult or ambiguous to define effective learning rates. On the contrary, the Hausdorff dimension is always a well-defined parameter that provides a quick way to compare different individuals or models. Unfortunately, in practice, calculating the Hausdorff dimension for complicated deterministic processes may be as difficult as deciding on effective learning rates. However, the fractal paradigm is not limited to the definition of a new parameter, and its advantages do not end here. (v) The rethinking of learning processes in geometric terms will allow us to reinterpret conditioning as a *mixture of excitatory and inhibitory processes* rather than a black-or-white selection of either. The degree of mixing will be determined by the value of the Hausdorff dimension (Section 6).

(vi) Also, we generalize the construction to random fractals, which are essential to describe experimental designs of Pavlovian conditioning where the characteristic of the stimuli are determined by random algorithms or the US is not presented at all trials (partial reinforcement). The Hausdorff dimension of the Cantor set is independent of the US intensity and it does not fully capture the efficiency of a process. This is obvious from Eq. (8) but (vii) we also give the counterexample of a partial-reinforcement program (know to be "less conditioning" than continuous reinforcement), where $\lambda = 0$ in some of the trial but α and β (hence, $d_{\rm H}$) are kept fixed throughout. Here the efficiency (the Hausdorff dimension, a pointwise geometric indicator) is less important than the determination of the geometric shape of the fractal, which offers a more global and useful perspective than the number $d_{\rm H}$. In fact, (viii) the mappings generating the fractal give a prediction on the learning curve: there will be plateaux in the curve with such and such distribution determined by the random algorithm employed to pick the value of the parameters at each trial. Different randomizations of Hull's model will illustrate the point. Finally, (ix) we make a preliminary connection with some results in the cognitive literature on performance variability, which was found to follow a multifractal pattern. With all due caution in comparing widely different paradigms, we simulate performance variability of internal origin by a Pavlovian conditioning model where the salience of the stimuli slightly changes at each trial, according to a random algorithm. Since $d_{\rm H} = d_{\rm H}(\alpha\beta)$ only depends on "internal" parameters determined by the type of subject and the type of stimuli presented, under a cognitive-interactionist perspective the Hausdorff dimension can be reinterpreted as the part of the efficiency of the process due to the characteristics of the subject in relation to the stimuli presented. Again, fractal geometry has the potential to open a new door of analysis.

The plan of the article is as follows. In Section 2, we recall some basic aspects of deterministic fractals. In Section 3, we apply this formalism to Hull's associative model of Pavlovian conditioning. Section 4 is devoted to the generalization of this one-dimensional case to more realistic models with many cues or deterministically varying parameters (CS salience, US magnitude), such as Rescorla–Wagner (Section 4.1), Mackintosh and Pearce–Hall (Section 4.2), and a new nonlinear model akin to Mackintosh (Section 4.3). Random fractals are the subject of Section 5; flexible conditioning programs are discussed in Section 5.1, where the fractal construction is extended to the very important case of random sets; a digression on cognitive experiments unveiling a multifractal pattern in task performance variability and its possible relation with our findings is discussed in Section 5.2. Section 6 briefly explores some applications of the fractal picture, both to the practical understanding of conditioning processes and to experimental predictions about response variability. Conclusions and future directions are in Section 7.