



HAL
open science

Does the enactive approach provide the framework for a new definition of learning?

Olivier Gapenne, Mehdi Khamassi

► To cite this version:

Olivier Gapenne, Mehdi Khamassi. Does the enactive approach provide the framework for a new definition of learning?. *Intellectica - La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo)*, 2024, Perspective(s) enactive(s) en sciences de l'éducation et de la formation, 80 (1), pp.21-58. hal-04670326

HAL Id: hal-04670326

<https://hal.science/hal-04670326v1>

Submitted on 12 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Does the enactive approach provide the framework for a new definition of learning?

L'approche éactive constitue-t-elle le cadre d'une nouvelle définition de l'apprentissage ?

Auteur/Author :

Olivier Gapenne – Université de technologie de Compiègne – Unité Costech

olivier.gapenne@utc.fr

Mehdi Khamassi – Institut des Systèmes Intelligents et de Robotique, CNRS, Sorbonne Université – mehdi.khamassi@sorbonne-universite.fr

Abstract:

What does learning mean for a living, thinking being? What are the mechanisms involved in mastering a new ability or transforming an existing one? For just over a century, the sciences that study intelligent behavior have proposed precise definitions and models of learning, favoring a behaviorist and internalist approach to learning, initially seen as an associative process, then as a process of acquisition, and finally as a process of self-organization. Is the enactive approach, and more generally the 4E approaches to cognition (embodied, embedded, enactive, extended), likely to change the definition of learning in the same way that these approaches are changing definitions in relation to cognition more generally? We first take a long look at the different ways in which learning has been understood over the last hundred or so years. We examine to what extent these successive new interpretations were useful, and we discuss their limitations. Then, in a second, shorter section, we attempt to lay the groundwork for an enactive definition of learning, which is conspicuously absent from the work of the pioneers.

Key Words: learning, enaction, association, connection, feedback, modeling

Résumé :

Que signifie apprendre pour un être vivant et pensant ? Quels sont les mécanismes favorables à la maîtrise d'une nouvelle capacité ou à la transformation d'une capacité existante ? Depuis un peu plus d'un siècle, les sciences qui étudient les comportements intelligents ont proposé des définitions et modélisations précises de l'apprentissage, privilégiant une approche comportementaliste et internaliste du mécanisme, initialement conçu comme un processus associatif, puis d'acquisition et enfin d'auto-organisation. L'approche éactive, et plus généralement les approches de la cognition 4E (*embodied, embedded, enactive, extended* ; en français, incorporée, située, éactive, étendue), sont-elles susceptibles de faire évoluer la définition de l'apprentissage comme elles le font plus globalement pour la cognition ? Nous proposons dans un premier temps de revenir longuement sur les principales conceptions de l'apprentissage pour en identifier les conquêtes successives et les limites. Au-delà de ces bilans, et dans un second temps plus court, nous tentons de formuler les premiers éléments d'une définition éactive de l'apprentissage dont il faut indiquer qu'elle fait explicitement défaut dans les travaux des pionniers.

I. Introduction

In this article we focus on the question of learning, showing how different theoretical approaches have gradually revealed the ways in which learning can take place. Here, we define learning as a mechanism, or set of mechanisms, by which a living organism increases its power through acquiring mastery of a new relationship with the world or with itself¹. A study of how behaviors emerge, and how they are transformed, will sometimes approach learning from a systems perspective, seeing learning as a network of *nodes* and *links*, and the history of this type of what we might call “learning science” is broadly the history of how ideas have changed as to how this network is constituted.

There is, however, a different way of approaching learning. In the place of a direct focus on learning as a mechanism, the focus may be indirect, examining instead the functions, effects, forms, and conditions of learning. An exhaustive review of the literature shows that, apart from a few original theoretical proposals and some computational modeling work, definitions of learning as a mechanism are rare and sometimes circular. Importantly, while it is indisputable that the notion of learning (or rather, of adaptation) permeates all of Varela and Maturana’s work on enaction and autopoiesis, a careful re-examination of their work has convinced us that this notion, as employed by Varela and Maturana, lacks a clear, enactive definition.

A change has occurred in the approach to learning from a systems perspective. In the scientific traditions of associationism, cognitivism and connectionism, the network of nodes and links referred to above was traditionally *internal*, *i.e.*, containing components limited to the agent and its brain, seen as a machine capable of producing representations from information about the environment. Recently, however, as a result of significant conceptual developments, learning systemics or dynamics have been considered as having a specific organization, subject to disturbances from elements coupled to and transcending the learning entity itself. Approaches based on the theory of non-linear dynamic systems, on the ecological approach, on cultural anthropology, on situated and distributed cognition, and on enaction all include physical and human environments, culture and technology in the determination of learning. The contribution that these approaches have made to the field of genetic constructivism², and in particular to approaches based on enaction, has enabled a clearer understanding of the mechanism of learning and its place in cognitive genesis. We consider learning to be a specific relational mode, with phylogenetic differences that point to difficulties (or even impossibilities) encountered by certain species in relation to learning. These difficulties reflect, on the one hand, limits in the extent to which organisms may be deformed or transformed, and, on the other, their relationship to the environment. The issue of the difficulty of learning becomes

¹ For the human species, the conquest of this power almost systematically mobilizes tools, instruments, devices and mediations. In other words, learning is technically augmented, which amplifies and constrains both the possibilities for relating to the world and to oneself, and the ways in which the flow of consciousness can be filled.

² On this point, we refer to Visetti (2004, pp. 231-233), who describes two forms of constructivism in particular: assembly constructivism and genetic constructivism. Visetti’s paper includes the following lines: "Here, every construction is understood as the growth, differentiation and complexification of a potential that is already organized from the outset, without, however, immanently determining the process that begins: it is not the property of a plan or a program, but first and foremost that of a dynamically organized support, at once closed and open, in a specific mode that is the creator of history. Local and global, structure and process, are mutually and dynamically determined. The individuality of the construction, like that of its parts, is never taken for granted. Rather, we describe reorganizations, or the unfolding of an organization sketched out from the outset, rather than the appearance of organizations from nothing. Organization always seems a primitive term: to be is to be organized, to pre-exist is to be pre-organized, etc."

particularly acute as regards perception and action, once we move beyond so-called sensitive periods (Knudsen, 2004). This suggests that where learning mobilizes a singular mode of coupling, it is probably not independent from more elementary modes of relationship specific to living organizations – such as autopoiesis – emerging earlier within the vital process.

II. Learning as passive bonding: associative learning

To our knowledge, the work of physiologist Ivan Pavlov (1927) was the first scientific investigation of learning. His work inspired the entire associationist and behaviorist traditions. Pavlov considered that events in the world (*e.g.*, signals sent to sensors) have their correspondents in the nervous system, and that learning consists in the creation of associative links between neurological entities. The first of the five explicit definitions of learning that we have identified in the literature may thus be stated as follows:

Learning definition #1: Learning consists in the creation of associative links between events within the nervous system that correspond to events in the world.

Pavlov further posited that the formation of these links occurs passively and is induced by the temporally contiguous repetition of at least two events. In concrete terms, Pavlov used a particular learning method, now known as Pavlovian or classical conditioning. This is a method, first used in dogs, whose aim is to take an already constituted relationship between what is termed an *unconditional stimulus* (US), *e.g.*, food, and an *unconditional response* (UR), salivation in the case of Pavlov's dogs, and to replace it with a relationship between a *conditioned stimulus* (CS), *e.g.*, sound or light, and a *conditioned response* (CR): the salivation in Pavlov's dogs becomes a CR after when it is a response triggered by the CS. Initially, the CS is a neutral stimulus (NS) not eliciting an UR, and subsequently the UR is elicited after repeated exposure to NS-US pairs. NS becomes CS, and UR becomes CR.

To explain the fact that UR and CR are identical, Pavlov hypothesized that an associative link is established between CS and US, which is reinforced through repetition. The strength of the link between CS and US is the measure of the strength of learning, and a great deal of research has been done into the factors determining this strength. As well as repetition and contiguity, two additional important factors have been identified, namely the salience of the CS and the power of the reinforcement. However, as we shall see later, other factors such as attention (Kamin, 1969; Mackintosh, 1975; Pearce & Hall, 1980; Panayi, Khamassi et al., 2021), can affect learning just as significantly in the same conditioning situation.

In the conditioning setup used by Pavlov, a metronome delivers a sound (NS) that the dog can hear, and a cannula inserted into the dog's cheek measures the amount of saliva produced (UR). The experimenter can also apply meat powder (US) to the dog's tongue. The procedure generally consists of the presentation of NS (which, at least initially, gives rise to an orientation response that has to be different from UR), followed by the presentation of US, which necessarily produces UR. Gradually (as Pavlov incidentally observed), the NS becomes CS. In other words, the delivery of CS produces UR, which we then term CR. Stabilization of learning occurs when the presentation of CS alone is able to produce CR. The sound of the metronome produces an increase in saliva production. This equivalence between UR and CR was a decisive element in the theorization of the situation. How could the fact that CS and US produce the same response (substitution hypothesis) be explained without the existence of a CS-US link?

All the experiments carried out with this protocol show that the effects are predictable and generalizable to all the species tested including humans (Balleine and O'Doherty, 2010). The shape of the learning curves is sufficiently invariable to have led scientists to imagine that they

were indeed holding the key to learning. However, the validity of these experiments requires the conditioning sequence NS - US - UR be rigorously observed. In addition (and this indicates that controlling this type of situation, despite its apparent simplicity, can be a very delicate matter) the CS is often composite: that is to say, while the sound of the metronome can induce salivation, events temporally associated with the sound can also do so. This can be demonstrated by combining two neutral stimuli – let us call them A and B – and presenting them together as a CS. After conditioning, either of the two triggers in isolation is shown to trigger CR, but to a lesser extent, as if the strength of the CS-US link were shared between A and B (Dickinson and Balleine, 2002). Finally, as numerous studies have shown, even on apparently elementary responses such as the palpebral response, there are systematic differences between individuals.

The inclusion in the initial model of three complementary phenomena has helped to delineate the mechanism assumed to be at play in the situation, *i.e.*, the formation of a link between CS and US. These three phenomena, outlined below, are *extinction*, *generalization*, and *discrimination*. Other phenomena that have enabled researchers to specify some of the mechanisms needed by computational models to account for Pavlovian conditioning are *blocking*, *conditioned inhibition*, and *conditioned suppression*.

Phenomenon 1: Extinction

Extinction is a gradual disappearance of CR when reinforcement (US) is no longer present. The interesting point about extinction is that it does not correspond to a disappearance of the binding, but rather to an inhibition of it. This has been confirmed by two elements, as follows. First, the first presentation of US, after a period of interruption, immediately gives rise to a response (re-learning is not necessary) with less latency than in the initial conditioning phase. Modern neuroscience has identified similar phenomena at the cellular level when inhibition is lifted. Second, when the period of reinforcement interruption is prolonged, a form of inhibition-forgetting can be observed: CS again elicits CR. Another noticeable feature is that some conditioning protocols where reinforcement is not delivered systematically create resistance to extinction. At this level, the question is whether it is the strength of the associative relationship that is greater, or the limited inhibition. What is more, while CR may disappear, it may occur for stimuli other than CS.

Phenomenon 2: Generalization

Composite stimuli have been mentioned above, but more interestingly as regards a possible modeling of the mechanism, it has been shown that CRs can be induced by CSs that were not initially present and that have a dimensional proximity (in terms, for example, of intensity or frequency) with the initial CS (Ghirlanda and Enquist, 2003). These induced CRs are generally attenuated in intensity, and the greater the distance from the initial stimulus, the greater the attenuation. Just as remarkably, it is possible to induce CRs with signals possessing the same characteristics as the initial CS but differing in the modality used (visual, auditory, or tactile). This is a very important point, and is likely to fuel new hypotheses in regard to the intrinsic intermodality of perceptual and, more broadly, cognitive activities.

Phenomenon 3: Discrimination

As well as having relevance to the study of generalization and categorization abilities, the classical learning setup is also relevant to the study of discrimination. It is possible to establish conditioning with two distinct CSs eliciting the same CR (and by implication, with the same US for reinforcement) and to show that producing extinction for one of the CSs (by removing the US) does not produce extinction for the other. This simple but remarkable finding indicates that the animal is able to discriminate between two CSs. From here, as in the case of

generalization, it becomes possible to implement psychophysical-type protocols to identify differential thresholds for a whole range of signal dimensions.

III. Computational models of associative learning

The learning model that Pavlov developed was directly inspired by the neurobiological work going on around him. Having observed the phenomenon of neuronal growth, Pavlov suggested a spawning model to explain the formation of associative links. At the time, this type of proposal was generally considered hypothetical, but contemporary neuroscientific research has largely proved him right (which, of course, is not to say that the question of learning has been settled). Certain functional rules have been established. These include Hebb's Law, which states that if two interconnected neurons are co-active, the connecting force of the synapse that links them must be potentiated (Hebb, 1949). The neurobiological mechanisms underlying Hebb's Law have in recent years been partially elucidated (Langille and Brown, 2018; Martens, Celikel and Tiesinga, 2015). Hebb's Law is a functional rule underlining the significance of the temporal contiguity of associated events, and as such lies at the heart of formal connectionist models and the mechanisms of neurogenesis that are to be found, for example, in the visual system. Hull's model (1943), although inspired by Pavlov's work and models, is a formal model mobilizing variables that cannot be directly verified experimentally. It is only deductions made from the model that may be verified empirically.

Hull advocated the use of intermediate variables, which he used extensively in his models. Intermediate variables allow for internal processes that are not directly observable, and which depend on associations. They are comparable to the latent variables that physicists have used to model unobservable phenomena. Models of this kind have been used to account for conditioning by including a latent variable as a hidden cause behind observations that may be CSs and USs (Courville et al., 2006; Gershman and Niv, 2012). From a computational point of view, this notion of latent variable has a wider scope than Pavlovian conditioning alone. It also resonates directly with the notion of latent learning espoused by Tolman, which we will explore further below: this is the idea that even in the absence of reward, an animal may learn from the relationships and properties of its environment. The learning may not at first be visible in the animal's behavior, hence the idea that it is latent or hidden, although we need to be able to account for the mechanisms underlying it. The relationships and properties learned by the animal (see our discussion below relating to *cognitive maps*) can be used by it to adapt its behavior more quickly in the presence of reward (see for example, Dollé et al., 2018).

Phenomenon 4: Blocking

A fourth phenomenon, blocking, sheds light on one of the key mechanisms in computational models of Pavlovian conditioning, namely *prediction error*. In blocking experiments the animal has already learned that a stimulus A (e.g., a sound) is followed by a food reward. A has therefore become a CS associated with a US. A second stimulus B (e.g., a light) is now presented simultaneously with A, followed by the US. This is repeated for a set of trials, followed by a test phase in which the animal is exposed to stimulus B only. It is found that B does not trigger salivation in the animal, suggesting that its conditioning has been blocked by the presence of an already reward-predicting stimulus A (Dickinson et al., 1976). In a variant of this experiment, A is only partially conditioned: the animal's training in the A-US association is interrupted after a limited number of trials. We then move on to phase 2, where A and B are presented simultaneously, followed by US. Here, it can be observed that B has indeed become a conditioned stimulus: its presentation alone triggers a salivation response from the animal, albeit a weaker one.

If a reinforcer is already fully predicted by the animal in virtue of a CS, then its appearance after the AB presentation is not surprising, and therefore does not trigger new conditioning: B is blocked. But when a certain level of surprise persists, conditioning can occur. This goes beyond Pavlov's original model, and below we will show precisely how it may be operationalized in computational models.

Phenomenon 5: Conditioned inhibition

Pavlov (1927) himself examined the limits of his initial model in certain situations, including conditioned inhibition. Conditioned inhibition is where an initial conditioning is established, and where a composite CS (the initial CS combined with a new CS) is then presented. This composite CS, which elicits a CR, will generally be subject to extinction if it is not followed by a US. However, if the initial CS is presented again, CR reappears. This suggests that the link between CS and CR initially established is not the object of inhibition as such, and that if extinction is due to inhibition, then this comes from the new CS.

Phenomenon 6: Conditioned suppression

Conditioned suppression is another example of conditioning that is difficult to interpret within the initial framework, showing that the CS-US link alone is not sufficient to account for learning. Here, the initial conditioning is of the Skinnerian or instrumental type (see below), *i.e.*, the animal learns to press a pedal to obtain positive reinforcement (food). Once the response rate has stabilized, the experimenter presents a new CS (*e.g.*, a sound) for a minute or more, followed by a negative reinforcement (*e.g.*, an electric shock). When this couple is repeated, the animal avoids pressing the pedal, *i.e.*, the CR disappears. The numbers of responses produced respectively with and without negative US over a given duration are compared, and a suppression ratio calculated which, if it tends towards 0, indicates that conditioning has been established. Its value reflects the strength of the association (the link here being inhibitory). Rescorla (1968), adopting this protocol, came up with the following idea: after establishing aversive CS-US conditioning (with an 80% probability of a shock being administered) for each of the animals in a sample, he divided the sample into 4 groups with different probabilities of shock during the CS-free phase. The most important result of this experiment is to show that the suppression ratio varies in the 4 groups. It is highest when the probability of shock is the same with or without CS, and it is lowest when the probability of shock is zero in the absence of CS. What is interesting is that in all 4 groups there are similar numbers of occurrences of the CS-US pair, a factor which cannot therefore explain the differences observed, as the initial Pavlovian model might suggest. This experiment shows that the contingency between CS and US would appear to be the critical factor in explaining these differences. The contingency is zero in the situation where the probability of US occurrence is the same with or without CS.

The Rescorla-Wagner model

On the basis of experiments like those described above in relation to blocking and to conditioned suppression, Rescorla and Wagner (1972) proposed a theory of conditioning where the critical factor determining the strength of the CS-US link is not so much repetition as the level of prediction of the US. Their model moreover excludes factors, proposed by Pavlov, relating to salience and power. Learning is thus the result of novelty and surprise. In this regard, Rescorla and Wagner (*ibid.*) operationalized a hypothesis proposed by Kamin (1969) that he termed "retrospective contemplation", with reference to the idea that the animals (such as rats) used in this type of experiment would somehow replay and review any situation in which they had been confronted with an unpredicted stimulus. This internal cognitive activity was seen as underpinning the active learning process. Interestingly, phenomena more recently studied in neuroscience, such as hippocampal reactivations – which appear to replay the reversed

sequence of actions just performed by the animal, from the US (where the animal is at the end of the trial) to the CS (where it was at the beginning) in mazes – are now being interpreted as possible substrates of this type of situational review that may contribute to learning (see, for example, Cazé et al., 2018). Possible computational mechanisms for this type of retrospective activity able to modulate learning have also been the subject of recent studies in humans (Moran et al., 2019).

Rescorla and Wagner, for their part, took a much simpler approach, focusing on a possible reinforcement mechanism based on a calculation of US prediction error given the stimuli presented. This mechanism is at the heart of all current models of reinforcement learning (Sutton and Barto, 2018; see the following section), including recent versions seeking to account for hippocampal reactivations that enable situations to be reviewed. The following equation summarizes the central learning mechanism in Rescorla and Wagner's model:

$$V_{t+1}(k) = V_t(k) + \alpha\beta_k \left(r_{t+1} - \sum_k V_t(k) \right)$$

The iterative process implemented by this equation can be described as follows: each stimulus k gradually acquires a predictive value for the reward, $V(k)$, which is increasingly reinforced over the course of successive trials (hence the representation of its changing value at time t , then at time $t+1$, etc.) through repeated association with the reward r . The reward itself is not enough to reinforce the value of stimulus k . What counts is the prediction error, *i.e.*, the difference between the amount of reward predicted (*i.e.*, expected) at time t and the amount r_{t+1} of reward actually obtained by the animal at time $t+1$. This predicted reward quantity is calculated here as the sum of the predicted reward values of the individual stimuli, as learned from previous trials. The parameters α and β respectively represent the rate (or speed) of learning (if α is close to 0, learning is slow, while if α is close to 1, learning is fast) and the salience of stimulus k (learning is faster with a highly salient stimulus i whose value β is close to 1 than with a less salient stimulus j whose value β is close to 0).

Below are a few examples³ to illustrate the central mechanism at play in this equation:

- Example 1: At the outset we associate a neutral stimulus A, whose initial value is therefore set to 0 (*i.e.*, $V_t(A) = 0$), with a reward arbitrarily set to 1 (*i.e.*, $r_{t+1}=1$). There is no other stimulus present, so $\sum_k V_t(k) = V_t(A) = 0$. The surprise associated with this first reward is at its greatest, and the model calculates a prediction error of $r_{t+1} - \sum_k V_t(k) = r_{t+1} - V_t(A) = 1 - 0 = 1$. Stimulus A is reinforced, its value at the next trial ($t+1$) being: $V_{t+1}(A) = V_t(A) + \alpha\beta_A(r_{t+1} - V_t(A)) = 0 + \alpha\beta_A(1 - 0) = \alpha\beta_A$. The value of stimulus A has therefore been increased by the quantity $\alpha\beta_A$.
- Example 2: We are now N trials later, *i.e.*, at time $t+N$, and stimulus A has been repeatedly associated with the reward for all these trials. It is now perfectly predictive of the reward (*i.e.*, $V_{t+N}(A) = r_{t+N} = 1$). What, according to this model, will happen on subsequent rewarded exposures to stimulus A? The model predicts that there is no longer any prediction error, because the reward is perfectly predicted, and so the stimulus value will not be further reinforced: $V_{t+N+1}(A) = V_{t+N}(A) + \alpha\beta_A(r_{t+N+1} - V_{t+N}(A)) = 1 + \alpha\beta_A(1 - 1) = 1 + 0 = 1$.
- Example 3: On the following trial, $t+N+2$, we present a new stimulus B, initially neutral, *i.e.*, $V_{t+N+2}(B) = 0$, simultaneously with stimulus A, and then give a reward. The Rescorla-Wagner model proposes the following explanation for the blocking

³ We refer readers wishing to go further to the book chapter (in French) *Introduction à la modélisation computationnelle* (Collins & Khamassi, 2021) and to the open-source code to which the chapter provides a link.

phenomenon: the value of stimulus B will not be reinforced because there is a zero prediction error, given that the reward is already perfectly predicted by stimulus A. Updating the value of B is expressed mathematically as follows: $V_{t+N+3}(B) = V_{t+N+2}(B) + \alpha\beta_B(r_t - \sum_k V_{t+N+2}(k)) = V_{t+N+2}(B) + \alpha\beta_B(r_t - (V_{t+N+2}(A) + V_{t+N+2}(B))) = 0 + \alpha\beta_B(1 - (1 + 0)) = 0 + 0 = 0$. The value of B remains zero.

Wagner, Rudy and Whitlow (1973) set up a more elaborate experiment to test other possible mechanisms with a view to extending their model. In their experiment, two clearly predictable situations are established following two CSs (A and B). An additional CS-US aversive conditioning is then performed, followed 10 seconds later by a post-trial event (PTE). The result of this manipulation of the predictability of the PTE demonstrated that despite an identical number of CS-US associations, conditioning is more difficult to establish when the PTEs are surprising. The presence of an unexpected PTE would prevent the process of post-trial repetition.

The story so far

Before going any further, let us consider what Pavlov teaches us about the learning mechanism. It is undeniable that the Pavlovian conditioning situation modifies an animal's behavior, or rather, has some effect on an animal's reactivity or apprehension domain, since the animal does not in fact produce a new response (UR and CR are similar), but this sheds only a very limited light on learning as a mechanism, since learning more generally involves producing new modes of exploration to apprehend new situations and objects. However, computational modelling of mechanisms possibly underlying the phenomena that Pavlov observed has since been able to identify types of neural activity, such as the response of dopaminergic neurons, correlated with reward prediction error signals (Schultz et al., 1997). Pavlov's legacy therefore provides a glimpse of plausible neurobiological mechanisms underlying this type of learning, which is admittedly very simple.

The learning paradigm of Pavlovian conditioning posits the possibility of creating arbitrary associations based on the temporally ordered contiguity of events. This time-constrained associative potential, which has largely inspired cognitive architectures in computational modeling (*cf.* Hebb's rule), leaves no room for movement, and addresses questions of perception and cognition purely in terms of passive reception and aggregation or association, a position that does not seem tenable when considering a living being, and which would appear to rule out a possible functional hierarchy of the multiple couplings (proprioceptive and others) with which animals are endowed.

So, what does the potential creation of new, arbitrary associations tell us? First of all, if we consider UR to be the result of an already constituted schema (*e.g.*, salivation), the possibility of producing this response again in relation to temporally contiguous events (still heralded by US, at least initially) means that some kind of structure of anticipation by recurrence, as proposed by Piaget (1967), has been created. But is not this simply the extension of an assimilating schema⁴, which would be incompatible with the idea that learning necessarily involves a transformation of the schema's overall structure? The difficulty here is knowing to what extent *the animal's capacity to learn* is functionally organized as an open loop or as a loop

⁴ For Piaget, "the essential fact to start from is that no knowledge, even perceptive knowledge, is a simple copy of reality, because it always involves a process of assimilation to previous structures. We take the term assimilation in the broad sense of integration with prior structures" (Piaget, 1967, p.22). An action pattern or scheme (*e.g.*, sucking or putting-together) can thus be said to be assimilative when the actions produced in the interaction do not encounter any notable resistance to their deployment. If this is not the case, the accommodation process is mobilized to modify the pattern.

including a retroaction. In other words, whether or not UR effects are linked to the occurrence of the next NS (future CS). We would argue against the idea of an open loop, as this would underestimate the role of the action of salivation and the often-overlooked positive reinforcement of satiety. As we see it, the formation of an associative bond needs to be interpreted within a broader framework, its significance being strongly associated with the repetition of an ordered scenario. This is a very general statement: the new relationship is not established as the result of a single occurrence. The very fact of repetition is an indication of the signifying circularity underlying the animal's existence. But it is also the constancy of the repetition of the sequence that enables the sequence to be perceived and the learning mechanism to be implemented, insofar as this learning mechanism corresponds to the creation and the stabilization of an anticipation.

To illustrate the point, we might mention recent findings in neuroscience suggesting that neural signals (for example, from dopaminergic neurons or in the prefrontal cortex) relating to reward predictions and reward prediction errors may reflect not only a current trial, but also past trials and sometimes anticipations (therefore estimations) of future trials (Seo and Lee, 2007; Enomoto et al., 2011; Wittmann et al., 2020). Such neural signals are therefore potentially useful for learning according to the Rescorla-Wagner model discussed above. These findings suggest that the animal does not learn the value of individual stimuli in isolation, but considers things in context, and tries to learn the overall value of this learning context, in addition to the value of the stimuli that make it up. Recent research in humans suggests that the learned value of context also influences behavior (Palminteri et al., 2015). A crucial question also arises here concerning the emergence of a particular, minimal form of reflexivity at the heart of learning and perhaps already present in the Pavlovian situation. As we mentioned above, this is what Kamin (1969) termed *retrospective contemplation*, a subject to which we will return later.

IV. Learning as active bonding: instrumental learning

Instrumental learning, due in particular to Thorndike (1913) and Skinner (1938), gave rise to different kind of research that placed the animal's activity at the forefront. What had been a passive vision of learning gave way to an active one, in which the exploratory dimension of behavior and the feedback (reinforcement) resulting from action took center stage. This brings us to our second definition of learning, seen from an associationist perspective:

Learning definition #2: Learning involves the creation of links between internal entities, resulting in the production or inhibition of a "response", depending on the nature of the reinforcement (reward or punishment) associated with it. The response is a behavior that must be instrumented (learned to use).

This definition reflects the view that learning leads to a progressive convergence towards organized, effective, satisfying behavior. Here, reinforcement is a basic form of feedback. Instrumental learning can take at least three forms:

- Serial learning, as in the case of behavior modeling (such as employed in dressage), where the positive reinforcement of individual elementary behaviors can be used to build up sequences of these behaviors. The qualitative aspect of serial learning also applies to sensory-motor learning.
- Discriminative learning (simple or complex) refers to all situations in which animals control their learning through actions that force them to make choices. The classic example of this is the maze, whether consisting simply of two branches one of which

contains a reward, or a more sophisticated device such as a Lashley vault or a multi-branch maze.

- Operant conditioning, proposed by Skinner (1938), has been studied mainly in rats (although it can be applied to any species, including humans) in setups featuring actuators (usually pedals) that can deliver reinforcement. Such setups are generic, insofar as only fairly simple technical adjustments are required in order to enable performances to be compared between different species. They are also able to exclude effects that might result from the proximity of a human experimenter.

Law of effect and "problem box" device

Although instrumental learning is most often associated with Skinner and his operant conditioning protocol, it should be remembered that this emphasis on action and its effects was formulated earlier by authors such as Thorndike (1913). The dominant view at the time was that the formation of associations in animals depends on the frequency of presentation of the events to be associated and on the temporal contiguity of their occurrence. Thorndike had a vision of learning causality that was clearly distinct from this dominant view. As we will outline below, his formulation of the law of effect also reflects a selectionist point of view. To test his theory that behavior is a function of its consequences, he placed a cat in a device known as a "problem box". The cat needed to find a way out of the box in order to receive positive reinforcement (food). Here we see the beginnings of Gestalt and cognitivist methodologies based on problem situations, which later gave rise to hypotheses on the internal mechanisms leading to the resolution of a problem. Like Skinner's rat-in-a-box a few years later, Thorndike's cat – which, it should be remembered, was kept hungry – produced the following stable sequence: the animal explores the limited space and produces a variety of behaviors (trial-and-error learning), it produces a response corresponding to the solution (opening the box), and it then learns to produce the "right response" (continuous learning = continuous decrease in the latency of the right response) to the detriment of other behaviors (selectionist aspect), ultimately producing only the right response.

Instrumental learning represents a significant shift in how learning was conceived, highlighting the agentive and relational dimension of the process. Learning, seen from this new perspective, is fully visible⁵. The animal manifests its own capacity to perceive relationships and organize its behavior to produce stable effects, albeit in a less than ecological way.

Reinforcement learning theory

From a computational point of view, instrumental learning means extending the Rescorla-Wagner model such that reward-predictive value is no longer linked only to stimuli, but also to sequences of actions leading to reward. This was formalized as the reinforcement learning (RL) theory (Sutton and Barto, 1998), in which it is assumed that animals learn by trial and error to select the actions that help them maximize the sum of future rewards. Importantly, the RL theory not only generalizes stimulus-value learning to action-value learning, but it also covers the reinforcement of actions whose positive consequences are delayed in time. In other words, while Rescorla-Wagner is able to reinforce only events (*i.e.*, stimuli) that were immediately followed by a reward, the RL theory can reinforce any event or action leading to a state of the task where the expected reward value is higher than the expectation before the event or action. Stated differently, any action bringing the animal closer to the reward is to be reinforced, even

⁵ This partly explains why Skinner was the first promoter of a large-scale educational technology project in the United States and beyond, and why attempts to return to a different, notably constructivist, understanding of the mechanisms of learning have been seen as neo-behaviorist (neo-Skinnerian, one might say) approaches, as in the case of enaction.

if subsequent actions are required before the reward is actually delivered. The reinforcement equation in the Rescorla-Wagner model thus needs to be rewritten as follows:

$$V_{t+1}(s) = V_t(s) + \alpha(r_{t+1} + \gamma \mathcal{W}_{t+1}(s') - V_t(s))$$

where a new parameter, γ (with $0 < \gamma < 1$) has been added. This is termed the discount factor and it enables a smaller weight to be given to $V_{t+1}(s')$, the expectation of future rewards, than to the current reward r_{t+1} .

Comparing this new model with the initial Rescorla-Wagner model, three important changes may be remarked. First, rather than distinguishing a value corresponding to each stimulus k , here there is a global value corresponding to each state s of the task, no matter how many stimuli this state might include. The global value of each state s can be decomposed into the specific values of the different stimuli k present in that state, but for simplicity we do not show this here. Second, we have removed the β parameter, which means that this formulation neglects saliency. Third and most importantly, the prediction error used for computing the reinforcement signal now has three terms rather than just two: instead of simply comparing the reward obtained r_{t+1} with the expected value $V_t(s)$, we now compare the reward obtained r_{t+1} plus the expectation of future reward $V_{t+1}(s')$ in the new state s' of the animal once it has performed an action (for instance, pressing a lever that turns on a light) with the previous reward expectation $V_t(s)$ in the state s of the animal before the action takes place (*e.g.*, while the light was off).

This third element is what enables RL models to reinforce actions leading to delayed rewards. In our example, once the animal has learned that state s' – where the light is on – is predictive of reward (*i.e.*, $V_{t+1}(s') > 0$), because the reward will be received at the next timestep: $r_{t+2} = 1$), any action that can lead the animal to state s' (*e.g.*, pressing a lever) gets reinforced because it yields a positive prediction error even when the reward has not yet been delivered (*i.e.*, $r_{t+1} = 0$): $r_{t+1} + \gamma \mathcal{W}_{t+1}(s') - V_t(s) = 0 + \gamma \mathcal{W}_{t+1}(s') - 0 = V_{t+1}(s')$, which is positive⁶.

For the same reasons, it is this third ingredient that enables RL models to learn sequences of actions leading to the reward. Sticking to our example, once the animal has learned that pressing the lever in state s is consistently followed by state s' , itself consistently followed by a reward, the reward-predictive value corresponding to state s becomes positive (*i.e.*, $V_t(s) > 0$). Then, any action that can lead the animal to state s (*e.g.*, pulling a chain) in turn starts getting reinforced. And so on, and so on.

The new model generalizes the same learning mechanism – learning from prediction errors – to both classical (Pavlovian) conditioning and operant (instrumental) conditioning. This is because learning from prediction errors enables any kind of event, whether a stimulus or an action, to be reinforced. The third element as described in the previous paragraph is what enables RL models to explain second-order conditioning, which Rescorla-Wagner fails to account for: if stimulus A is followed by stimulus B, itself followed by a reward r , in RL models the value of stimulus A will be reinforced because it is followed by a positive prediction error: $r_{t+1} + \gamma \mathcal{W}_{t+1}(B) - V_t(A) = 0 + \gamma \mathcal{W}_{t+1}(B) - 0 = V_{t+1}(B)$, which has become positive after B has been repeatedly followed by reward. In contrast, in Rescorla-Wagner, $r_{t+1} - V_t(A) = 0 - 0 = 0$, thus the value of A remains zero because it is never immediately followed by reward.

Strikingly, the RL theory has played a key role in identifying the neural mechanisms of associative learning. In the case of a monkey performing a Pavlovian conditioning task, dopaminergic neurons were found to respond according to a reward prediction error signal consistent with the three-term equation in the RL theory (Schultz et al., 1997): they respond to

⁶ For simplicity, here we just show how the value of states $V(s)$ is learned, and not how the specific value of actions performed in those states $Q(s,a)$ is learned. In the RL theory they are updated in the same way, *i.e.*, as a function of prediction errors. Interested readers can refer to Sutton and Barto (1998).

unexpected reward, but no longer respond when the reward is well predicted by the animal; after learning, they respond to stimuli that have acquired a reward-predictive value. Interestingly, during second-order conditioning (of the type A->B->reward), the same recorded dopaminergic neuron initially responds to the reward, then progressively shifts its response to A, which is the first event causing the animal to anticipate reward delivery (Maes et al., 2020).

Overall, the RL theory has contributed not only to formalizing the associative mechanisms by which stimuli and actions become reinforced through learning, but also to more precisely describing the neural substrates of this learning process.

Criticisms of associative learning

While the behaviorist framework may be suitable for thinking about simple instrumental learning, or even perceptual learning (*e.g.*, Roelfsema et al., 2010), it is manifestly inadequate when it comes to addressing what are seen as more abstract forms of cognition. In the following section we discuss what makes the behaviorist framework unsuitable for this purpose. What is more, a behaviorist paradigm of learning is essentially centered on one individual alone, and the contribution of other individuals, not just as perceived objects but as actors in the learning process itself, is not problematized. But let us return briefly to the question of perception. Although instrumental learning may be linked to movement, movement is not seen as playing a role in the formation of percepts. The role of movement is essentially exploratory. It has no organizing or morphogenetic function in relation to stimuli other than to define a temporal framework within which associative links can be constructed. Here we have an echo of an overarching empiricist world-view in which perception is essentially a matter of unconscious processes that do not involve, or presuppose, action. To put it bluntly, in the behaviorist tradition there is no theory of the internal mechanisms of perception and learning other than an aggregative mechanism of sensations. If we are to explore these internal mechanisms, we need instead to look to cognitive approaches. One of the main criticisms levelled at behaviorist studies of learning is precisely that they necessarily consider only the inputs and outputs of the animal-system (conceived as a black box) in their attempts to characterize laws based on the coupling of stimuli and responses. The behaviorist position has also taken extreme forms, certain authors adopting what we might qualify as a militantly empirical approach and describing their work as anti-theoretical. However, as Richelle vigorously points out (*e.g.*, 1953, p.53), we must guard against over-hasty and caricatural judgments. We have to recognize that behaviorist work is by no means homogeneous, and the tensions that divided the community in the past are in some respects still relevant today. Researchers like Hull (with an interest in formalisms and theory, and putting forward hypotheses on internal mechanisms that involve intermediate variables) are clearly at odds with the behaviorist stereotype. An important development was a gradually increasing interest in representing internal events as embodying knowledge, to the detriment of the associations between these events.

Finally, a crucial point, often raised in the ethology community, concerns the non-ecological nature of all these situations and the difficulty, if not the impossibility, of replicating them when animals are observed in their natural, meaningful environment. Gardner (1991), for example, considers that the behavior of animals in the Skinner box, and in particular the pressing of the pedal, is intertwined with a stimulation of the palmar surface of the paw, an action that animals habitually perform in their natural environment, and therefore needs to be interpreted accordingly. In addition, there is evidence to suggest that conditioned animals continue to press the pedal even when offered food in close proximity, which might point to an almost "pathological" aspect to this behavior.

Among the criticisms of associative learning, we could add a more recent objection, namely that reinforcement learning models focus on maximizing expected rewards while neglecting the interest of reducing uncertainty about perceived states of the world (Friston et al., 2009). This objection might be countered by including in the learning model some notion of uncertainty minimization alongside that of reward maximization: instead of learning to associate a stimulus or action or event simply with an average predictive reward value (as is the case in the RL theory), we could use a Bayesian framework to learn to associate it with an entire probability distribution. From a learned probability distribution, we could then extract: (1) the mean, corresponding to the average value learned by reinforcement learning models; and (2) the variance (i.e., the inverse of the distribution's "precision"), representing a level of uncertainty in this estimate. This uncertainty could be a key factor in modulating not only the strength of learning (e.g., from a normative point of view in Bayesian theory, greater uncertainty in prior knowledge should motivate faster learning), but also the animal's level of attention (replacing the additional term potentially added to the Rescorla-Wagner model in Panayi, Khamassi et al, 2021, for example). An assessment of uncertainty could serve as a motor for an animal's curiosity, a high level of uncertainty increasing its eagerness to learn. This may be related to the notion of directed exploration, or curious exploration, which has been formulated as "active inference" in Bayesian terms (Friston et al., 2015), and to which we will return later. While Bayesian formulations of reinforcement learning (learning probability distributions rather than mean values) have since been proposed (e.g., Ghavamzadeh et al., 2015), this initial critique of reinforcement learning theory – and associative learning in general – has had the merit of highlighting the need to consider the role of animals' active motivations to learn, to explore in order to acquire new knowledge. We discuss this topic in detail in the following section.

V. Learning as active motivation to acquire knowledge

We now step aside from what has been a chronological account, in order to present some recent contributions from computational models that seek to include the active dimension of learning described above. We believe that it is important to highlight some of the mechanisms postulated before we move on to other forms of learning, such as cognitive map learning.

An initial observation is that in most behavioral experiments involving animals, the animals will seldom behave identically in 100% of the trials. Where an animal has learned to press a lever to obtain food, there will always be a trial here and there where it does something different. We could possibly consider this as noise, statistically speaking, and simply assume that our computational models can never account for 100% of observed phenomena. But there seems to be something else, something qualitative that has the potential to inform us about the active nature of learning.

In computer science, we have long known that for a learning algorithm to work, it must have a mechanism that makes it do something different from time to time, even after learning has apparently converged on an acceptable solution. The reason for this is that if nothing new is ever attempted, the optimality of the status quo is never called into question. More importantly, trying something new, or at least different, from time to time, even after learning, when we think we know the right solution, enables us to adapt more quickly to changes in the rules governing a task, or to changes in the environment around us. We call this an exploration/exploitation trade-off. Most computational conditioning models employ a very simple but effective solution: they include a parameter representing an exploration rate ϵ . For example, 1% of the time the model will choose an action completely at random – so as to encourage exploration of different actions – and 99% of the time it will choose the action that

is associated with the highest predictive reward value V . In the RL theory this is known as an epsilon-greedy strategy (Sutton and Barto, 2018).

More recent computational models, however, have proposed mechanisms to make this process more dynamic, which can make them even more effective in terms of maximizing gains, and which can give a better account of certain experimental data. For example, it may be supposed that animals will vary their rate of exploration: when they detect that the task or the world has changed, they could deliberately explore more; conversely, when they see that the task and the world are stable and familiar, they could deliberately explore less, so as to minimize surprise and maximize their gains. Mechanisms for varying the rate of exploration ϵ according to the estimated stability of the task have already been proposed, and seem to better account for the behavior of macaques in regard to certain tasks, as well as for neural activities in their prefrontal cortex that appear to underlie dynamic exploration (Khamassi et al., 2015).

Another way making the process dynamic is by having the animal choose not only its rate of exploration, but also specific actions that favor exploration. If actions are chosen randomly during exploratory trials, then previously explored actions will sometimes be re-performed, with no new information being gained. If, however, estimates of action-reward uncertainties are retained in memory, then exploration can be directed towards actions judged to be more uncertain. We call this a mechanism of directed exploration, where greater value is given to uncertain actions whose execution could help reduce uncertainty and thus increase knowledge. This type of exploration would appear to exist in humans (Wilson et al., 2014) and other species (Costa et al., 2019). Directed exploration has been linked to active exploration, curiosity in children, and developmental psychology (Gottlieb et al., 2013). There are also links with Karl Friston's active inference theory (Pezzulo et al., 2015), which holds that a central motivation behind action is seeking to minimize uncertainty in our internal representations.

The mechanisms proposed in computational models to account for active exploration take us even further away from the idea of a passive mode of learning in which there is first perception and subsequently action. Here, we have quite the opposite: we consider that the animal acts, even possibly in the absence of the possibility of reward (*i.e.*, without a reinforcer in the sense of operant conditioning), with an intrinsic motivation to acquire knowledge. And when a reward is available in the environment, this mechanism links up with the central learning mechanism (such as that of the Rescorla-Wagner model seen above) to produce active exploration.

VI. Learning as the constitution of knowledge nodes

As we have seen, learning was initially conceived as the constitution of links between entities, usually presumed to be neurobiological, with this constitution resulting from a repeated exposure to the CS-US couple and/or from observing the effects of action (feedback having the status of positive or negative reinforcement). However, against this initial conception, we have an accumulation of more recent empirical findings showing, on the one hand, that learning can involve phenomena that are apparently strictly internal, independent of stimuli and reinforcements, and, on the other, that the transformations specific to the learning mechanism concern not only the links but also, and perhaps above all, the nodes of the network. In the emerging cognitivist tradition, the nodes are seen as being representations that drive behavior.

The cognitive map hypothesis

Tolman (1948) was an early proponent of an alternative to behaviorist positions on learning. Two of his experimental situations, namely the three-pass maze situation (spatial learning), and the latent learning situation, are often cited in opposition to behaviorism. These two experiments tend to show that learning involves internal events that can be qualified as

cognitive and that can be assumed to mobilize representations and “calculations” relating to these representations. Learning is conceived here as the acquisition of knowledge about the external world, to be mobilized to achieve a goal. Tolman and Honzik (1930a) taught rats to move through a maze including three branches (A, B, and C) of different lengths. After a phase of free exploration of the whole maze, the rats are divided into two groups, and each group has to learn how to get around an obstacle. The predictions of behaviorist theories based on associative force postulate that rats will systematically opt for the shortest path, given the obstacle present in direct path A. This is indeed what was observed for rats in group 1, for whom the obstacle was close by. However, this is not the case for group 2. Rats prefer to take path C, which, despite a longer route, allows them to bypass the obstacle. From this, Tolman developed the idea of a cartographic representation of space (a *cognitive map*), formed during the exploratory phase, and used by the rat to organize its movements and reach its goal. Computational modeling has seized on this notion and formalized it as a different type of learning from the reward-based learning represented in associationist models of the Rescorla-Wagner type (Redish and Touretzky, 1997; Arleo and Gerstner, 2000). The learning suggested by Tolman, and which he termed *latent learning*, occurs during a phase where there is no need for a reward. It consists in learning the probability of transition from one state to another as a function of action. In other words, it consists in learning the arrows between the nodes of a cognitive map. Then, once the position of a reward is known, a mental simulation drawing on a cognitive map enables the rat to find the shortest path between its current position and the reward. Interestingly, despite a seemingly irreconcilable opposition between behaviorists – advocates of Rescorla-Wagner-type associationist learning mechanisms – and cognitivists – advocates of cognitive map-type mechanisms – a growing body of work suggests that mechanisms linked to the two types of learning co-exist in the brain (Daw et al., 2005; Khamassi & Humphries, 2012). What is more, it would appear that certain behaviors observed in complex tasks combining both spatial cognition and generalization-type phenomena (*i.e.*, phenomenon 2 presented above) *can only* be explained by models that include both types of learning, *i.e.*, Rescorla-Wagner and cognitive map (Dollé et al., 2018). The cognitive map hypothesis suggests the possibility of latent learning of structural regularities in the task or environment, which can then be reused to adapt behavior to new situations.

Latent learning

In the latent learning experiment (Tolman and Honzik, 1930b), rats were taught over a period of 18 days to navigate a complex maze to obtain food rewards. Three groups were then formed: rats in the first group were no longer rewarded, rats in the second continued to be rewarded systematically, while rats in the third were rewarded only from day 11 onwards. It was found that the rats in the third group performed at the same level as those in the second from the very first reinforcement on day 11, and that their performance was even better after that. The experimenters concluded that the rats were able to learn even where learning was not reinforced, and therefore that reinforcement does not condition learning itself, but rather its behavioral expression.

More recent experimental situations have also fueled this conceptual renewal (Holland and Straub, 1979). These include the so-called conditioning reappraisal situation, where classical conditioning is first established, after which the presentation of US is associated with a nausea-inducing toxin. When CS is subsequently presented, it no longer induces CR, even though the CS-US association has not been manipulated or modified. This suggests that US modification alone is sufficient to alter behavior and thus associative functioning. Results obtained with the so-called mediated learning protocol have implications similar to those that can be drawn from results obtained previously (Holland, 1984).

In this way, the refocusing on internal representations and on ways in which these representations may be manipulated has enabled the behaviorist community⁷:

- to integrate results obtained later showing that event manipulation, and not only inter-event links, is likely to constrain learning,
- increasingly show that even the architectures enriched by Hull do not succeed in capturing the full range of results,
- and, above all, to open up new interpretative opportunities for managing phenomena that are not directly observable in behavioral terms, but which undoubtedly have effects⁸.

Cognitivist approaches

In the nascent cognitivist tradition following Tolman's experiments, the question of learning tended to become marginalized and eclipsed by questions relating to memory and information processing operations.

Learning definition #3: Learning is a mechanism for acquiring and/or modifying knowledge.

This definition reflects, for example, the viewpoint of Lindsay and Norman (1972), who describe their conception of learning as the incorporation of new information into semantic structures already stored in memory. Bruner (1957) sees learning, particularly perceptual learning, as an active process of internal categorization, while Bartlett (1932) and Vernon (1954) see it as the construction of a schema. Cognitivist work on learning has been highly inventive in terms of hypotheses and models, once again to account for internal operations that are difficult to observe, all the more so since the interpretation of certain results suggests a production/acquisition of new knowledge that does not correspond to external events.

The marginalization of learning as a subject of study *per se* also led to a relativization of its perceived role in the modulation of behavior; new models focused in particular on attention and motivation, leading to a distinction made between learning and performance. Learning, as a mechanism⁹, is defined in this context as a transformation operation encapsulated in rules specifying actions to be performed in response to successive elements of information received (system input).

The role of feedback

As already mentioned, the cognitivist tradition has preserved the idea of acquisition through action. Another legacy of instrumental learning is the idea that reaffirmed information linked to activity can have the status of feedback. The notion of feedback has evolved significantly, in

⁷ It is interesting to note, in the light of recent work on animal behavior, that authors in the field such as Pickens and Holland (2004) no longer hesitate to redefine conditioning as involving associations between internal representations: "We described conditioning as involving the establishment of associations between internal representations of the CS and US, such that the CS comes, through learning, to activate a representation of the US".

⁸ Although epistemologically distinct from the cognitivist tradition, we recall that the Gestalt school of psychology, in particular Köhler's (1925) work with great apes that sought to reveal interspecific discontinuities in learning, specifically studied the phenomenon of insight, by which is meant the impromptu resolution of a problem that suggests the presence of an underlying internal mechanism. Specifically, Köhler examined the use by chimpanzees of interlocking sticks in order to extend their reach and attain a food reward. This ability is thought not to have emerged in evolutionarily earlier species. Subsequent work has clearly shown that the emergence of this behavior occurs only in animals that have previously acquired the interlocking technique; this does not subtract from the interest of the observation, but it does limit the scope of the supposed internal activities.

⁹ In this case, a particular combination of symbolic processing operations that produce or change knowledge.

line with the conceptual developments that cognitivism has heralded. Feedback, as an informational fact (*i.e.*, carrying a certain quantity of information, in the Shannon and Weaver sense) continues to be thought of as being above all a means of measuring success or failure, or deviation from a goal. But feedback can be involved in more complex, more internal modes of operation, such as in comparisons, anticipations or inferences¹⁰. Feedback information linked to an agent's action is often tested against or supplemented by knowledge relating to the rules of action or the properties of objects. Wiener (1954), the father of cybernetics, saw learning as being essentially a form of feedback, in which a pattern of behavior is modified by past experience. For him it is a type of feedback that is complicated in its form, exerting its influence not only on action, but also on the pattern of action. In other words, the whole activity of an agent is affected by learning that appears to be local in nature.

Mechanisms for acquiring new knowledge

Numerous mechanisms have been envisaged to account for the learning/acquisition of new knowledge¹¹. They have all given rise to models and empirical studies that it would take too long to describe here. What is indisputable is that by considering learning as comprising operations that cannot be directly observed, the cognitivist tradition has to some extent exposed itself to intrusion and inflation, which has been damaging insofar as it has not always provided itself with the means to control what has been claimed in its name. The mechanisms that have been proposed can be divided into three groups:

- Incremental mechanisms that concern the quantitative modification of knowledge (*e.g.*, increasing the availability of a lexical item given its frequency of use).
- Generative knowledge mechanisms based on episode storage (coding and selection according to multiple criteria such as novelty, relevance, etc.).
- General generative mechanisms (inference and memorization) such as assimilation, analogical transfer, induction, hypothesis testing, and generalization.

Implicit learning

Among the different mechanisms that have been put forward involving an acquisition of knowledge via processes that are not directly observable, implicit learning can be seen as something of an extreme case. Implicit learning is where learning takes place without there being any attention paid to, or any awareness of, the object of learning. A field of research has emerged around the question, which would appear to show that despite the apparent paradox in the idea of learning about things without knowing about them, implicit learning setups do in fact mobilize subjects' attention, and attention seems to be central to learning. In this research, subjects are confronted with a situation structured according to complex rules that they cannot

¹⁰ In this way, we can see that the computational-representational approach has made a shift in relation to earlier approaches, since feedback plays a clearly secondary role; the point of view is completely centralist and internalist, and behavior is determined by one or more goals.

¹¹ From this point of view, Anderson's (1983) ACT theory has been a reference model for modeling cognition and knowledge acquisition. Anderson classically distinguishes two types of knowledge (procedural and declarative) and postulates three memory instances (declarative, procedural and working memory). Acquisition is the permanent storage in working memory of processes that are taking place. Storage mechanisms for declarative knowledge and matching mechanisms for procedural knowledge ensure transfer. Modeling can become sophisticated, as in the case of procedural knowledge acquisition, where following an initial interpretation phase (the selection of declarative information to allow a choice of actions to be made), a second compilation phase transforms the declarative elements into real know-how. Two mechanisms, proceduralization and composition, ensure this compilation. Other, more local models have been proposed, such as the BAIRN model (Klahr and Wallace, 1976), where the system creates knowledge by creating new nodes either from the combination of several nodes, or from sequentially stable traces of operation, or by eliminating redundancy.

apprehend (nor are they asked to). A classic result (Lewicki, Hill and Bizot, 1988) has been to show that in sequential reaction time tasks (where subjects are simply asked to press keys on a keyboard arranged in a topographically similar way to the display of targets (4 on a screen), subjects have significantly shorter reaction times when the succession of positions is governed by a rule than when it is random. Although subjects are unable to make explicit the rule (*e.g.*, repetition of the same sequence of 10 items) that determines the appearance and position of targets, and may even have the feeling that the display is random, they manage to build up some anticipation. While it has been shown that attention is important in obtaining this effect (by adding an additional task that makes it disappear), an analysis of the succession of positions reveals that dynamic regularities are nevertheless perceptible. As some authors have suggested and shown (Perruchet, 1997), the object of learning is "the very object of phenomenal perception of the world, which changes under the effect of repetition". These experiments exploring the learning of skills appear to show that this learning does not concern the abstract rules that govern the production of objects perceived, but rather the possibility of stabilizing and identifying regularities. However, the artificial nature of these situations obviously calls into question the significance of the results obtained. Those results might suggest that learning cannot take place without attention and, above all, without the possibility of perceiving, either in the behavior of objects or in the subject/object relationship, certain regularities that can subsequently favor successful anticipation. An interesting point is that these perceived regularities are apprehended by subjects at a level distinct from that of the causal system that generates them (that is to say, a system that can be expressed in the form of complex rules). The conditions allowing these perceived regularities to be constituted remain to be clarified. It is possible, as Perruchet and Vinter (2002) have done, to appeal to Gestalt by invoking processes of automatic structuring of the perceptual field, but other explanatory frameworks, which we will discuss later, are also conceivable. However, these results are interpreted, they indicate quite clearly that learning, as a basis for successful anticipation, is possible even in environments presenting local regularities that the subject manages in a pre-reflective mode. Where subjects are unable to state the rule for determining the succession of target positions, it is only a retrospective analysis of the situation that may enable them to pinpoint local regularities such as to constitute learning at a perceptual level.

Learning by observation

There is another family of situations where learning seems to take place not only on the basis of internal processing independent of directly perceived events, but also without any involvement or action on the part of the subject (human or animal) in the situation; here, learning appears to follow merely from observing the behavior of others. We speak of observational learning when an animal-learner, having observed a certain sequence of events involving an animal-model in context, manages to perform an activity faster and better than in a control situation where observation is not possible. Through observing the other animal and the situation, it as if the animal had performed the activity itself. In some cases, the animal reproduces the observed behavior identically (imitation), but in other cases there is some distance between the observed behavior and the animal's own behavior following observation. Observation can even be an opportunity for an animal to learn which behaviors are best avoided. This was demonstrated in the seminal experiment by John, Chesler, Bartlett and Victor (1968), in which the learners (cats) observed an aversive conditioning situation and produced the avoidance response, some on the very first try. One interesting conclusion to be drawn from their results is that observing "naïve" models can be more beneficial than observing "expert" models, which corroborates the importance that the observation of mistakes has for this type of learning. More recent work on mice shows that observational learning has some interesting intra- and interspecific features. In a series of studies, Carlier and Jamon (2006) compared two

groups of mice in a dual task in which the mouse first has to push the food reward to the end of an angled tube, then open a drawer at the other end of the tube. It thus has to learn a sequence of two distinct actions to be executed in two distinct locations. The model mice selected are over-trained experts that are also capable of performing the task in front of other mice (whose presence they are aware of via their scent). It is worth noting that some experts are disturbed by the presence of the others (which leads the experimenters always to have a pair of models just in case). In this situation, the two groups – experimental and control – are distinguished by the fact that only the experimental group can see the models perform the task. The controls can see the device and have an exchange with the models, but never see them perform the task. The results show a significant learning facilitation effect in the experimental group that is not seen in the control group; however, not all the mice in the experimental group accomplish the complex task on the first try (only 6 out of 15). The authors suggest that learning during the observation phase focuses not so much on actions as on the models' areas of interest, and is coupled with observational conditioning that enables the learners to acquire the sequence of actions required. It should be noted that in this type of situation, it is generally impossible to control what the experimental animals actually observe. Even where the experimental space is restricted, this does not guarantee that the animals' attention is focused on the behavior of the models. Moreover, there are differential effects depending on the species. Work on similar problems has been carried out with turtle doves and blackbirds (Carlier and Lefebvre, 1997). Corvids are an interesting case, as these long-lived animals (around 40 years) display i) long ontogeny (close to that of primates and humans), ii) cerebral hypertrophy (greater than that of humans) of certain cortical territories, and iii) rapid learning capacities throughout life. Studies show that the blackbirds can learn very quickly by observing a conspecific or an experimenter (Carlier and Lefebvre, *ibid.*); it is often necessary to hide the first manipulation of the device, as this may be enough for the animal to satisfy the task of opening a box on the first try. Doves, on the other hand, may take several weeks to complete this learning process once they have had the opportunity to see a model. Furthermore, the authors (*ibid.*) were also able to show, by comparing two turtle-dove families (gregarious versus territorial) of the same species, that only one of the two families benefited from observational learning even though they co-evolved together¹². Underlying this type of research on observational learning and the diversity of its forms are obviously some highly relevant questions that cannot be developed here concerning the role played by the perception of other individuals, the actualization of perceived and imagined action at the level of the subject, and the effect of this actualization on the subject's own actions.

Learning to learn

To conclude our list of works showing how sophisticated internal cognitive activities are implemented in learning situations, we might recall the results obtained by Harlow (1949) with non-human primates showing that animals are capable of learning to learn, or in other words, that they appear capable of progressively generalizing rules that reduce the learning time for different sequences. Within the cognitivist tradition Harlow's results are noteworthy, insofar as they point to internal processing possibilities whose inferential nature may be said to exceed the association of perceived events.

This notion of learning to learn has been modeled computationally under the label of meta-learning. It has its origins in methods initially developed in the field of artificial intelligence for revealing structure in a data sets, so as to accelerate learning when data of the same type are

¹² Studies of dogs show that they outperform their ancestors (wolves and jackals) in orientation tasks. This is especially noticeable in the case of sheepdogs and hunting dogs. A similar superiority is found in primates executing the same type of task.

encountered in the future (Schmidhuber, 1987; Giraud-Carrier et al., 2004). This search for structure is generally carried out off-line, *i.e.*, not in real time. It may, for example, involve exploring the best models or learning algorithms for this type of data, or the most appropriate learning parameter values.

These methods, adapted for online use, have been applied to the modeling of learning behaviors in humans and other animals in studies seeking to shed light on potential behavioral adaptation dynamics accompanying the execution of tasks. Some work has focused on online learning of the values of learning model parameters, such as exploration rate, thus enabling the creation of an active exploration process such as that described above (Schweighofer and Doya, 2003; Khamassi et al., 2011). Interestingly, these types of computational mechanisms are able to reproduce the behavioral dynamics that Harlow observed in monkeys: the time required to learn which of two objects is linked to a reward is seen to decrease from sequence to sequence. Another body of work has investigated mechanisms for arbitrating between different types of learning in the brain (Daw et al., 2005; Lee et al., 2014): mechanisms of this kind detect which type of learning (stated simply, learning either with or without a cognitive map) is most effective at different points in a task. In some variants of these models, a meta-controller, learns – and therefore memorizes – which type of learning has proved most effective in which situation, and can therefore contribute to generalization in new, similar situations (Dollé et al., 2018).

Finally, recent work has modeled certain meta-learning mechanisms in the prefrontal cortex as corresponding to a reconfiguration of a deep neural network that enables implicit encoding of network operating rules, which can then mimic learning-like behavioral dynamics without involving any further need for a learning mechanism *per se* (Wang et al., 2018).

Provisional remarks on cognitivist approach of learning

Within the cognitivist tradition, despite only a limited interest in the question of learning itself and a greater focus on memory mechanisms, elements have nevertheless been identified and clarified that have given rise to some computational modeling.

The first of these elements is the role of action as a condition for learning, although cognitivists recognize that the main function of action, insofar as it is itself guided by already-constituted knowledge, is to extract information about the external world – information that is specifically coded in such a way as to make it compatible with, and contribute to, cognitive computation. We have seen how active exploration mechanisms can guide action, with the aim of reducing uncertainty in our internal representations. Models of such mechanisms often feature an exploration bonus in terms of information entropy (in the Shannon sense), reflecting the amount of information that can be acquired by taking action. This type of approach explicitly neglects the question of action (cf. Marr's scientific project, 1982). What is more, the form of an action and the effect of this form on the construction of cognitive experience has only rarely been taken into account. And when it has been, the form of the action itself has generally received less attention than the sequence of actions to which that action belongs, insofar as this sequence reveals the implementation of internal problem-solving strategies (Bastien, 1997).

A second element is the consideration of feedback and its potentially reconfiguring effect on cognitive structure. The reconfiguration that Wiener (1954) considered necessarily global is linked both to the meaning of coupling and to the constitution of anticipation, which corresponds to the possibility of changing function. From this point of view, what Wiener wrote is extremely relevant, even though his conception of feedback applies to formal, discrete cognitive architectures, the limits of which are well known today (c.f. the body of critical work on the computational-representational approach). In mathematical terms, the question of feedback has often been linked to the notion of recursivity, since it involves "reasoning about functions in which at least one variable is the value provided by a previous application of the

function" (Livet, 2004, p.130). Linking feedback and recursivity in this way must nevertheless remain subject to the following constraint, highlighted by Livet: "We cannot consider as a recursive function a function that would be a function of both its input x and its output y at the same instant, so that x depends on y which depends on x . In order to find it, we obviously have to decompose the function into two variables. To find it again, we obviously have to shift the inputs and outputs in time, and admit a t_0 where x is given without y yet having any influence, which it will only have at time $t+1$ on input x in $t+1$ (assuming feedback takes place without delay)" (Livet, *ibid.*, p.130).

The third element is that learning appears to mobilize mechanisms that are essentially internal and, for the most part, unconscious, but which nonetheless presuppose attentive/perceptive awareness on the part of the subject. For our part, we consider it highly problematic (from an epistemological and a scientific standpoint) when work on perception takes no account of subjects' own descriptions of their perceptual experience. Such descriptions should constitute the very object of a science of experience, particularly perceptual experience. The works of Varela, Thompson and Rosch (1993) and Shanon (1993), among others, are quite explicit about this, having paved the way for a substantial body of theoretical and methodological work on experience in relation to the 3rd and the 1st person (Lenay, 2006; Stewart and Gapenne, 2004; Varela, 1996). However, we believe that it is perfectly possible to think of certain fascinating observations that have been made concerning implicit or observational learning, or insight phenomena, within a constructivist framework of the genetic type¹³. An examination of these different situations shows that if internal cognitive activity is taking place, it is at least alternating, if not simultaneously, with phases of engagement. Moreover, we believe (and this is a central point) that this internalization of activity is the result of a necessary, and sometimes very long, history of moto-sensory couplings, and that it basically calls into question the internal nature of the processes. Perception is often seen as an active process of encoding and representing information, and it is our belief that this process is an internal activity in which representations are constructed. This conception is obviously distinct from a conception whereby perception is a relational act, which continues to be an act even where it does not mobilize an overt commitment on the part of the organism.

A fourth and final element, complementary to the previous one, is that learning is sensitive to cognitive and situational context. This is at the heart of many theories of perception, in particular. Perception, as a fundamental cognitive act, is structured in relation to a social, cultural and technical context, and in relation to knowledge constituted by the subject. Thus, while learning involves processing incoming information (inputs) relating to the precise object of the subject's activity, this processing is constantly modulated by incoming elements that are nevertheless present in the situation, although potentially irrelevant to the realization of the task, and by knowledge possessed by the subject in relation to this situation it is already known.

VII. Learning as the constitution of an attractor in a network of interconnected elements

The 1970s saw the spread a host of hitherto unfamiliar ways of thinking about how phenomena are *caused*. Concepts such as emergence, complexity, qualitative physics, and distributed causality, began to percolate into numerous fields of science, and the question of learning was no exception, with a study of the behavior of formal neural networks, extending earlier work in neurocybernetics, being an area of predilection. These artificial networks, inspired by biological

¹³ We make this point because many researchers would have been quite happy with a sensorimotor framework reserved for constructivism and a separate conceptual/abstract framework reserved for cognitivism. Such a separation is obviously unacceptable, but the fact that it was considered desirable by some reveals a monolithic vision of constructivism that is highly problematic.

neural networks, have remarkable learning capabilities, and modeling them has been and still is the subject of a great deal of research (particularly in the field of machine learning). The study of these networks has had the great advantage of renewing the mathematization of the learning process:

Learning definition #4: Learning involves mechanisms at play in a system comprising nodes (to which functions apply) and links, and in which weights change over time (*cf.* Hebb's Law) such as to constitute a major constraint on the learning mechanism as an auto-organized convergence.

In this sense, connectionism proposes the introduction of a sub-symbolic¹⁴ algorithmic layer likely to clarify certain symbolic boxological models which basically present themselves as a special case.

Neural networks

Classifying and finding relationships are the two mathematical problems posed by learning in neural networks. A neural network is technically defined as a universal function approximator. It has an architecture and a learning algorithm (of which there are several) to build up a finite base of examples. One of the interesting properties of these networks is their ability to self-organize, which may lead us to change the way we think about cognitive models of learning, and enhance our ability to estimate mechanisms at the so-called microcognitive scale. . In this vein, we may note a set of connectionist models that address the process of learning of cognitive maps in the form of self-organizing maps (Kohonen, 1991), which not only allow self-organization between the nodes of the graph to account for the topology between places in the surrounding space, but which can also incrementally increase the number of nodes and links according to the need for information processing (*e.g.*, when new locations are discovered in the environment; Khamassi et al., 2006). Alexandre (2004) has proposed a clear synthesis of the possible correspondences between the mechanisms (supervised or unsupervised) and architectures (single- or multi-layer, unidirectional or recurrent) of formal and biological learning.

While neural networks represent a remarkable step forward in our understanding of cognition, insofar as they present a vision of cognitive micro-computation that includes notions relating to distributed dynamics and the genesis/emergence of these dynamics, these abstract models have remained based on a profoundly associationist, internalist and representationalist conception of learning; the stable configurations (that may even be expressed in terms of energy functions) that result from learning are not seen as anything other than representations of events in the world. Neural networks are currently enjoying a new lease of life, as well as renewed popularity thanks to the recent success of deep learning – *i.e.*, training a neural network with many hidden layers to solve statistical problems such as classification and regression on large amounts of data. This is yet another example of the statistical efficiency of self-organization and distributed dynamics. In addition to the initial neurobiological inspiration behind neural networks (formal neurons having originally been inspired by biological neurons and the strength of their synaptic interconnections) we note that they have also drawn a finer-grained inspiration from more complex mechanisms, such as convolutional networks (LeCun and Bengio, 1995). Convolutional networks integrate the results of parallel processing on a set of neurons so as to recognize perceptual patterns at different spatial and temporal scales.

¹⁴ "Information processing is not described as algorithmic, but as a spread of activity through cortical areas leading to self-organization of highly distributed spatiotemporal patterns which represent the computational result." (Engel and König, 1993).

Developments such as these have played an important role in recent successes. Artificial deep neural networks are now being used to ascertain which of their computational mechanisms might help understand vision-related sequential processing (in particular, but not exclusively) in different brain areas (Yamins and DiCarlo, 2016).

However, while some neural networks have fully acquired the status of applications in a number of industrial fields, and irrespective of their origin and their relevance to learning in a living cognitive agent, new areas of relevant research have continued to open up. These include neuroscientific work on synchronization phenomena in neural networks and on the modalities of extra-synaptic transmission by diffusion (cf. work on "gasnet", Husbands, Smith, Jakobi, and O'Shea, 1998), which seems especially relevant to the question of learning, insofar as the phenomena correlated with this type of transmission have long time constants.

A provisional assessment and the beginnings of a self-organizing vision of learning

Systems thinking, culminating in connectionism, has stressed the importance of understanding the dynamics of coupled elements. However, this thinking has remained representationalist and cephalocentric. A conceptual rejuvenation has been most noticeable at the crossroads of ecological approaches and the theory of non-linear dynamic systems applied to development sciences (Thelen and Smith, 1994). An important first point is that dynamics (and their modeling) cannot be limited to the organism, but must include the organism's situation (both present and historical) together with all the components of that situation (environmental, social, and cultural). The causality of learning goes beyond the subject. A second point is that the dynamic linking the various elements requires the subject to be active in order to be stabilized from that subject's perspective. And today, numerous scientific studies are attempting to characterize how a subject's perceptual and cognitive invariants are constituted, which once again only makes sense if we consider the subject/world relationship. This relationship may or may not be technically instrumented. A third point is that the learning of these invariants mobilizes, in some cases, social and technical devices designed to make explicit the path to be taken. In this case, the distributed dimension of learning is multiplied tenfold (see the question of support systems, Bachimont, 2010), making the analysis of learning much more difficult, as shown by certain ethnological (Mondada, 2005) and anthropological (Bril and Roux, 2002) studies, for example. The introduction of non-linear dynamical systems theory into certain areas of the human sciences, particularly in the perceptual-motor development sciences (cf. the numerous important works of Thelen and his collaborators; Thelen and Smith, 1994), has profoundly influenced conceptions of ontogeny and learning. This approach is in line with the historical extension of Piagetian constructivism, which posits the origin of perception and cognition in the acted relationship to the world. From the outset, the organism is considered to be coupled with its environment; the characterization of the subject's behavior and of how this behavior changes draws on multiple elements that transcend the subject itself. All the elements are considered to be heterarchically coupled (in other words, no single element alone is the cause of the phenomenon, and the part played by different elements is subject to situation-related weighting). The transformation of behavior is therefore not the result of a program or an environmental footprint, but of the self-organization (constraint emergence) inherent in the system's temporal deployment. What is more, mathematical models – most of them qualitative – have been produced to characterize the properties of the system's dynamics, in particular its non-linearities (bifurcations) and stability (attractors). One of the crucial contributions of this approach is that it posits variability as intrinsic, and considers regularities as the result of coupling history. Learning is conceived as a particular mode of transformation enabling living and thinking forms to construct new behaviors different from those defining the species' basic repertoire. In this definition, proposed by Kelso in 1995, learning implies the organism's active coupling with its environment and, at this level, takes up Wiener's hypothesis of global, not just

local, system change. In many situations, it is possible to show that the heterogeneity of initial dynamics gives way to a gradual convergence towards the coupling strategy that appears optimal (Haken, Kelso and Bunz, 1985; Thelen, Corbetta and Kamm, 1993). This has led this tradition to consider ontogeny and learning as part of selectionist processes, which can be traced at the biological level and modeled (*e.g.*, Sporns, Tononi and Edelman, 2000 or the work of the "Epigenetic Robotics" community on this subject). One of the questions that remains open in this approach is the question of the direction (and motivation) of behavioral transformation. How can we account, in a physicalist approach of this kind, for the desire to learn and the choice (preference, direction) of what is learned?

To begin addressing this question, we note that in almost all the computational approaches to learning that we know of, whether in artificial intelligence, robotics (including epigenetic/developmental robotics), or computational neuroscience, the repertoire of actions is fixed. In other words, the modeler chooses in advance which actions will have relevance for an agent seeking to accomplish a task, and it remains for the agent to learn what the values of these actions are in different states of the task, *i.e.*, to learn how to select the relevant actions. In some models there is a dynamic, incremental repertoire of states to represent the different parts of the task (Khamassi et al., 2006; Redish 2007), which is already less constrained by the *a priori* knowledge included by the modeler. But most of the time, there is no possibility of discovering new actions. There is therefore still much room for exploration in regard to finding efficient computational mechanisms for the discovery of new actions, and investigating to what extent these mechanisms might account for the discovering of new actions in living organisms. It is interesting to note that the epigenetic robotics community has recently begun to propose representation redescription mechanisms, whereby events experienced by an agent executing a task are mentally re-simulated and analyzed as a means of identifying possible representations of states and actions to best explain the dynamics of the events observed during the task (Doncieux et al., 2018).

VIII. Learning as the constitution of a plastic and specific structure of anticipation

Inspired by the dynamicist tradition, a number of research groups have been attempting in recent years to develop a "genetic constructivist" approach that would represent a decisive shift in what has up to now been an informational, computational and internalist conception of cognitive experience. Proponents of the enactive approach, which is sometimes qualified as *non-representational*, do not see cognition as resulting from internal mechanisms that reconstitute the world on the basis of one or more sensory channels. Rather, their objective is identifying the minimal structural and functional conditions required for the formation of an experience likely to enable the deployment of autonomous, adapted and effective behavior. Moreover, seen through the enactive lens, it is the very functioning of an agent that contributes to that agent's own transformation. This is a spiral, rather than circular, conception of cognitive functioning, the elementary form of which, in this case, is the sensory-motor¹⁵, or rather, moto-sensory loop. Varela (1989, p.199) proposes the concept of coupling by closure¹⁶, which "*consists in thinking that the nervous system is defined essentially by its various modes of internal coherence, which derive from its interconnectivity.*" Signals from the various neural

¹⁵ While a number of theoretical frameworks highlight the constitutive role of movement, the developmental or ontogenetic dimension is rarely addressed. There are, however notable exceptions, such as Piagetian theory, and Maturana and Varela's autopoiesis-enaction theory.

¹⁶ It will be remarked that "closure" is not the same as "closed". Here we are talking about open, organizationally closed systems. It is a concept is very close, at least in spirit, to several others, such as schema (Piaget), functional circles (Von Uexhüll), dynamic invariant (Thelen), and transductive relation (Simondon).

circuits (retina, cochlea, neuromuscular spindles, semicircular canals, olfactory receptors, etc.) induce organizational disturbances of varying amplitude. This brings us to our final definition:

Learning definition #5: Learning consists in organizing the succession of actions in relation to a new coherence of reafferent signals.

In other words, learning is a mechanism for building up a structure of anticipation, leading to a transformation and enrichment of the learner system's own behaviors. A substantial body of existing research makes it possible to specify the modalities of this mechanism through modeling (*e.g.* Philipona, O'Regan and Nadal, 2003; Stewart and Gapenne, 2004) and experimentation (*e.g.* Lenay and Stewart, 2012; Roy *et al.*, 2019).

Against this background we also note the emergence of new approaches in computational modeling, such as "radical interactionism", whose originality lies in the reformulation of the "observation/action" model of the reinforcement learning theory (described above) into an "experience/result" model (also called "interaction"; Gay *et al.*, 2017). The objective is then to endow an agent with the ability to construct variables estimating spatiotemporal regularities relative to the context, enabling it to predict, based on limited sensory feedback, which interactions with the environment are possible, so as to better enable learning that furthers the agent's intrinsic motivations. The model considers perception and action to be inseparable, forming a sensorimotor schema. The work undertaken to date with this model has sought in particular to identify ways in which the contextual part of these sensorimotor schemas may be enriched through the learning of associations between spatio-temporal regularities, and the construction of structures characterizing near (peri-personal) and far (extra-personal action) space, by means of a spatial memory capable of processing interactions.

Hypothesis

Let us now try to clarify matters. Our hypothesis in this enactive framework is that the particular nature of learning lies in the nature of the feedback (or feedbacks) that guide action, allowing anticipation to be created, and enabling a novel mode of control, known as prospective control, within the feedback-action coupling. By prospective control we mean the idea of temporal extension of a self-organizing, metastable process (in the sense of Simondon, 2005). This presupposes tolerance of future functional uncertainties (the principle of functional non-univocity), which in turn presupposes a certain structural plasticity such that conflicts within the feedback-action coupling can be resolved. Although we must act in order to learn, not all actions give rise to learning, since not all reafferences of action have the potential to reconfigure the organization of the unfolding structure. In the enactive approach, reafferences essentially have the status of disturbances¹⁷ that modulate bodily dynamics and which are able to exercise control over those bodily dynamics only when they acquire the status of constraints. Reafferences are not defined as informational inputs or outputs. Thus, our enactive definition of learning is in a sense compatible with von Foerster's (1969) vision of recursivity, which posits that memory is a memory without explicit information storage, insofar as the dynamic or operation carries the entirety of its history at all times.

Reflexive loop

¹⁷ Simondon (2005) points out that stimulation is what enables a being to initiate the transition from the potential dynamic state to the "actual" state of accumulated energy. He also points out that stimulation can only be effective if it encounters a being in an unstable state.

To this definition, which is rooted in the biological and moto-sensory organization of a living agent, we need to add an element to account for the precondition that the subject "knows" that it is learning in order to learn, in other words, that it experiences its own transformation. The aim here is to characterize the more specifically human dimension of learning, using the same level logic as for language 1 and 2 (communication versus narrative) or consciousness 1 and 2 (direct versus reflexive): what would "learning 2" be like? Adopting a dynamicist perspective, seeking to specify the nature of the variables (control parameters) that constrain and empower dynamics, we still need to understand how the subject (organism or system) apprehends through its actions the elements that will enable the reorganization of its behavior. One solution would be to introduce a reflexive loop that would enable the subject to self-assess (even implicitly) and to know (or, more simply, feel) whether and how it is changing. This would have the merit of corresponding to a certain supposed or observed phenomenology (reminiscent of Kamin's (1969) retrospective contemplation mentioned earlier), but it entails the classic risk of tangling different levels of experience. As Rosenthal (2004) reminds us, the stumbling block of classical approaches to learning is encapsulated in a paradox voiced in Plato's Socratic dialogue *Meno*: learners do not know what they need to know. And even where they acquire initially unknown knowledge, how can they know that it is what they need? Here we have the problem of the orientation of the learning mechanism and the evaluation of transformative pathways. There are two possible, complementary solutions to this apparent paradox: the first is to consider that there is a morphogenetic orientation intrinsic to dynamics, and that any schema can become anticipatory once it has been constructed through the deployment of what Piaget calls extrapolation and recurrence (1967, p.188-189). In the latter case, the anticipatory structure unfolds in both directions of time, relative to the temporal core constituted by the primary pattern. This first solution can be supplemented by a second one, which considers the question of help (Gapenne, 2006) and thus the technical, social and cultural dimensions of learning; the second solution can overcome this paradox, but only from a certain point of view, since, ultimately, learning must be assumed and experienced by the learner itself. These two solutions lead us briefly to the social aspects of learning, including its socio-technical components.

Learning, development and socio-technical environment

This is precisely what was at the heart of the Vygotskian approach. One of Vygotsky's (1985) preoccupations was to define learning and development as two clearly distinct but not independent mechanisms; for him, their dependent relationship operates as a basis for ontogenetic development by learning, as a relationship to tools and to other learners¹⁸. A first element of differentiation is to consider that learning can take place on a variety of time scales, which can be temporally circumscribed and included in the vital cycle that constitutes the time scale of development, which, for its part, nevertheless presents non-linearities. Moreover, although people today often talk about lifelong learning, we should not confuse this with development. Learning does not always succeed (contrasting with the robustness of the living ontogenetic process), and it requires an effort and a commitment (different from a living being's commitment to its own survival). Human beings are embroiled in the vital cycle and assume

¹⁸ Yvon and Clot (2003) point out that at least three others aspects have been highlighted, namely independence, identity, and symmetrical dependence. As our position is to define learning as a specific relational (or coupling) mode, distinct from the more elementary modes of relationship specific to living beings, we propose considering learning and development as two distinct mechanisms, functionally interdependent and therefore influencing each other. This position was elaborated in the Gestalt tradition, in particular by Kofka. Kelso (1995) takes a similar stance when he writes "*People enter the learning situation with a certain degree of pre-organization that constraints the form that learning takes* (p.184)". The dynamics of this pre-organization obviously have their own survival constraints.

their permanent transformation; but they also have the opportunity, also permanent, to organize their transformation. Perception serves as good example of this. As Eleanor Gibson (1969) reminds us, situations in which perceptual learning occurs even in adult subjects are very frequent, and indicate that perception can be constructed at any age, provided that the conditions of biological plasticity and commitment allow it. However, certain situations show that perceptual learning is not always possible. One example is the inability of certain species (including chickens and toads) to adapt to a prismatic deviation. Hatwell (1966) and Jeannerod (1975) note that it is very rare that congenitally blind people are able to attain a full visual experience following surgery. Vision generally remains very limited¹⁹, to such an extent that those concerned often decide to forego this learning opportunity. It would appear that a set of very favorable conditions is required for learning to take place (motivation, social support, education). From this point of view, Vygotsky's work on the proximal zone of development, a notion taken up by Bruner in the concept of scaffolding, strongly supports the importance of the social in the individual dynamic of learning.

An important area of study that is contributing to our understanding of the situated, distributed and socialized approach to learning is examining how specialized technical skills such as turning or stone-cutting, and also certain domestic skills, are learned. Bril and Roux (2002), for example, showed that the way in which a social "calendar" of apprenticeships is created, the way in which technical gestures themselves are learned, and the way in which these things are brought to bear on the economic life of the community, have structural aspects in common. In addition – and this is a crucial point – at the heart of this kind of system of learning is a relationship between the learner and the teacher, which can be described as helping, and which mobilizes many of the mechanisms at play in scaffolding situations. This has clear echoes of the notion of the zone of proximal development, mentioned above. And going beyond the teacher, it is the co-presence of learners and "experts" (learning often takes place in workshops) that needs to be considered. We expect this relatively recent intellectual current to be reflected in the scientific orientation of future work on learning, with researchers looking beyond the subject and considering learning in its instrumental but also and above all social dimension.

Computational neuroscience has recently turned its attention to what are, for the moment, very simple dimensions of social cognition, seeking to integrate these dimensions with the already well-known mechanisms of reinforcement learning in non-social contexts. The laboratory tasks implemented often involve the presence of an acting conspecific, which must be directly taken into account by the learning agent, in the sense of an imitation that adds to the classical variables of reinforcement learning models (Burke et al., 2010). While these approaches need to be developed further to provide a better understanding of the complexity of possible teacher-learner relationships in settings that are more ecological than that of the laboratory, we remark that this work appears to have already identified, in certain parts of the prefrontal cortex, a degree of reuse of some of the reinforcement learning mechanisms usually used for non-social contexts, and here extended to the learning of the trust that can be attributed to this or that agent/teacher (Behrens et al., 2009).

IX. Conclusion

¹⁹ Where the operation is not done during the first two or three years of life, amblyopic vision would appear to be the maximum possible visual attainment. Post-operative phenomena such as nystagmus and painful glare may persist and impair learning.

We have seen that the quest to understand the learning process scientifically is a recent undertaking, which essentially began with the work of Ivan Pavlov in the early 20th century. The way that thinking about learning subsequently evolved may be condensed into a handful of succinct definitions, corresponding to the great meta-theoretical frameworks of associationism, computationism, emergentism and constructivism, of which the enactive approach is a particular form, doubly heir to Piagetian work and the phenomenological tradition in philosophy and psychology. Perhaps more noticeably than research into other so-called cognitive phenomena, research into learning is above all based on biological considerations. And, as is often the case in cognitive science and technology, computational and, to some extent, emergent approaches have provided formal and algorithmic tools for simulating learning behaviors. However, this behavioral mimicry says nothing about the difference in the material realizations of learning in these two types of beings, living and computational, and, moreover, says nothing about the necessary distinction between ontogeny and learning. And although some researchers in these fields have sought to base the operation and the constraints of their artificial entities on those of natural systems, in so doing they have generally seen living systems as producing coupling by input in accordance with Varela's hypothesis that "*the nervous system functions on the basis of the informative content of instruction from the environment, and it functions by elaborating an operational representation of this environment.*" (Varela, 1989, p.199).

On the strength of the various considerations that we have outlined above, we are inclined to give a positive answer to the question formulated in the title of this article, and to acknowledge that the enactive approach does indeed provide the framework for a new definition of learning. After the "prediction-association", "acquisition-representation" and "emergence-auto-organization" pairs, a pair linking "anticipation and autonomy" (and implying a coupling by closure) seems to us to be an appropriate scientific focal point for extending the enactive approach to learning and initiating new computational modeling.

Acknowledgements

The authors would like to warmly thank David Lewis for his detailed comments and precious help with English.

References

- Anderson, J.R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Arleo, A., & Gerstner, W. (2000). Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biological cybernetics*, 83(3), 287-299.
- Bachimont, B. (2010). *Le sens de la technique : Le numérique et le calcul*. Paris, éditions Encre Marine.
- Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, 35(1), 48-69.
- Bartlett, F.C. (1932). *Remembering*. Cambridge: Cambridge University Press.
- Bastien, C. (1997). *Les connaissances de l'enfant à l'adulte*. Paris: Armand Colin.
- Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behavior. *science*, 324(5931), 1160-1164.

- Bril, B. & Roux, V. (2002). Le geste technique : réflexions méthodologiques et anthropologiques. Numéro spécial de la revue *Technologies, Idéologies, Pratiques, Volume 14*.
- Bruner, J. (1957). *Going beyond the information given*. In *Contemporary approaches of cognition*. Cambridge: Harvard University Press
- Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences*, 107(32), 14431-14436.
- Carlier, P. & Jamon, M. (2006). Observational learning in C57BL/6j mice. *Behavioural Brain Research*, 174, 125-131.
- Carlier, P. & Lefebvre, L. (1997). Ecological differences in social learning between adjacent, mixing populations of Zenaida doves. *Ethology*, 103, 772-784.
- Cazé, R., Khamassi, M., Aubin, L., & Girard, B. (2018). Hippocampal replays under the scrutiny of reinforcement learning models. *Journal of neurophysiology*, 120(6), 2877-2896.
- Costa, V. D., Mitz, A. R., & Averbeck, B. B. (2019). Subcortical substrates of explore-exploit decisions in primates. *Neuron*, 103(3), 533-545.
- Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in cognitive sciences*, 10(7), 294-300.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8(12), 1704-1711.
- Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, 2(4), 313.
- Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. *Stevens' handbook of experimental psychology*, 3, 497-533.
- Dollé, L., Chavarriaga, R., Guillot, A., & Khamassi, M. (2018). Interactions of spatial strategies producing generalization gradient and blocking: A computational approach. *PLoS computational biology*, 14(4), e1006092.
- Doncieux, S., Filliat, D., Díaz-Rodríguez, N., Hospedales, T., Duro, R., Coninx, A., ... & Sigaud, O. (2018). Open-ended learning: a conceptual framework based on representational redescription. *Frontiers in neurobotics*, 12, 59
- Engel, A.K. & König, P. (1993). Paradigm shifts in the neurobiology of perception. In U. Ratsch, M.M.Richter & I.O. Stamatescu (Eds), *Intelligence and artificial intelligence* (pp.178-192). Berlin: An Interdisciplinary Debate. Springer.
- Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T. K., Ueda, Y., ... & Kimura, M. (2011). Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proceedings of the National Academy of Sciences*, 108(37), 15462-15467.
- Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PloS one*, 4(7), e6421.
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience*, 6(4), 187-214.
- Gapenne, O. (2006). Système d'aide et transformation cognitive. *Intellectica*, 44, 7-16.
- Gay, S. L., Mille, A., Georgeon, O. L., & Dutech, A. (2017). Autonomous construction and exploitation of a spatial memory by a self-motivated agent. *Cognitive Systems Research*, 41, 1-35.
- Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning & behavior*, 40, 255-268.
- Ghavamzadeh, M., Mannor, S., Pineau, J., & Tamar, A. (2015). Bayesian reinforcement learning: A survey. *Foundations and Trends in Machine Learning*, 8(5-6), 359-483.

- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, N.J.: Prentice-Hall Inc.
- Giraud-Carrier, C., Vilalta, R., & Brazdil, P. (2004). Introduction to the special issue on meta-learning. *Machine learning*, 54, 187-193.
- Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour*, 66(1), 15-36.
- Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, 17(11), 585-593.
- Haken, H., Kelso, J.A.S. & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51, 347-356.
- Harlow, H.F. (1949). The formation of learning sets. *Psychological Review*, 56, 51-65.
- Hatwell, Y. (1966). *Privation sensorielle et intelligence*. Paris: PUF.
- Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior*, 4(60), 78-60.
- Holland, P.C. (1984). Acquisition of representation-mediated conditioned food aversion. *Learning and Motivation*, 12, 1-18.
- Holland, P.C. & Straub, J.J. (1979). Differential effects of two ways of devaluing the unconditioned stimulus after first- and second-order appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior and Processes*, 5, 65-78.
- Hull, C.L. (1943). *Principles of behavior*. New York:Appleton-Century-Croft.
- Husbands, P., Smith, T.M.C., Jakobi, N. & O'Shea, M. (1998). Better living through chemistry: evolving GasNets for robot control. *Connection Science*, 10(3-4):185-210.
- Jeannerod, M. (1975). Déficit visuel persistant chez les aveugles-nés opérés. *Année Psychologique*, 1, 69-95.
- John, E.R., Chesler, P., Bartlett, F. & Victor, I. (1968). Observation Learning in cats. *Science*, 159, 1489 – 1491.
- Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B.A. Campbell & R.M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, pp. 279-296.
- Kelso, J.A.S. (1995). *Dynamic Patterns. The Self-Organization of Brain and Behavior*. Cambridge: The MIT Press
- Khamassi, M., Martinet, L. E., & Guillot, A. (2006, September). Combining self-organizing maps with mixtures of experts: application to an actor-critic model of reinforcement learning in the basal ganglia. In *International Conference on Simulation of Adaptive Behavior* (pp. 394-405). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Khamassi, M., Wilson, C., Rothé, R., Quilodran, R., Dominey, P. F., & Procyk, E. (2011). Meta-learning, cognitive control, and physiological interactions between medial and lateral prefrontal cortex. *Neural basis of motivational and cognitive control*, 351-370.
- Khamassi, M., & Humphries, M. D. (2012). Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. *Frontiers in behavioral neuroscience*, 6, 79.
- Khamassi, M., Quilodran, R., Enel, P., Dominey, P. F., & Procyk, E. (2015). Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. *Cerebral cortex*, 25(9), 3197-3218.
- Klahr, D. & Wallace, J.G. (1976). *Cognitive development: An information-processing view*. Hillsdale: Erlbaum
- Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behaviour. *Journal of Cognitive Neuroscience*, 16, 1412–1425
- Köhler, W. (1925). *The mentality of apes*. New York: Harcourt.

- Kohonen, T. (1991). Self-organizing maps: Optimization approaches. In *Artificial neural networks* (pp. 981-990). North-Holland.
- Langille, J.J. & Brown, R.E. (2018). The synaptic theory of memory: A historical survey and reconciliation of recent opposition. *Frontiers in Systems Neuroscience*, 12, 52.
- LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10), 1995.
- Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81(3), 687-699.
- Lenay, C. (2006). La méthode minimaliste : phénoménologie et science objective. *Intellectica*, 43, 125-134.
- Lenay, C. & Stewart, J. (2012). Minimalist approach of perceptual interactions. *Frontiers in Human Neurosciences*, 6.
- Lewicki, P., Hill, T. & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, 20, 24-37.
- Lindsay, P.H. & Norman, D.A. (1972). *Human information processing: An introduction to psychology*. New York: Academic Press.
- Livet, P. (2004). La notion de récursivité, de la première cybernétique au connexionnisme. *Intellectica*, 39, 125-137.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychol Rev*, 82(4), 279-298.
- Maes, E. J. P., Sharpe, M. J., Usypchuk, A. A., Lozzi, M., Chang, C. Y., Gardner, M. P., Schoenbaum, G., & Iordanova, M. D. (2020). Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nature neuroscience*, 23(2), 176-178.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman Publishers
- Martens, M.B, Celikel, T & Tiesinga, T.H.E (2015). A developmental switch for Hebbian plasticity. *Plos Computational Biology*, 19 p.
- Mondada, L. (2005). La constitution de l'origo déictique comme travail interactionnel des participants : une approche praxéologique de la spatialité. *Intellectica*, 41/42, 75-100.
- Moran, R., Keramati, M., Dayan, P., & Dolan, R. J. (2019). Retrospective model-based inference guides model-free credit assignment. *Nature communications*, 10(1), 750
- Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature communications*, 6(1), 8096.
- Panayi, M.C., Khamassi, M., & Killcross, S. (2021). The rodent lateral orbitofrontal cortex as an arbitrator selecting between model-based and model-free learning systems. *Behavioral Neuroscience*, 135 (2), pp.226-244.
- Pavlov, I. P. (1927). *Conditioned reflexes*. Londres : Oxford University Press.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev*, 87(6), 532-552.
- Perruchet, P. (1997). Le rôle de l'attention dans les apprentissages implicites. In D. Mellier & A. Vom Hofe (Eds.), *Attention et contrôle cognitif*. Rouen : Presses Universitaires de Rouen, pp. 19-27.
- Perruchet, P. & Vinter, A. (2002). The Self-Organizing Consciousness (Target paper). *Behavioral and Brain Sciences*, 25, 297- 330.
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in neurobiology*, 134, 17-35.
- Piaget, J. (1967). *Biologie et connaissance*. Neuchâtel et Paris : Delachaux et Niestlé.
- Pickens, C. L., & Holland, P. C. (2004). Conditioning and cognition. *Neuroscience and Biobehavioral Reviews*, 28(7), 651-661.

- Philipona, D., O'Regan, J.K., & Nadal, J.P. (2003). Is there something out there? Inferring space from sensorimotor dependencies. *Neural Computation*, 15, 2029-2049.
- Redish, A. D., & Touretzky, D. S. (1997). Cognitive maps beyond the hippocampus. *Hippocampus*, 7(1), 15-35.
- Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychological review*, 114(3), 784.
- Rescorla, R.A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 66, 1-5.
- Rescorla, R.A & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A.H. Black & W. F. Prokasy (Eds.), *Classical conditioning II*. New York: Appleton-Century-Crofts, pp.64-99.
- Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in cognitive sciences*, 14(2), 64-71.
- Rosenthal, V. (2004). Formes, sens et développement : quelques aperçus de la microgenèse. *Texto!* revue électronique (http://www.revuetexto.net/Inedits/Rosenthal_Formes.html).
- Roy, V., Menant, O., Hacquemand, R., Vallier, C., Enschede, E., Rovira, K. & Gapenne, O. (2019). A case of visuo-auditory sensory substitution in rats. *Behavioral Processes*, 164, 157-166.
- Schmidhuber, J. (1987). *Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook* (Doctoral dissertation, Technische Universität München).
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599.
- Schweighofer, N., & Doya, K. (2003). Meta-learning in reinforcement learning. *Neural Networks*, 16(1), 5-9.
- Seo, H., & Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *Journal of neuroscience*, 27(31), 8366-8377.
- Shanon, B. (1993). *The representational and the presentational*. New York: Harvester Wheatsheaf.
- Simondon, G. (2005). *L'individuation à la lumière des notions de forme et d'information*. Grenoble: Millon.
- Skinner, B.F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Smith, T., Husbands, P., Philippides, A. & O'Shea (2003). Neuronal plasticity and temporal adaptivity: GasNet robot control. *Adaptive Behavior*, 10, 161-183.
- Sporns, O., Tononi, G., and Edelman, G.M. (2000). Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex* 10, 127-141.
- Stewart, J. & Gapenne, O. (2004), Reciprocal modelling of active perception of 2-D forms in a simple tactile-vision substitution system, *Minds and Machines*, 14, 3, 309-330.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press. 2nd Edition.
- Thelen, E. & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.
- Thelen, E. Corbetta, D. & Kamm, K. (1993). The transition to reaching: matching intention and intrinsic dynamics. *Child Development*, 64, 1058-1098.
- Thorndike, E.L. (1913). *The psychology of learning*. New York: Teachers College, Columbia University.
- Tolman E.C. & Honzik, C.H. (1930a). Insight in rats. *University of California Publications in Psychology*, 4, 215-232.

- Tolman E.C. & Honzik, C.H. (1930b). Introduction and removal of reward, and performance in rats. *University of California Publications in Psychology*, 4, 257-275.
- Varela, F.J. (1989). *Autonomie et connaissance. Essai sur le vivant*. Paris : Seuil.
- Varela, F.J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, 3, 330-349.
- Varela, F.J., Thompson, E. & Rosch, E. (1993). *L'inscription corporelle de l'esprit*. Paris: Seuil.
- Vernon, M.D. (1954). *A further study of visual perception*. Cambridge: Cambridge University Press.
- Visetti, Y.M. (2004). Constructivismes, émergences : Une analyse sémantique et thématique. *Intellectica*, 39, 229-259.
- Von Foerster, H. (1969). What is memory that it may have hindsight and foresight as well. *The Future of Brain Sciences*. Bogoch ed, Plenum Press, pp. 19-64.
- Vygotsky, L.S. (1985). *Pensée et langage*. Paris : Editions sociales
- Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., ... & Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature neuroscience*, 21(6), 860-868.
- Wagner, A.R., Rudy, J. & Whitlow, J.W. (1973). Rehearsal in animal conditioning. *Journal of Experimental Psychology*, 97, 407-426.
- Wiener, N (1954). *The human uses of human beings (Cybernetics and Society)*. Houghton Mifflins.
- Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, 143(6), 2074.
- Wittmann, M. K., Fouragnan, E., Folloni, D., Klein-Flügge, M. C., Chau, B. K., Khamassi, M., & Rushworth, M. F. (2020). Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. *Nature communications*, 11(1), 3771
- Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3), 356-365.
- Yvon, F. & Clot, Y. (2003). Apprentissage et développement dans l'analyse du travail enseignant. *Pratiques Psychologiques*, 1, 19-35.